

# Math 231: Linear Algebra

Darshan Patel

Fall 2016

## Contents

## 1 Systems of Linear Equations and Matrices

### 1.1 Introduction to Systems and Linear Equations

Let  $ax + by = c$  where at least one of the constants is nonzero. This equation will have a straight line as its graph. Consequently, every straight line has an equation that can be written in this form. Because of this, we refer to an equation in this form as a linear equation (in the variables  $x$  and  $y$ ).

Consider  $l_1$  and  $l_2$ , two straight lines, in a plane. there are 3 possibilities:

1.  $l_1$  and  $l_2$  are parallel
2.  $l_1$  and  $l_2$  intersect at a single point
3.  $l_1$  and  $l_2$  coincide

Let  $a_1x + b_1y = c_1$  and  $a_2x + b_2y = c_2$  be equations for  $l_1$  and  $l_2$  respectively. Consider the system of two linear equations in the two variables  $x$  and  $y$  as:

$$\begin{aligned}a_1x + b_1y &= c_1 \\a_2x + b_2y &= c_2\end{aligned}$$

Then

1. If  $l_1$  and  $l_2$  are parallel, then there is no ordered pairs  $(x, y)$  that satisfy the system.
2. If  $l_1$  and  $l_2$  intersect (properly), then there is a unique ordered pair  $(x, y)$  that satisfy the system.
3. If  $l_1$  and  $l_2$  coincide, then there are infinitely many ordered pairs  $(x, y)$  that satisfy the system.

Summary: Given the system of two linear equations in the two variables  $x$  and  $y$  shown as:

$$\begin{aligned}a_1x + b_1y &= c_1 \\a_2x + b_2y &= c_2\end{aligned}$$

exactly one of the following must be true:

1. the system has no solutions
2. the system has a unique solution
3. the system has infinitely many solutions

Note: In (3), we call the system inconsistent, In (1) and (2), we call the system consistent.

To generalize, we have the following definitions:

- A linear equation in  $n$  variables,  $x_1, x_2, \dots, x_n$ , is an equation of the form  $a_1x_1 + a_2x_2 + \dots + a_nx_n = b$ .
- A solution of a linear equation in the  $n$  variables,  $x_1, x_2, \dots, x_n$ , is a sequence of  $n$  numbers,  $s_1, s_2, \dots, s_n$  with the property that the following makes each equation a true statement:  $x_1 = s_1, x_2 = s_2, \dots, x_n = s_n$ .
- A system of linear equations in the  $n$  variables,  $x_1, x_2, \dots, x_n$ , is a finite collection of equations.
- A solution of a system in the  $n$  variables,  $x_1, x_2, \dots, x_n$ , is a sequence of values that satisfy each and every equation in the system.

If we begin with a system of linear equations in  $n$  variables, we can obtain an “equivalent” system (some solution set) by performing any/all of the following “permissible” maneuvers:

- interchange a pair of equations in the system
- multiply an equation in the system by a nonzero constant
- add a multiple of one equation in the system to another equation in the system

General linear system of  $m$  equations in  $n$  variables  $x_1, x_2, \dots, x_n$ :

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

Associated with this system its an object called its augmented matrix:

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

Note: There are  $m$  rows and  $n + 1$  columns.

Corresponding to the permissible maneuvers, listed previously, we have 3 elementary row operations that can be performed on the augmented matrix:

- interchange 2 pair of rows
- multiple a row by a nonzero constant
- add a multiple of one row to another row

## 1.2 Gaussian Elimination

To transform a matrix into reduced row-echelon form (rref), the following conditions must be met:

1. If the matrix contains any rows containing of all zeros (zero rows), they appear at the bottom of the matrix
2. In any nonzero row, the first nonzero entry reading left to right is a 1 (leading 1)
3. In two successive nonzero rows, the leading 1 in the lower row appears to the right of the leading 1 in the upper row
4. In any column containing a leading 1, all other entries are 0

Note: It can be shown that the reduced row echelon form of a matrix is unique.

If conditions 1-3 are met only, we say that the matrix is in row-echelon form (ref). The process used to transform a matrix into its reduced row echelon form is Gauss-Jordan elimination.

Procedure for Gauss-Jordan Elimination:

1. Step 0: Write the augmented matrix of the system.
2. Step 1: Identify the left most column in the augmented matrix that contains a nonzero entry in it.
3. Step 2: If necessary or desirable, bring a nonzero entry to the top of the column designated in Step 1.
4. Step 3: Multiply the top row by the reciprocal of the entry at the top of the augmented column, thus placing a “1” in that position.
5. Step 4: Add appropriate multiple of the top row to the other ones in order to create zeros in all positions in the designated column.
6. Step 5: Cover the top row and repeat the process. Remember to remove the cover when Step 4 is performed.

Note: If we do not remove the cover in Step 5, the procedure will yield row echelon form (not unique!) and the process is then called Gaussian elimination.

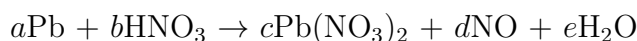
**Example 1.1.** Solve this linear system.

$$\begin{aligned}
 x_2 + 2x_3 + 3x_4 + x_6 &= -1 \\
 2x_1 + 2x_2 + 4x_3 + 8x_4 + 12x_6 &= 2 \\
 -4x_1 + 3x_2 + 6x_3 + 5x_4 + x_5 - 14x_6 &= -10
 \end{aligned}$$

$$\begin{bmatrix} 0 & 1 & 2 & 3 & 0 & 1 & -1 \\ 2 & 2 & 4 & 8 & 0 & 12 & 2 \\ -4 & 3 & 6 & 5 & 1 & -14 & -10 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 2 & 4 & 8 & 0 & 12 & 2 \\ 0 & 1 & 2 & 3 & 0 & 1 & -1 \\ -4 & 3 & 6 & 5 & 1 & -14 & -10 \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} 1 & 1 & 2 & 4 & 0 & 6 & 2 \\ 0 & 1 & 2 & 3 & 0 & 1 & -1 \\ -4 & 3 & 6 & 5 & 1 & -14 & -10 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 2 & 4 & 0 & 6 & 2 \\ 0 & 1 & 2 & 3 & 0 & 1 & -1 \\ 0 & 7 & 14 & 21 & 1 & 10 & -2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 5 & 2 \\ 0 & 1 & 2 & 3 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 3 & 1 \end{bmatrix}$$

**Example 1.2.** Balance the following chemical equation:



$$\text{Pb: } a = c, \text{ H: } b = 2c, \text{ N: } b = 2c + d, \text{ O: } 3b = 6c + d + e$$

System of equations:

$$\begin{aligned}
 a - c &= 0 \\
 b - 2c &= 0 \\
 b - 2c - d &= 0 \\
 3b - 6c - d - e &= 0
 \end{aligned}$$

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 1 & -2 & -1 & 0 & 0 \\ 0 & 3 & -6 & -1 & -1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & -2 & -1 & 2 & 0 \\ 0 & 0 & -6 & -1 & 5 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & -2 & -1 & 2 & 0 \\ 0 & 0 & 0 & 2 & -1 & 0 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & -1 & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & 1 & \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 2 & -1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{3}{4} & 0 \\ 0 & 1 & 0 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & -\frac{3}{4} & 0 \\ 0 & 0 & 0 & 1 & -\frac{1}{2} & 0 \end{bmatrix}$$

Then

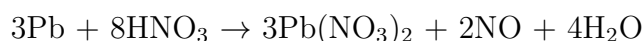
$$a - \frac{3}{4}e = 0 \rightarrow a = \frac{3}{4}e \text{ and } a = 3$$

$$b - 2e = 0 \rightarrow b = 2e \text{ and } b = 8$$

$$c - \frac{3}{4}e = 0 \rightarrow c = \frac{3}{4}e \text{ and } c = 3$$

$$d - \frac{1}{2}e = 0 \rightarrow d = \frac{1}{2}e \text{ and } d = 2$$

$$e = 4$$



A system is said to be homogeneous if and only if all of the b's are equal to zero.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= 0 \end{aligned}$$

A homogenous system is necessarily consistent, since  $x_1 = x_2 = \dots = x_n = 0$  is a solution of the system. This is called the trivial solution of the homogenous system. Then there are only two possibilities for the solution set of such a system:

1. The trivial solution is the unique solution of the system
2. The system has infinitely many solutions, one of which is the trivial one

**Example 1.3.** Solve the homogeneous system.

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_1 + 2x_2 + 3x_3 + 4x_4 &= 0 \\ x_1 - 3x_2 + 5x_3 - 7x_4 &= 0 \end{aligned}$$

$$\begin{aligned} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 4 & 0 \\ 1 & -3 & 5 & -7 & 0 \end{bmatrix} &\rightarrow \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & -4 & 4 & -8 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -1 & -2 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \end{bmatrix} \\ &\rightarrow \begin{bmatrix} 1 & 0 & -1 & -2 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 1 & \frac{1}{3} & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & -\frac{5}{3} & 0 \\ 0 & 1 & 0 & \frac{7}{3} & 0 \\ 0 & 0 & 1 & \frac{1}{3} & 0 \end{bmatrix} \end{aligned}$$

Then:

$$\begin{aligned} x_1 - \frac{5}{3}x_4 &= 0, \quad x_1 = \frac{5}{3}x_4 \\ x_2 + \frac{7}{3}x_4 &= 0, \quad x_2 = -\frac{7}{3}x_4 \\ x_3 + \frac{1}{3}x_4 &= 0, \quad x_3 = -\frac{1}{3}x_4 \\ x_4 &= x_4 \end{aligned}$$

This system has infinitely many solutions.

**Theorem 1.1.** Consider a homogeneous linear system of  $m$  equations in  $n$  variables. If  $m < n$ , then the system has infinitely many solutions. If  $m \geq n$ , then no prediction can be made in general.

*Proof.* Form the augmented matrix of the given system. It has  $m$  rows and  $n + 1$  columns. Then use Gaussian-Jordan elimination to obtain the reduced row echelon form of this matrix. If  $r$  = the number of nonzero rows in reduced row echelon form, then  $r \leq m$ . By hypothesis,  $m \leq n$ . From these two facts, conclude that  $r \leq n$ . If  $r$  = the number of leading 1s in reduced row echelon form,  $r$  = the number of leading variables in the system. Let  $n$  = the total number of variables in the system. So  $(n - r)$  = the number of free variables in the system. Since  $r \leq n$ ,  $n - r$  is an integer that is  $\geq 0$ , that means the system has at least one free variable. Since there are infinitely many choices for a free variable, there are infinitely many solutions of the system.

Note: If we remove the word "homogeneous" from the statement of the theorem, there are no solutions even though  $m < n$ . It is true that if we know a linear system to be consistent, then  $m < n$  and there are infinitely many solutions.  $\square$

### 1.3 Matrices and Matrix Operations

**Definition 1.1.** Matrix: a rectangular array of objects (for our purposes, these objects will be real numbers or symbols that represent real numbers)

Each object in the array is called an entry in the matrix.

If a matrix has  $m$  rows and  $n$  columns, we say that it is an  $m \times n$  matrix. This is called the size of the matrix.

Note: If  $A$  represents such a matrix, then the entry in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $A$  is represented by  $a_{ij}$ .

**Definition 1.2.** Square Matrix: a matrix where  $m = n$  and is thus of order  $n$

The entries  $a_{11}, a_{22}, \dots, a_{nn}$  constitute the main diagonal of square matrix  $A$  of order  $n$ .

**Definition 1.3.** Upper Diagonal Matrix: a matrix where all of the entries in  $n \times n$  matrix  $A$  that are below the main diagonal are equal to 0

**Definition 1.4.** Lower Diagonal Matrix: a matrix where all of the entries in matrix  $A$  that are above the main diagonal are equal to 0

**Definition 1.5.** Diagonal Matrix: a matrix that is both upper triangular and lower triangular

**Definition 1.6.** Scalar Matrix: a diagonal matrix  $A$  that has the property that all of its entries on the main diagonal are equal

**Definition 1.7.** Identity Matrix: a scalar matrix  $A$  that has all entries on its main diagonal equal to 1

**Definition 1.8.** Trace: the sum of the entries of the entries on the main diagonal of  $n \times n$  matrix  $A$ , written as  $\text{tr}(A)$

**Definition 1.9.** Transpose: the  $m \times n$  matrix obtained from  $n \times n$  matrix  $A$  by interchanging its rows with its columns

Let  $A$  and  $B$  be  $2 \ m \times n$  matrices. We say that  $A$  and  $B$  are equal, written as  $A = B$ , if and only if their entries are identical positionally

Let  $A$  and  $B$  be  $2 \ m \times n$  matrices. Then the sum of  $A$  and  $B$ , written as  $A + B$ , is the  $m \times n$  matrix determined from  $A$  and  $B$  by adding their entries positionally

Let  $A$  be a  $m \times n$  matrix and let  $c$  be a scalar (real number). Then  $cA$  is the  $m \times n$  matrix determined from  $A$  by multiplying each of the entries in  $A$  by  $c$

Let  $A$  be a  $m \times n$  matrix. Then the negative of  $A$ , written as  $-A$ , is the  $m \times n$  matrix given by:  $-A = (-1)A$

Let  $A$  and  $B$  be  $2 \ m \times n$  matrices. Then the difference of  $A$  and  $B$ , written as  $A - B$ , is given by:  $A - B = A + (-1)B$

Let  $A$  be a  $m \times p$  matrix and let  $B$  be a  $p \times n$  matrix. Then the product of  $A$  and  $B$ , written as  $AB$ , is the  $m \times n$  matrix whose entry in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column is found as follows: multiple the entries in the  $i^{\text{th}}$  row of  $A$  by the corresponding entries in the  $j^{\text{th}}$  column of  $B$  and add the remaining products. Note: In order to be able to compute  $AB$ , it must be true that: the number of columns of  $A$  = the number of rows of  $B$ .

Assuming that  $AB$  can be computed, the number of rows of  $AB$  = the number of rows of  $A$ . Also, the number of columns of  $AB$  = the number of columns of  $B$

To obtain the  $i^{\text{th}}$  row of  $AB$ , compute  $A_i B$  where  $A_i$  is the  $i^{\text{th}}$  row of  $A$ . Similarly, to obtain the  $j^{\text{th}}$  column of  $AB$ , compute  $AB_j$  where  $B_j$  is the  $j^{\text{th}}$  column of  $B$ .

The general linear system of  $m$  equations in the  $n$  variables,  $x_1, x_2, x_3, \dots, x_n$  can be written as follows:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

This is defined as  $AX = B$  where  $A$  is the coefficient matrix of the system,  $X$  is the matrix of the variables and  $B$  is the matrix of constants

## 1.4 Non-Rules and Rules of Matrix Arithmetic

Non-rules:

1. It is not generally true that  $AB = BA$ . That is, there is no commutative law for matrix multiplication
2. There is no zero product law for matrix. That is, if  $AB = 0$  (zero matrix), it is not necessarily true that  $A = 0$  or  $B = 0$
3. There is no cancellation law for matrices

Rules: (Throughout this list, uppercase letters represent matrices, lowercase letter represent scalars and all indicated operations are assumed be able to be done)

1. Commutative Law of Matrix Addition:  $A + B = B + A$
2. Associative Law of Matrix Addition:  $A + (B + C) = (A + B) + C$
3. Existence of an Additive Identity Element:  $A + 0 = A, 0 + A = A$
4. Existence of an Additive Inverse:  $A + (-A) = 0, (-A) + A = 0$
5. Distributive Law:  $c(A + B) = cA + cB$
6. Distributive Law:  $(c + d)A = cA + dA$
7. Associative Law of Scalar Multiplication:  $(cd)A = c(dA)$
8. Existence of a Scalar Multiplicative Identity Element:  $1A = A$
9. Distributive Law:  $A(B + C) = AB + AC$
10. Distributive Law:  $(A + B)C = AC + BC$
11. Associative Law of Matrix Multiplication:  $A(BC) = (AB)C$
12. Zero Factor Law:  $A0 = 0, 0A = 0$
13. Portability of the Scalar:  $c(AB) = (cA)B = A(cB)$

*Proof.* We will prove that for any linear system, the solution set is either empty, a singleton or infinite.

Consider the linear system  $AX = B$ . If the solution set is either empty or a singleton, we're done. It remains to show that if the system has more than one solution, it must have infinitely many solutions. Suppose  $X_1$  and  $X_2$  are 2 distinct solutions of this system. This means that  $AX_1 = B$  and  $AX_2 = B$ . Define  $X^* = tX_1 + (1 - t)X_2$ , where  $t$  is a real number. Then:

$$\begin{aligned}
 AX^* &= A(tX_1 + (1 - t)X_2) \\
 &= t(AX_1) + (1 - t)AX_2 \\
 &= tB + (1 - t)B \\
 &= (t + (1 - t))B \\
 &= 1B \\
 &= B
 \end{aligned}$$

This shows that  $X^*$  is a solution set of the given system. Since  $t$  is arbitrary, there are infinitely many  $X^*$ 's, one for each choice of  $t$ .  $\square$

If  $A$  is a  $m \times n$  matrix,  $I_n$  is a  $n \times n$  matrix and  $I_m$  is a  $m \times m$  matrix, then  $AI_n = A$  and  $I_mA = A$ . This justifies calling  $I$  an identity matrix for matrix multiplication.

Let  $A$  be an  $n \times n$  matrix. We say that  $A$  is invertible if and only if we can find an  $n \times n$  matrix  $B$  such that  $AB = I_n$  and  $BA = I_n$ . If such a matrix  $B$  can be found, we call it an



inverse of  $A$ .

Note: Not every square matrix is invertible.

If  $A$  is an invertible  $n \times n$  matrix, it has a unique inverse.

*Proof.* Suppose to the contrary, that  $A$  has 2 distinct inverses. Let  $B$  and  $C$  be the 2  $n \times n$  matrices such that  $AB = I_n$ ,  $AC = I_n$  and  $B \neq C$ . Also, let  $BA = I_n$ ,  $CA = I_n$ . We can then say:

$$B = I_n B = (CA)B = C(AB) = C(I_n) = C$$

This contradicts the distinctness of  $B$  and  $C$ . Conclusion:  $A$  can only have 1 inverse.  $\square$

If  $A$  is invertible, we can now speak about this inverse of  $A$  as  $A^{-1}$ . Thus:  $AA^{-1} = I_n$  and  $A^{-1}A = I_n$ .

If  $A$  is a  $2 \times 2$  matrix, say  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , and if  $ad - bc \neq 0$ , then  $A$  is invertible and

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

**Theorem 1.2.** Let  $A$  and  $B$  be 2 invertible  $n \times n$  matrices and let  $c$  be a nonzero scalar, then

1.  $A^{-1}$  is invertible and  $(A^{-1})^{-1} = A$
2.  $cA$  is invertible and  $(cA)^{-1} = \frac{1}{c}A^{-1}$
3.  $AB$  is invertible and  $(AB)^{-1} = B^{-1}A^{-1}$

*Proof.* In each case, we will verify that (matrix)(proposed inverse) =  $I_n$ . We need also show that (proposed inverse)(matrix) =  $I_n$  but later in the chapter, we will know that if the first is true, then the second is true as well.

1.  $(A^{-1})(A) = I_n$  So  $A^{-1}$  is invertible and  $(A^{-1})^{-1} = A$
2.  $(cA)(\frac{1}{c}A^{-1}) = (c)(\frac{1}{c})(A)(A^{-1}) = 1I_n = I_n$  So  $(cA)^{-1}$  is invertible and  $(cA)^{-1} = \frac{1}{c}A^{-1}$
3.  $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = A(I_n)A^{-1} = AA^{-1} = I_n$  So  $(AB)^{-1}$  is invertible and  $(AB)^{-1} = B^{-1}A^{-1}$

$\square$

Generalization of (iii): If  $A_1, A_2, A_3, \dots, A_k$  are  $k$  invertible  $n \times n$  matrices, then  $A_1 A_2 A_3 \dots A_k$  is invertible and

$$(A_1 A_2 A_3 \dots A_k)^{-1} = A_k^{-1} \dots A_2^{-1} A_1^{-1}$$

Suppose  $A^k$  is an  $n \times n$  matrix. If  $k$  is a positive integer, then

$$A^k = AAA \dots A$$

( $k$  copies of  $A$ )

Suppose we wish to define  $A^0$ . It is defined as:  $A^0 = I_n$ , provided  $A \neq 0$ .

Finally, we can define  $A^{-k}$ . It is defined as:  $A^{-k} = (A^{-1})^k$

With these definition and with appropriate conditions satisfied, the familiar laws of exponents remain in effect. That is, if  $A$  is a square matrix and if  $r$  and  $s$  are integers, then

$$A^r A^s = A^{r+s}$$

and

$$(A^r)^s = A^{rs}$$

In the generalization of (iii), suppose that  $A_1 = A_2 = A_3 = \cdots = A_k = A$ . Then we get  $(AAA \dots A)^{-1} = A^{-1}A^{-1}A^{-1} \dots A^{-1}$ , which means  $(A^k)^{-1} = (A^{-1})^k$ . That is, if  $A$  is invertible, then so is  $A_k$  where  $k$  is a positive integer.

**Example 1.4.** If  $A = \begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix}$ , find  $A^2, A^3, A^{-2}$ .

$$A^2 = \begin{bmatrix} 14 & 5 \\ 25 & 9 \end{bmatrix}$$

$$A^3 = AA^2 = \begin{bmatrix} 67 & 24 \\ 120 & 43 \end{bmatrix}$$

$$A^{-2} = (A^2)^{-1} = \begin{bmatrix} 14 & 5 \\ 25 & 9 \end{bmatrix}^{-1} = \begin{bmatrix} 9 & -5 \\ -25 & 14 \end{bmatrix}$$

Let  $A$  be a  $n \times n$  matrix. Suppose  $f(x) = c_k x^k + c_{k-1} x^{k-1} + \cdots + c_2 x^2 + c_1 x + c_0$ , then  $f(A) = c_k A^k + c_{k-1} A^{k-1} + \cdots + c_2 A^2 + c_1 A + c_0$ .

**Example 1.5.** If  $f(x) = x^3 - 4x^2 - 2x + 1$ , compute  $f(A)$  if  $A = \begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix}$ .

$$f(A) = A^3 - 4A^2 - 2A + I_2 = \begin{bmatrix} 6 & 2 \\ 10 & 4 \end{bmatrix} = 2 \begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix} = 2A$$

Note:

$$A^3 - 4A^2 - 2A + I_2 = 2A$$

$$I_2 = 4A + 4A^2 - A^3 = A(4I_2 + 4A - A^2)$$

. This means that  $A^{-1} = 4I_2 + 4A - A^2$ .

If  $D$  is a  $n \times n$  matrix  $\begin{bmatrix} d_1 & & & \\ & d_2 & & 0 \\ & & \ddots & \\ 0 & & & d_{n-1} & \\ & & & & d_k \end{bmatrix}$ , then

$$D^k = \begin{bmatrix} d_1^k & & & 0 \\ & d_2^k & & \\ & & \ddots & \\ 0 & & & d_{n-1}^k & \\ & & & & d_k^k \end{bmatrix}$$

and if  $d_i \neq 0$  for  $1 \leq i \leq n$ , then  $D$  is invertible and

$$D^{-1} = \begin{bmatrix} \frac{1}{d_1} & & & 0 \\ & \frac{1}{d_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{d_{n-1}} & \\ & & & & \frac{1}{d_k} \end{bmatrix}$$

## 1.5 Elementary Matrices: A Method for finding $A^{-1}$ if $A$ is invertible

An  $n \times n$  matrix  $E$  is called elementary if it can be obtained from  $I_n$  by performing exactly ONE elementary row operation on  $I_n$ .

The following is true:

1. If  $E$  is an elementary matrix obtained from  $I_n$  by performing a single elementary row operation on  $I_m$  and if  $A$  is an  $m \times n$  matrix,  $EA$  is the  $n \times n$  matrix obtained from  $A$  by performing the same elementary row operation on  $A$ .

**Theorem 1.3.** Every elementary matrix is invertible and its inverse is another elementary matrix of the same type.

**Example 1.6.** How do you transform matrix  $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$  to matrix  $B = \begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix}$ ?

To transform  $A$  to  $B$  using elementary row operations, we can first interchange the rows of  $A$  and then multiply the second row of  $B$  by 3.

$$\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} A = B$$

. If we call the first matrix  $E_2$  and the second matrix  $E_1$ , then:

$$E_2 E_1 A = B$$

or

$$A = E_1^{-1}E_2^{-1}B$$

Two  $n \times n$  matrices  $A$  and  $B$  are said to row equivalent, written  $A \sim B$  if and only if each can be rewritten from the other by performing a finite sequence of elementary row operations.

**Example 1.7.**

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \sim \begin{bmatrix} 3 & 4 \\ 3 & 6 \end{bmatrix}$$

**Theorem 1.4.** Equivalence Theorem: Let  $A$  be an  $n \times n$  matrix. Then the following statements are equivalent:

1.  $A$  is invertible
2. The homogeneous system  $AX = 0$  has only the trivial solution
3.  $A \sim I_n$

*Proof.* We will prove this by proving (1) implies (2), (2) implies (3), and (3) implies (1).

1. Consider  $AX = 0$ . Since  $A$  is invertible,  $A^{-1}$  exists. Then:

$$\begin{aligned} AX &= 0 \\ A^{-1}AX &= A^{-1}0 \\ (A^{-1}A)X &= A^{-1}0 \\ IX &= 0 \\ X &= 0 \end{aligned}$$

2. Since we can obtain  $I_n$  from  $A$  by performing a finite sequence of elementary row operations, we can conclude that  $A \sim I_n$
- 3.

$$\begin{aligned} A &\sim I_n \\ E_k \dots E_2 E_1 A &= I_n \\ A &= E_1^{-1} E_2^{-1} \dots E_k^{-1} I_n \end{aligned}$$

Since  $A$  is expressed as a product of elementary matrices,  $A$  must be invertible

□

Since  $A$  is invertible,  $A \sim I_n$ . This means that we can find a finite sequence of elementary row operations  $E_1 E_2 \dots E_k$  such that  $E_k \dots E_2 E_1 A = I_n$  and so  $A^{-1} = E_k \dots E_2 E_1$ .

**Example 1.8.** If  $E_2 E_1 A = I_n$ , then  $A^{-1} = E_2 E_1 I_n$ .

$[A \mid I_n]$  is the super augmented of  $A$  which has  $n$  rows and  $2n$  columns. In addition,  $[A \mid I_n]$  can be transformed into  $[I_n \mid A^{-1}]$  using elementary row operations.

**Example 1.9.** Given that  $\begin{bmatrix} -1 & 0 & 1 \\ -5 & 1 & 3 \\ 7 & -1 & -4 \end{bmatrix}$  is invertible, find its inverse.

$$\begin{aligned} & \begin{bmatrix} -1 & 0 & 1 & 1 & 0 & 0 \\ -5 & 1 & 3 & 0 & 1 & 0 \\ 7 & -1 & -4 & 0 & 0 & 1 \end{bmatrix} \rightarrow \\ & \begin{bmatrix} 1 & 0 & -1 & -1 & 0 & 0 \\ -5 & 1 & 3 & 0 & 1 & 0 \\ 7 & -1 & -4 & 0 & 0 & 1 \end{bmatrix} \rightarrow \\ & \begin{bmatrix} 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & -2 & -5 & 1 & 0 \\ 0 & -1 & 3 & 7 & 0 & 1 \end{bmatrix} \rightarrow \\ & \begin{bmatrix} 1 & 0 & -1 & -1 & 0 & 0 \\ 0 & 1 & -2 & -5 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 & 1 \end{bmatrix} \rightarrow \\ & \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & -1 & 3 & 2 \\ 0 & 0 & 1 & 2 & 1 & 1 \end{bmatrix} \end{aligned}$$

$$A^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 3 & 0 \\ 2 & 1 & 3 \end{bmatrix}$$

**Example 1.10.**

$$\begin{aligned} -x + z &= 0 \\ -5x + y + 3z &= 0 \\ 7x - y - 4z &= 0 \end{aligned}$$

$$\begin{bmatrix} -1 & 0 & 1 \\ -5 & 1 & 3 \\ 7 & -1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By the Equivalence Theorem, since  $A$  is invertible, the only solution of the homogeneous system is the trivial solution. Thus,  $x = y = z = 0$ .

**Example 1.11.**

$$\begin{aligned}x + 2y + z &= 0 \\x + 3y - 2z &= 0 \\-x - y - 4z &= 0\end{aligned}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & -2 \\ -1 & -1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

By the Equivalence Theorem, since  $A$  is not invertible, this homogeneous system has infinitely many solutions, one of which is the trivial solution.

**1.6 Further Results on Systems of Equations and Invertibility**

**Theorem 1.5.** Let  $A$  be an  $n \times n$  invertible matrix. Then the system  $AX = B$  has a unique solution for every  $B$ .

**Theorem 1.6.** Let  $A$  be an  $n \times n$  matrix. If  $AX = B$  has a unique solution for every  $B$ , then  $A$  is invertible.

*Proof.* Since  $AX = B$  is uniquely solvable for every  $B$ , let  $B = 0$ . Then  $AX = 0$  has a unique solution. Since the trivial solution always satisfy the homogeneous system  $AX = 0$ , by Equivalence Theorem,  $A$  is invertible.  $\square$

From these theorems, we can conclude that if  $A$  is an  $n \times n$  matrix,

$$A \text{ is invertible} \iff AX = B \text{ is uniquely solvable for every } B$$

This adds a fourth statement to the Equivalence Theorem.

**Theorem 1.7.** Suppose  $A$  is an  $n \times n$  matrix. If  $B$  is an  $n \times n$  matrix with the property that  $BA = I_n$ , then  $B$  is an inverse of  $A$ .

*Proof.*

$$\begin{aligned}BA &= I_n \\BA(A^{-1}) &= I_n(A^{-1}) \\B(AA^{-1}) &= A^{-1} \\BI_n &= A^{-1} \\B &= A^{-1}\end{aligned}$$

This is not a valid proof since it is not told that  $A$  is invertible. But if we can prove that  $A$  is invertible, then we can use this argument to establish the truth of the theorem. Consider the homogeneous system  $AX = 0$ . Then

$$\begin{aligned} B(AX) &= B(0) \\ (BA)X &= 0 \\ I_n X &= 0 \\ X &= 0 \end{aligned}$$

So statement 2 in the Equivalence Theorem is true. All statements in that theorem are true. In particular, 2 is true. That is,  $A$  is invertible. Then use the previous argument to complete this proof.  $\square$

**Theorem 1.8.** If  $A$  is an  $n \times n$  matrix and if  $B$  is an  $n \times n$  matrix such that  $BA = I_n$ , then  $B = A^{-1}$ . (a "left inverse" of  $A$  is the inverse of  $A$ ).

Corollary: If  $A$  and  $B$  are  $n \times n$  matrices such that  $AB = I_n$ , then  $B = A^{-1}$ . (a "right inverse" of  $A$  is the inverse of  $A$ ).

*Proof.* From Theorem 3, we can say that  $A = B^{-1}$ . Then taking inverse of both sides:  $A^{-1} = (B^{-1})^{-1} = B$ .  $\square$

**Example 1.12.** Determine the conditions on  $b_1, b_2, b_3$  such that the system is invertible.

$$\begin{aligned} x_1 + 2x_2 + x_3 &= b_1 \\ x_1 + 3x_2 - 2x_3 &= b_2 \\ -x_1 - x_2 - 4x_3 &= b_3 \end{aligned}$$

$$\begin{bmatrix} 1 & 2 & 1 & b_1 \\ 1 & 3 & -2 & b_2 \\ -1 & -1 & -4 & b_3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 1 & b_1 \\ 0 & 1 & -3 & b_2 - b_1 \\ 0 & 1 & -3 & b_1 + b_3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 1 & b_1 \\ 0 & 1 & -3 & b_2 - b_1 \\ 0 & 0 & 0 & b_1 + b_3 - (b_2 - b_1) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 2 & 1 & b_1 \\ 0 & 1 & -3 & b_2 - b_1 \\ 0 & 0 & 0 & 2b_1 - b_2 + b_3 \end{bmatrix}$$

For consistency,  $2b_1 - b_2 + b_3 = 0$ .

## 2 Determinants

### 2.1 Determinants by Cofactor Expansion

Let  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ . Then the determinant of  $A$ , written  $\det(A)$ , is the number:  $a_{11}a_{22} - a_{12}a_{21}$ . In other words, if  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$  and  $\det(A) \neq 0$ , then  $A$  is invertible and

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Let  $A$  be an  $n \times n$  matrix and let  $a_{ij}$  be the entry in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $A$ . Then:

1. The minor of  $a_{ij}$ , written  $M_{ij}$ , is the determinant of the  $(n-1) \times (n-1)$  matrix obtained from  $A$  by crossing out its  $i^{\text{th}}$  row and  $j^{\text{th}}$  column.
2. The cofactor of  $a_{ij}$ , written  $C_{ij}$ , is given by:  $C_{ij} = (-1)^{i+j} M_{ij}$ .

**Example 2.1.**

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 3 & -2 & -1 \\ -1 & 2 & 1 \end{bmatrix}$$

$$M_{23} = \det\left(\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}\right) = 4 - (-1) = 5$$

$$M_{31} = \det\left(\begin{bmatrix} 1 & 4 \\ -2 & -1 \end{bmatrix}\right) = -1 - (-8) = -7$$

$$C_{23} = (-1)^{2+3}(5) = -5$$

$$C_{31} = (-1)^{3+1}(-7) = 7$$

If  $A = [a_{11}]$ , then  $\det A = a_{11}$ .

Consider the general  $2 \times 2$  matrix  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ .

Then:

1.  $M_{11} = a_{22}$
2.  $C_{11} = (-1)^{1+1}a_{22} = a_{22}$
3.  $M_{12} = a_{21}$
4.  $C_{12} = (-1)^{1+2}a_{21} = -a_{21}$
5.  $M_{21} = a_{12}$
6.  $C_{21} = (-1)^{2+1}a_{12} = -a_{12}$



7.  $M_{22} = a_{22}$

8.  $C_{22} = (-1)^{2+2}a_{11} = a_{11}$

By definition:

1.  $\det(A) = a_{11}a_{22} - a_{12}a_{21} = a_{11}C_{11} - a_{12}(-C_{12})$

2.  $\det(A) = a_{11}C_{11} + a_{12}C_{12}$

3.  $\det(A) = a_{11}a_{22} - a_{12}a_{21} = a_{22}C_{22} - a_{21}(-C_{21})$

4.  $\det(A) = a_{21}C_{21} + a_{22}C_{22}$

5.  $\det(A) = a_{11}a_{22} - a_{21}a_{12} = a_{11}C_{11} - a_{21}(-C_{21})$

6.  $\det(A) = a_{11}C_{11} + a_{21}C_{21}$

7.  $\det(A) = a_{11}a_{22} - a_{21}a_{12} = a_{22}C_{22} - a_{12}(-C_{12})$

8.  $\det(A) = a_{12}C_{12} + a_{22}C_{22}$

Note: Regardless of the choice of row or column in  $2 \times 2$  matrix  $A$ , the number obtained by multiplying its entries by the corresponding cofactors and adding the resulting products is the same. This result, which has been verified for  $2 \times 2$  matrices, remains validated for  $n \times n$  matrices in general.

Let  $A$  be an  $n \times n$  matrix. Then we can define  $\det(A)$  in any of the following ways:

1. Cofactor expansion along  $i^{\text{th}}$  row of  $A$ :  $\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \sum_{j=1}^n a_{ij}C_{ij}$  where  $1 \leq i \leq n$ .
2. Cofactor expansion along  $j^{\text{th}}$  column of  $A$ :  $\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{mj}C_{mj} = \sum_{i=1}^m a_{ij}C_{ij}$  where  $1 \leq j \leq m$ .

**Example 2.2.** Compute  $\det(A)$  if  $A = \begin{bmatrix} 2 & 1 & 4 \\ 3 & -2 & -1 \\ -1 & 2 & 1 \end{bmatrix}$  using cofactor expansion along

1. the first row of  $A$ :  $\det(A) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = 2C_{11} + 1C_{12} + 4C_{13} = 2(0) + 1(-2) + 4(4) = 14$
2. the third column of  $A$ :  $\det(A) = a_{13}C_{13} + a_{23}C_{23} + a_{33}C_{33} = 4C_{13} + (-1)C_{23} + 1C_{33} = 4(4) + (-1)(-5) + 1(-1) = 14$

**Example 2.3.** Find  $\det(A)$  if  $A = \begin{bmatrix} 2 & 3 & 0 & 1 \\ -1 & 0 & 4 & 2 \\ 5 & 0 & -1 & 0 \\ 4 & 0 & 2 & 0 \end{bmatrix}$ .

$$\det(A) = 3C_{12} + 0C_{22} + 0C_{32} + 0C_{42}$$

$$= -3(\det(\begin{bmatrix} -1 & 4 & -2 \\ 5 & -1 & 0 \\ 4 & 2 & 0 \end{bmatrix}))$$

$$= -3(-2 \det(\begin{bmatrix} 5 & -1 \\ 4 & 2 \end{bmatrix}))$$

$$= (-3)(-2)(14) = 84$$

**Theorem 2.1.** Let  $A$  be an  $n \times n$  triangular matrix. Then,

$$\det(A) = a_{11}a_{22} \dots a_{nn} = \prod_{i=1}^n a_{ii}$$

Basket-Weave Method (only for  $3 \times 3$  matrices):

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$$

## 2.2 Evaluating Determinants using Row Reduction

**Theorem 2.2.** Let  $A$  be an  $n \times n$  matrix that contains at least one row or one column where its entries are all zeros. Then,  $\det(A) = 0$ .

*Proof.* Use cofactor expansion of  $\det(A)$  along the zero row or zero column.  $\square$

Recall that if  $A$  is an  $n \times n$  matrix, then the transpose of  $A$ , written  $A^T$ , is the  $n \times n$  matrix obtained from  $A$  by interchanging its rows with its columns.

Properties of the Transpose

1.  $(A^T)^T = A$
2.  $(A + B)^T = A^T + B^T$
3.  $(cA)^T = cA^T$
4.  $(AB)^T = B^T A^T$

Suppose  $A$  is an  $n \times n$  invertible matrix. Let  $B = A^{-1}$ . Then  $(AA^{-1})^T = (A^{-1})^T A^T = I_N^T = I_N$ . So,  $A^T$  is invertible and  $(A^T)^{-1} = (A^{-1})^T$ .

A square matrix  $A$  is said to be symmetric if and only if it has the property that  $A^T = A$ .

Properties of Symmetric Matrices:

1. If  $A$  is symmetric, so is  $A^T$
2. If  $A$  and  $B$  are symmetric matrices, then  $A + B$  is symmetric
3. If  $A$  is symmetric, then  $cA$  is symmetric

Let  $A$  be an  $m \times n$  matrix and consider  $AA^T$ . Is it true that  $AA^T$  is always symmetric? Yes.

$$(AA^T)^T = AA^T$$

Conclusion:  $AA^T$  must be symmetric.

Suppose  $A$  is an  $n \times n$  matrix. Is it true that  $A + A^T$  is symmetric? Is it true that  $(A + A^T)^T = A + A^T$ ?

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

Conclusion: Yes.

Suppose  $A$  is an  $n \times n$  matrix.

1. If we interchange a pair of rows of  $A$  to obtain  $n \times n$  matrix  $B_1$ , then  $\det(B_1) = -\det(A)$
2. If we multiply a row of  $A$  by a nonzero scalar  $c$  to obtain  $n \times n$  matrix  $B_2$ , then  $\det(B_2) = c \det(A)$
3. If we add a multiple of one row of  $A$  to another row of  $A$  to obtain  $n \times n$  matrix  $B_3$ , then  $\det(B_3) = \det(A)$

Suppose  $E$  is an elementary matrix. ( $\det(I_N) = 1$ ). If  $E$  is type 1, then  $\det(E) = -1$ . If  $E$  is type 2, then  $\det(E) = c$ . If  $E$  is type 3,  $\det(E) = 1$ .

**Example 2.4.** Evaluate:  $\det\left(\begin{bmatrix} 1 & -3 & 4 & 2 \\ 2 & -4 & 0 & 1 \\ -1 & 2 & 3 & 3 \\ 3 & 1 & -1 & 2 \end{bmatrix}\right)$ .

$$\begin{aligned}
 \det\left(\begin{bmatrix} 1 & -3 & 4 & 2 \\ 2 & -4 & 0 & 1 \\ -1 & 2 & 3 & 3 \\ 3 & 1 & -1 & 2 \end{bmatrix}\right) &= \det\left(\begin{bmatrix} 1 & -3 & 4 & 2 \\ 0 & 2 & -8 & -3 \\ 0 & -1 & 7 & 5 \\ 0 & 10 & -13 & -4 \end{bmatrix}\right) \\
 &= -\det\left(\begin{bmatrix} 1 & -3 & 4 & 2 \\ 0 & -1 & 7 & 5 \\ 0 & 2 & -8 & -3 \\ 0 & 10 & -13 & -4 \end{bmatrix}\right) \\
 &= -\det\left(\begin{bmatrix} 1 & -3 & 4 & 2 \\ 0 & -1 & 7 & 5 \\ 0 & 0 & 6 & 7 \\ 0 & 0 & 57 & 46 \end{bmatrix}\right) \\
 &= -\det\left(\begin{bmatrix} 1 & -3 & 4 & 2 \\ 0 & -1 & 7 & 5 \\ 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & \frac{41}{2} \end{bmatrix}\right) \\
 &= -(1)(-1)(6)\left(-\frac{41}{2}\right) \\
 &= -123
 \end{aligned}$$

**Example 2.5.** Without using the basket weave method, evaluate

$$\begin{aligned}
 &\det\left(\begin{bmatrix} 2 & 0 & -6 \\ 5 & -3 & -15 \\ 1 & 4 & 4 \end{bmatrix}\right) \\
 \det\left(\begin{bmatrix} 2 & 0 & -6 \\ 5 & -3 & -15 \\ 1 & 4 & 4 \end{bmatrix}\right) &= \det\left(\begin{bmatrix} 2 & 0 & 0 \\ 5 & -3 & 0 \\ 1 & 4 & 4 \end{bmatrix}\right) = (2)(-3)(4) = -24
 \end{aligned}$$

Because  $\det(A^T) = \det(A)$ , we can perform "elementary column operations" to compute  $\det(A)$ . These have the same effect as the corresponding elementary row operations have on the determinants.

**Example 2.6.** Evaluate  $\det\left(\begin{bmatrix} 3 & 1 & -4 & 2 & 5 \\ 5 & 7 & 1 & -3 & 6 \\ 2 & 6 & 7 & 8 & -4 \\ 2 & 6 & 7 & 8 & -4 \\ -6 & -2 & 8 & -4 & -10 \\ 1 & 9 & 5 & -12 & -3 \end{bmatrix}\right)$

Adding -2 times the 1st row to the 4th row produces a row of zeros in the result and have no effect on the determinant. Thus  $\det(A) = 0$ .

## 2.3 Properties of Determinants, Cramer's Rule

**Theorem 2.3.** Let  $A$  be an  $n \times n$  matrix and let  $c$  be a scalar. Then  $\det(cA) = c^n \det(A)$ .

*Proof.* Since  $cA$  is obtained from  $A$  by multiplying each of the  $n$  rows by  $c$  and since each such multiplication introduces a factor of  $c$  in the determinant, we get  $\det(cA) = c^n \det(A)$ .  $\square$

Suppose  $A$  and  $B$  are  $n \times n$  matrices. Is it necessarily true that  $\det(A + B) = \det(A) + \det(B)$ ? No.

Suppose  $A$  and  $B$  are  $n \times n$  matrices. Is it necessarily true that  $\det(AB) = \det(A)\det(B)$ ? Yes.

**Theorem 2.4.** If  $A$  and  $B$  are  $n \times n$  matrices, then  $\det(AB) = \det(A)\det(B)$

*Proof.* We first consider a special case of this result in which  $A = E$ , an elementary matrix. Suppose  $E = E_1$ . Then  $\det(E_1B) = -\det(B) = \det(E_1)\det(B)$ . Suppose  $E = E_2$ . Then  $\det(E_2B) = c\det(B) = \det(E_2)\det(B)$ . Suppose  $E = E_3$ . Then  $\det(E_3B) = 1\det(B) = \det(E_3)\det(B)$ . This completes the proof of the special case.

Claim: Let  $A$  be an  $n \times n$  matrix. Then  $A$  is invertible if and only if  $\det(A) \neq 0$ .

Proof of Claim: Let  $R$  be the reduced row echelon form of  $A$ . This means that there elementary matrices  $E_1, E_2, \dots, E_k$  such that  $E_k \dots E_2 E_1 A = R$ . So then  $\det(E_k \dots E_2 E_1 A) = \det(R)$ . Then  $\det(E_k) \dots \det(E_2) \det(E_1) \det(A) = \det(R)$ . So either  $\det(A)$  and  $\det(R)$  are both not equal to zero or both equal to zero. Suppose  $A$  is invertible. By the Equivalence Theorem,  $A \sim I_N$ . In other words, if  $R$  is the reduced row echelon form of  $A$ , then  $R = I_N$ . Since  $\det(R) = 1 \neq 0$ , it follows that  $\det(A) \neq 0$ .

Conversely, if  $\det(A) \neq 0$ ,  $\det(R) \neq 0$  where  $R$  is the reduced row echelon form of  $A$ . This implies that  $R$  has no zero rows in it which means that  $R = I_N$ . By Equivalence Theorem,  $A$  must be invertible. This adds a fifth statement to the Equivalence Theorem:  $A$  is invertible  $\iff \det(A) \neq 0$ . Now we are in the position to prove the theorem in the general case.

Part 1: Suppose  $A$  is not invertible.

Claim:  $AB$  is not invertible.

Suppose  $AB$  was invertible. Then we would be able to write  $AB$  as a product of elementary matrices;  $AB = E_1 E_2 E_3 \dots E_k$ . This implies that matrix  $A$ ,  $n \times n$ , can be expressed as a product of elementary matrices as well, which would mean that  $A$  would be invertible. Thus,  $AB$  is not invertible. Then  $\det(AB) = 0$ ,  $\det(BA) = 0$ . So,  $\det(AB) = \det(A)\det(B)$ .

Part 2: Suppose  $A$  is invertible. Then  $A = E_k \dots E_2 E_1$  where  $E_k, \dots, E_2, E_1$  are elementary matrices. So then  $AB = E_k \dots E_2 E_1 B$ . Then

$$\begin{aligned} \det(AB) &= \det(E_k \dots E_2 E_1 B) \\ &= \det(E_k) \dots \det(E_2) \det(E_1) \det(B) \\ &= \det(E_k \dots E_2 E_1) \det(B) \\ &= \det(A) \det(B) \end{aligned}$$

$\square$

**Theorem 2.5.** Let  $A$  be an  $n \times n$  matrix. If we multiply the entries in one row of  $A$  by the corresponding cofactors in a different row of  $A$  and we add the resulting products, the sum is always zero.

*Proof.* Define  $n \times n$  matrix  $A'$  so that the entries of  $A'$  are identical to those of  $A$  with the following exception: we replace the  $i^{\text{th}}$  row of  $A$  by a second copy of the  $j^{\text{th}}$  of  $A$ . Then  $\det(A') = 0$ , since  $A'$  has 2 identical rows. But we compute  $\det(A)$  by using a cofactor expansion of our choosing. Let's select the cofactor expansion of the  $k^{\text{th}}$  row of  $A'$ .

$$\det(A') = a_{i1}C_{k1} + a_{i2}C_{k2} + \cdots + a_{in}C_{kn} = 0$$

□

The adjoint of  $A$ , written  $\text{adj}(A)$ , is the transpose of the matrix obtained from  $A$  by replacing each of its entries by its cofactor.

**Theorem 2.6.** If  $A$  is an  $n \times n$  matrix,  $A(\text{adj}(A)) = \det(A)I_N$  or  $(\text{adj}(A))A = \det(A)I_N$ .

*Proof.*  $A(\text{adj}(A)) =$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} \det(A) & 0 & \cdots & 0 \\ 0 & \det(A) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \det(A) \end{bmatrix} \\ = \det(A)I_N$$

□

**Example 2.7.** Let  $A = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & -2 \\ -1 & 2 & -3 \end{bmatrix}$ .

$$\text{adj}(A) = \begin{bmatrix} 1 & 8 & 5 \\ 7 & -4 & -5 \\ -5 & 0 & -5 \end{bmatrix}^T = \begin{bmatrix} 1 & 7 & -5 \\ 8 & -4 & 0 \\ 5 & -5 & -5 \end{bmatrix}$$

$$\det(A) = 20$$

$$A(\text{adj}(A)) = \begin{bmatrix} 1 & 3 & -1 \\ 2 & 1 & -2 \\ -1 & 2 & -3 \end{bmatrix} \begin{bmatrix} 1 & 7 & -5 \\ 8 & -4 & 0 \\ 5 & -5 & -5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 20I_3$$

Suppose that  $A$  is invertible. By the Equivalence Theorem,  $\det(A) \neq 0$ . Multiply by  $\frac{1}{\det(A)}$ .

$$\frac{1}{\det(A)}A(\text{adj}(A)) = \frac{1}{\det(A)}\det(A)I_n$$

and so

$$A\left(\frac{1}{\det(A)}\text{adj}(A)\right) = I_N$$

and thus:

$$A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$$

**Example 2.8.** In the previous example,

$$A^{-1} = \begin{bmatrix} \frac{1}{20} & \frac{7}{20} & -\frac{1}{4} \\ \frac{2}{5} & -\frac{1}{5} & 0 \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{bmatrix}$$

This formula can involve an inordinate amount of computations. For a  $4 \times 4$  matrix, we would have to compute the determinants of 16  $3 \times 3$  matrices and 1  $4 \times 4$  matrices. The result is however useful to prove other results, one of which follows.

**Theorem 2.7.** Cramer's Rule: Consider a system of  $n$  linear equations in  $n$  variables where the coefficient matrix of the system has a nonzero determinant.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_m \end{aligned}$$

Then  $AX = B$  looks like:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & b_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

The unique solution of this system is given by:

$$x_1 = \frac{\det(A_1)}{\det(A)}$$

$$x_2 = \frac{\det(A_2)}{\det(A)}$$

$$x_n = \frac{\det(A_n)}{\det(A)}$$

where  $A_j$  is the  $n \times n$  matrix obtained from  $A$  by removing its  $j^{\text{th}}$  column and replacing it with  $B$  ( $1 \leq j \leq n$ ).

**Example 2.9.** Solve:

$$\begin{aligned} 3x_1 + x_2 + 4x_3 &= 5 \\ x_1 + 3x_2 + 2 + 2x_3 &= -1 \\ 2x_1 + x_2 + 3x_3 &= -2 \end{aligned}$$

Then:

$$AX = B = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 3 & 2 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix}$$

$$\det(A) = 2$$

$$x_1 = \frac{\det\begin{bmatrix} 5 & 1 & 4 \\ -1 & 3 & 2 \\ -2 & 1 & 3 \end{bmatrix}}{\det(A)} = \frac{54}{2} = 27$$

$$x_2 = \frac{\det\begin{bmatrix} 3 & 5 & 4 \\ 1 & -1 & 2 \\ 2 & -2 & 3 \end{bmatrix}}{\det(A)} = \frac{8}{2} = 4$$

$$x_3 = \frac{\det\begin{bmatrix} 3 & 1 & 5 \\ 1 & 3 & -1 \\ 2 & 1 & -2 \end{bmatrix}}{\det(A)} = \frac{-40}{2} = -20$$

*Proof.* Since  $\det(A) \neq 0$ , we know that this system has a unique solution given by  $X = A^{-1}B$ . Since  $\det(A) \neq 0$ , the previous result tells us that

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$$

So then

$$\begin{aligned} x &= \frac{1}{\det(A)} \text{adj}(A)B \\ &= \frac{1}{\det(A)} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix} B \end{aligned}$$

Therefore

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$x_j = \frac{1}{\det(A)} (b_1 C_{1j} + b_2 C_{2j} + \cdots + b_n C_{nj})$$

Thus:

$$A_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & b_1 & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & b_2 & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & b & \cdots & a_{nn} \end{bmatrix}$$

$$\det(A_j) = b_1 C_{1j} + b_2 C_{2j} + \cdots + b_n C_{nj}$$

Substitution gives  $x_j = \frac{1}{\det(A)} \det(A_j) = \frac{\det(A_j)}{\det(A)}$ . □

One of the virtues of Cramer's Rule is that it allows us to solve for a specific variable without solving for all of the variables.



### 3 Vector Spaces

#### 3.1 Definitions and Examples

**Definition 3.1.** Let  $\mathbb{R}^n$  be the set of all ordered  $n$ -tuples of real numbers

For  $u = (u_1, u_2, \dots, u_n)$  and  $v = (v_1, v_2, \dots, v_n)$ :

1. Addition:  $u + v = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n)$
2. Scalar Multiplication:  $ku = (ku_1, ku_2, \dots, ku_n)$

Properties:

1. Closure Property of Addition:  $u, v \in \mathbb{R}^n \rightarrow u + v \in \mathbb{R}^n$
2. Commutative Property of Addition:  $u + v = v + u$
3. Associative Property of Addition:  $u + (v + w) = (u + v) + w$
4. Existence of an Additive Identity Element in  $\mathbb{R}^n$ : There is an element in  $\mathbb{R}^n$ , namely  $(0, 0, \dots, 0)$  with the property that  $u + (0, 0, \dots, 0) = u$  for every  $u \in \mathbb{R}^n$ . This element is called the additive identity of  $\mathbb{R}^n$  and is represented by 0.  
 $u + 0 = u$  (and  $0 + u = u$ ) for all  $u \in \mathbb{R}^n$
5. Existence of an Additive Inverse Element in  $\mathbb{R}^n$ : For every  $u \in \mathbb{R}^n$ , there is an element  $-u \in \mathbb{R}^n$  with the property that  $u + (-u) = 0$  (and  $(-u) + u = 0$ )
6. Closure Property of Scalar Multiplication:  $ku \in \mathbb{R}^n$  if  $u \in \mathbb{R}^n$  and  $k = \text{scalar}$
7. Distributive Property:  $k(uv) = ku + kv$
8. Distributive Property:  $(k + l)u = ku + lu$
9. Associative Property of Scalar Multiplication:  $k(lu) = (kl)u$
10. Existence of a Scalar Multiplication Identity Element:  $1u = u$  for all  $u \in \mathbb{R}^n$

**Definition 3.2.** A vector space  $V$  is a set of objects, together with two operations called scalar addition and scalar multiplication defined on this set such that the ten properties (axioms) listed above are all satisfied ( $\mathbb{R}^n$  is a vector space under the standard operations of addition and scalar multiplication). The objects in a vector space are called vectors.

if we let  $M_{m \times n}$  represent the set of all  $m \times n$  matrices with real identities and we define addition and scalar multiplication in the conventional manner, then the ten axioms are satisfied. Consequently,  $M_{m \times n}$  is a vector space under these operations.

**Definition 3.3.** Define  $P_n$  to be the set of all polynomial functions of degrees less than or equal to  $n$  together with the zero polynomial. Define, in one variable, addition and scalar multiplication in the usual manner.

$$p \in P^n, q \in P^n \rightarrow p + q \in P^n \text{ and } kp \in P^n$$

This is a vector space under the indication operations.

**Example 3.1.** Let  $S$  be the set of all functions that satisfy the differential equation

$$f'' + 5f' + 4f = 0$$

Define addition in the usual manner for the function.

$$\begin{aligned} f \in S &\rightarrow f'' + 5f' + 4f = 0 \\ g \in S &\rightarrow g'' + 5g' + 4g = 0 \\ (f + g)'' + 5(f + g)' + 4(f + g) &= f'' + g'' + 5f' + 5g' + 4f + 4g \\ (f'' + 5f' + 4f) + (g'' + 5g' + 4g) &\rightarrow f + g \in S \end{aligned}$$

**Example 3.2.** Let  $V \in \mathbb{R}^2$ . Define addition  $\oplus$  and scalar multiplication  $\otimes$  as follows:

$$(x_1, y_1) \oplus (x_2, y_2) = (x_1 + y_1 + 1, x_2 + y_2 + 1)$$

**Example 3.3.**  $(2, 3) \oplus (4, 5) = (7, 9)$

$$k \otimes (x_1, y_1) = (kx_1 + k - 1, kx_2 + k - 1)$$

**Example 3.4.**  $5 \otimes (2, 3) = (14, 19)$

Is  $\mathbb{R}^2$  a vector space under these conditions?

$$1. (x_1, y_1) \oplus (a, b) = (x_1, y_1)$$

$$\begin{aligned} (x_1 + a + 1, y_1 + b + 1) &= (x_1, y_1) \\ \rightarrow x_1 + a + 1 &= x_1 \\ y_1 + b + 1 &= y_1 \\ \rightarrow a = -1, b &= -1 \end{aligned}$$

2. Additive Identity Element in  $\mathbb{R}^2$ :

$$(-1, -1) = 0$$

$$3. (x_1, y_1) \oplus (a, b) = (-1, -1)$$

$$\begin{aligned} (x_1 + a + 1, y_1 + b + 1) &= (-1, -1) \\ x_1 + a + 1 = -1 &\rightarrow a = -x_1 - 2 \\ y_1 + b + 1 = -1 &\rightarrow b = -y_1 - 2 \end{aligned}$$

Additive Inverse of  $(x_1, y_1) : (-x_1 - 2, -y_1 - 2) = -(x_1, y_1)$

**Example 3.5.**  $(3, -8) \oplus (-5, 6) = (-1, -1) = 0$

4.  $k \otimes ((x_1, y_1) \oplus (x_2, y_2)) = (k \otimes (x_1, y_1)) \oplus (k \otimes (x_2, y_2))$  LHS:

$$k \otimes (x_1 + x_2 + 1, y_1 + y_2 + 1)$$

$$(k(x_1 + x_2 + 1) + k - 1, k(y_1 + y_2 + 1) + k - 1)$$

$$(kx_1 + kx_2 + 2k - 1, ky_1 + ky_2 + 2k - 1)$$

RHS:

$$(kx_1 + k - 1, ky_1 + k - 1) \oplus (kx_2 + k - 1, ky_2 + k - 1)$$

$$(kx_1 + k - 1 + kx_2 + k - 1 + 1, ky_1 + k - 1 + ky_2 + k - 1 + 1)$$

$$(kx_1 + kx_2 + 2k - 1, ky_1 + ky_2 + 2k - 1)$$

5.  $(k + l) \otimes (x_1, y_1) = (k \otimes (x_1, y_1)) \oplus (l \otimes (x_1, y_1))$

6.  $k \otimes (l \otimes (x_1, y_1)) = (kl) \otimes (x_1, y_1)$

$$k \otimes (lx_1 + l - 1, ly_1 + l - 1)$$

$$(k(lx_1 + l - 1) + k - 1, k(ly_1 + l - 1) + k - 1)$$

$$((kl)x_1 + (kl) - 1, (kl)y_1 + (kl) - 1)$$

$$(kl) \otimes (x_1, y_1)$$

7.  $1 \otimes (x_1, y_1) = (x_1, y_1)$

$$(1x_1 + 1 - 1, 1y_1 + 1 - 1)$$

$$(1x_1, 1y_1) = (x_1, y_1)$$

**Theorem 3.1.** Let  $V$  be a vector space with operations defined. Then

1.  $0u = 0$  for all  $u \in V$
2.  $k0 = 0$  for all scalars  $k$
3.  $(-1)u = -u$  for all  $u \in V$
4. If  $ku = 0$ , then either  $k = 0$ ,  $u = 0$ , or both

*Proof.* 1.

$$0 + 0 = 0 \rightarrow (0 + 0)u = 0u$$

$$\rightarrow 0u + 0u = 0u$$

$$\rightarrow (-0u) + (0u + 0u) = (-0u) + (0u)$$

$$\rightarrow ((-0u) + 0u) + 0u = (-0u) + 0u$$

$$\rightarrow 0 + 0u = 0$$

$$\rightarrow 0u = 0$$

2.

$$\begin{aligned}
 0 + 0 &= 0 \\
 k(0 + 0) &= k0 \\
 &\text{etc...}
 \end{aligned}$$

3.

$$\begin{aligned}
 1 + (-1) &= 0 \\
 \rightarrow (1 + (-1))u &= 0u \\
 \rightarrow 1u + (-1)u &= 0u \\
 \rightarrow u + (-u) &= 0 \\
 \rightarrow -u + (u + (-1)u) &= -u + 0 \\
 \rightarrow (-u + u) + (-1)u &= -u \\
 \rightarrow 0 + (-1)u &= -u \\
 \rightarrow (-1)u &= -1
 \end{aligned}$$

4.

$$ku = 0$$

- (a) If  $k = 0$ , we're done. In what follows, assume  $k \neq 0$ .  
 (b) If  $u = 0$ ,

$$\begin{aligned}
 \frac{1}{k}(ku) &= \frac{1}{k}(0) \\
 \rightarrow \left(\frac{1}{k}k\right)u &= 0 \\
 \rightarrow 1u &= 0 \\
 \rightarrow u &= 0
 \end{aligned}$$

□

## 3.2 Subspaces

**Example 3.6.** Suppose  $V \in \mathbb{R}^3$ . Assume operations defined. Define  $W_1 = \{(x, y, z) | y = 2x, z = 3x\}$ . Then:  $(1, 2, 3) \in W_1$ ,  $(1, 7, 2) \notin W_1$ . Define  $W_2 = \{(x, y, z) | y = 2x, z = 3\}$ . Then:  $(1, 2, 3) \in W_1$ ,  $(1, 2, 4) \notin W_2$ . Is either of these positive nonempty subsets of  $V$  a vector space with respect to the operators defined in  $V$ ?

$W_1$ :

- Choose  $(x, y, z)$  and  $(x', y', z') \in W_1$  Then:

$$z = 3x$$

$$y = 2x$$

$$z' = 3x'$$

$$y' = 2x'$$

Thus:

$$(x, y, z) + (x', y', z') \in W_1$$

Therefore:

$$(x + x', y + y', z + z') \in W_1$$

From this,

$$y + y' = 2x + 2x' = 2(x + x')$$

and

$$z + z' = 3x + 3x' = 3(x + x')$$

Conclusion:  $W_1$  is closed under addition.

•

$$(x, y, z) \rightarrow (-x, -y, -z)$$

Then

$$-z = 3(-x)$$

and

$$-y = 2(-x)$$

All 10 of the axioms are satisfied by the elements of  $W_1$ . Thus,  $W_1$  is itself a vector space under the operations defined on  $V$ .

$W_2$ :

- Axioms 1, 4, 5, 6 are all violated.
- Other 6 axioms continue to hold.

Conclusion:  $W_2$  is not a vector space under the operations defined on  $V$ .

**Definition 3.4.** Let  $V$  be a vector space. A nonempty subset  $W_1$  of  $V$  is called a subspace of  $V$  if and only if  $W_1$  is itself a vector space under the operations defined on  $V$ .

Let  $V$  be a vector space:

- The zero space,  $\{0\}$  is a subspace of  $V$
- $V$  is a subspace of itself
- Every vector space has at least 2 subspaces (only true if  $V$  is not the zero space)

**Theorem 3.2.** Let  $V$  be a vector space and let  $W$  be a nonempty subset of  $V$ . If  $W$  is closed under addition and scalar multiplication (that is, if axioms 1 and 6 are true), then  $W$  is a subspace of  $V$ .

*Proof.* It remains to show that axioms 4 and 5 must hold. For axiom 4, choose a vector  $w \in W$ . Then  $0w \in W$  (from axiom 6). Since  $0w = 0$ ,  $0 \in W$ . For axiom 5, choose a vector  $w \in W$ . Then  $(-1)w \in W$  (from axiom 6). But  $(-1)w = -w$ . Therefore if  $w \in W$ ,  $-w \in W$ .  $\square$

**Example 3.7.** In each of the following, determine if the given subset of a vector space is a subspace.

- $V = M_{n \times n}$  where  $W = \{A \in V | A \text{ is symmetric}\}$ 
  - Let  $A, B \in W$ . Then  $A + B \in W$ . This means there is closure under addition.
  - $A \in W$ ,  $k$  scalar. Then  $kA \in W$ . Also,  $A \in W$ . Then since  $A^T = A$ ,  $(kA)^T = kA^T = kA$ . This means there is closure under scalar multiplication.

Therefore  $W$  is a subspace of  $V$ .

- $V = M_{n \times n}$  where  $W = \{A \in V | A \text{ is invertible}\}$ 
  - Let  $A, B \in W$ . Then  $A, B$  is invertible. But  $A + B$  is not necessarily invertible. Thus,  $W$  is not closed under addition.

Therefore  $W$  is not subspace of  $V$ .

- $V = P_n$  where  $W = \{f \in P_n | f(a) = f(b) \text{ for specified values of } a, b\}$ 
  - Let  $f, g \in W$ . Then  $(f + g) \in W$  and so  $f \in W$  and  $f(a) = f(b)$ . Also  $g \in W$  and so  $g(a) = g(b)$ .
  - Thus  $(f + g)(a) = f(a) + g(a) = f(b) + g(b) = (f + g)(b)$  and thus  $(f + g) \in W$

Thus  $W$  is a subspace of  $V$ .

- Let  $f \in W$ ,  $k = \text{scalar}$ .
  - $f \in W$  and so  $f(a) = f(b)$
  - $(kf)a = k(f(a)) = k(f(b)) = (kf)(b)$
  - Then  $kf \in W$ .

Thus  $W$  is a subspace of  $V$ .

- $V = \mathbb{R}^n$ . If  $A$  is a given  $m \times n$  matrix,  $W = \{X \in \mathbb{R}^n | AX = 0\}$ .
  - Let  $X, X' \in W$ . Then  $X \in W$  and so  $AX = 0$ .
  - Let  $X' \in W$ . Then  $AX' = 0$ .
  - Then  $A(X + X') = AX + AX' = 0 + 0 = 0$
  - Therefore  $X + X' \in W$ .
  - $X \in W$ ,  $k = \text{scalar}$ . Then  $X \in W$  and so  $AX = 0$
  - Then  $A(kX) = k(AX) = k0 = 0$ .
  - Thus  $kX \in W$ .

Thus  $W$  is a subspace of  $V$ .

**Definition 3.5.** Let  $V$  be a vector space and let  $v_1, v_2, \dots, v_r$  be  $r$  distinct vectors chosen from  $V$ . A vector  $v \in V$  is called a linear combination of  $v_1, v_2, \dots, v_r$  if and only if we can find a set of scalars  $k_1, k_2, \dots, k_r$  such that

$$u = k_1 v_1 + k_2 v_2 + \dots + k_r v_r$$

**Definition 3.6.** If it is true that every vector can be expressed as a linear combination of the vectors  $v_1, v_2, \dots, v_r$ , we say that these vectors span  $V$ .

**Example 3.8.** • Show that  $(1, 0)$  and  $(0, 1)$  span  $\mathbb{R}^2$ .

$$(a, b) = a(1, 0) + b(0, 1)$$

- Is it true that  $(1, 2)$  and  $(2, 1)$  span  $\mathbb{R}^2$ ?

$$(a, b) = k_1(1, 2) + k_2(2, 1) = (k_1 + 2k_2, 2k_1 + k_2)$$

$$\begin{aligned} k_1 + 2k_2 &= a \\ 2k_1 + k_2 &= b \end{aligned} \rightarrow \begin{bmatrix} 1 & 2 & a \\ 2 & 1 & b \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & \frac{2b-a}{3} \\ 0 & 1 & \frac{2a-b}{3} \end{bmatrix}$$

$$(a, b) = \left(\frac{2b-a}{3}\right)(1, 2) + \left(\frac{2a-b}{3}\right)(2, 1)$$

- Show that  $(1, 0), (0, 1), (1, 1)$  span  $\mathbb{R}^2$ .

$$(a, b) = (a - c)(1, 0) + (b - c)(0, 1) + c(1, 1) = a(1, 0) + b(0, 1) + 0(1, 1)$$

- Show that  $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$  span  $M_{2 \times 2}$ .

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

- Show that  $1, x, x^2, x^3$  span  $P_3$ .

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 = a_0 1 + a_1 x + a_2 x^2 + a_3 x^3$$

Note: The last 2 examples can be generalized to  $M_{n \times n}$  and  $P_n$  respectively.

**Theorem 3.3.** Let  $V$  be a vector space containing  $r$  distinct vectors  $v_1, v_2, \dots, v_r$ . Define  $W$  to be the subset of  $V$  consisting of all linear combinations of  $v_1, v_2, \dots, v_r$ . Then  $W$  is a subspace of  $V$ .

*Proof.* To prove this, we show that  $W$  is closed under addition and scalar multiplication.

- Choose  $w, w' \in W$ . Then  $w \in W$  and so  $w = k_1 v_1 + k_2 v_2 + \dots + k_r v_r$ . Also,  $w' = k'_1 v_1 + k'_2 v_2 + \dots + k'_r v_r$ . Thus  $w + w' = (k_1 + k'_1)v_1 + (k_2 + k'_2)v_2 + \dots + (k_r + k'_r)v_r = l_1 v_1 + l_2 v_2 + \dots + l_r v_r$ .

- Choose  $w \in W$  and  $k = \text{scalar}$ . From the first part of the argument,  $w = k_1v_1 + k_2v_2 + \cdots + k_rv_r$ . Then  $kw = (kk_1)v_1 + (kk_2)v_2 + \cdots + (kk_r)v_r = m_1v_1 + m_2v_2 + \cdots + m_rv_r \in W$ .

Conclude that  $W$  is a subspace of  $V$ .  $\square$

Notation: We represent  $W$  by  $\text{span}\{v_1, v_2, \dots, v_r\}$  and call it the space spanned by  $v_1, v_2, \dots, v_r$ .

**Example 3.9.** Determine whether or not  $(1, 3, 6)$  is the space spanned by  $(1, -3, 2)$  and  $(2, -4, 8)$ . Equivalent question: Can we express  $(1, 3, 6)$  as a linear combination of  $(1, -3, 2)$  and  $(2, -4, 8)$ ?

$$(1, -3, 6) = k_1(1, -3, 2) + k_2(2, -4, 8) = (k_1 + 2k_2, -3k_1 + 4k_2, 2k_1 + 8k_2)$$

$$\begin{aligned} k_1 + 2k_2 &= 1 \\ -3k_1 + 4k_2 &= 3 \\ 2k_1 + 8k_2 &= 6 \end{aligned} \rightarrow \begin{bmatrix} 1 & 2 & 1 \\ -3 & -4 & 3 \\ 2 & 8 & 6 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 6 \\ 0 & 0 & -2 \end{bmatrix}$$

Inconsistency! So we cannot define scalars  $k_1, k_2$  to meet the required condition.

Conclusion:  $(1, 3, 6)$  is not in the space spanned by  $(1, -3, 2)$  and  $(2, -4, 8)$ .

**Theorem 3.4.** Let  $V$  be a vector space containing the  $r$  distinct vectors  $v_1, v_2, \dots, v_r$ . If  $W$  is a subspace of  $V$  that contains these vectors, then  $\text{span}\{v_1, v_2, \dots, v_r\}$  is a subspace of  $V$ . In other words,  $\text{span}\{v_1, v_2, \dots, v_r\}$  is the smallest subspace of  $V$  containing the vectors  $v_1, v_2, \dots, v_r$ .

*Proof.* We need to show that  $\text{span}\{v_1, v_2, \dots, v_r\}$  is a subset (hence a subspace) of  $W$ . Choose  $u \in \text{span}\{v_1, v_2, \dots, v_r\}$ . We need to show that  $u \in W$ .

$$u = k_1v_1 + k_2v_2 + \cdots + k_rv_r$$

Since  $v_i \in W$  and  $W$  is closed under scalar multiplication,  $k_iv_i \in W$  ( $1 \leq i \leq r$ ). Then since  $W$  is closed under addition,  $\sum_{i=1}^r k_iv_i = u \in W$ . This completes the proof.  $\square$

**Theorem 3.5.** Let  $V$  be a vector space and let  $S = \{v_1, v_2, \dots, v_r\}$  and  $S' = \{w_1, w_2, \dots, w_k\}$  be 2 subsets of vectors chosen from  $V$ . Then

$$\text{span}\{v_1, v_2, \dots, v_r\} = \text{span}\{w_1, w_2, \dots, w_k\}$$

if and only if every  $v_i$  is expressible as a linear combination of the  $w$ 's and every  $w_j$  is expressible as a linear combination of the  $v$ 's.

*Proof.* Left to Right:

$$v_i \in \text{span}\{v_1, v_2, \dots, v_r\} \text{ where } 1 \leq i \leq r$$

By hypothesis,  $v_i \in \text{span}\{w_1, w_2, \dots, w_k\}$ . By definition,  $v_i$  is expressible as a linear combination of  $w_1, w_2, \dots, w_k$ , where  $1 \leq j \leq k$ . By an analogous argument, each  $w_j$  is expressible as a linear combination of  $v_1, v_2, \dots, v_r$ .

Right to Left: Suppose that every  $v_i$  can be expressed as a linear combination of  $w_1, w_2, \dots, w_k$ . This means that spaces spanned by  $w_1, w_2, \dots, w_k$  contain the vectors  $v_1, v_2, \dots, v_r$ . From



the theorem,  $\text{span} \{v_1, v_2, \dots, v_r\} \subseteq \text{span} \{w_1, w_2, \dots, w_k\}$ . Using the other part of the hypothesis and the same type of argument, conclude that  $\text{span} \{w_1, w_2, \dots, w_k\} \subseteq \text{span} \{v_1, v_2, \dots, v_r\}$ .

Thus,

$$\text{span} \{v_1, v_2, \dots, v_r\} = \text{span} \{w_1, w_2, \dots, w_k\}$$

□

### 3.3 Linear Independence

**Definition 3.7.** Let  $V$  be a vector space. Let  $v_1, v_2, \dots, v_r$  be  $r$  distinct vectors chosen from  $V$ . We say that  $S = \{v_1, v_2, \dots, v_r\}$  is a linearly independent set if and only if the only linear combination of  $v_1, v_2, \dots, v_r$  that yields the zero vector is the one where each of its scalars equal zero. Accordingly,  $\{(1, 0), (0, 1)\}$  is a linearly independent set,  $\{(1, 0), (0, 1), (1, 1)\}$  is not. If a set is not linearly independent, we say it is linearly dependent.

Independence Condition: If

$$k_1 v_1 + k_2 v_2 + \dots + k_r v_r = 0$$

then

$$k_1 = k_2 = \dots = k_r = 0$$

**Example 3.10.** In each of the following, determine if the given set is linearly independent or linearly dependent.

- Choose  $v \in V (v \neq 0)$ ,  $S = \{v\}$ . Consider  $k_1 v = 0$ . Since  $v \neq 0$ ,  $k = 0$ . Thus  $S$  is linearly independent. This tells us that any nonzero singleton must be linearly independent.
- Let  $S = \{v_1, v_2, \dots, v_r\}$  where one of the  $v$ 's is the zero vector. To be definite, assume  $v_1 = 0$ . Then  $S = \{0, v_2, v_3, \dots, v_r\}$ . Then

$$k_1 0 + k_2 v_2 + k_3 v_3 + \dots + k_r v_r = 0$$

If we choose  $k_1 = 1$ ,  $k_2 = k_3 = \dots = k_r = 0$ , we can satisfy the requirement. Thus  $S$  is linearly dependent. This tells us that any set of vectors that contain the zero vector is linearly dependent.

- $V = \mathbb{R}^4$ ,  $S = \{(1, 3, -1, 2), (2, 5, 4, 1), (-1, 1, 3, 2)\}$ . Then

$$k_1(1, 3, -1, 2) + k_2(2, 5, 4, 1) + k_3(-1, 1, 3, 2)$$

or

$$\begin{aligned} k_1 + 2k_2 - k_3 &= 0 \\ 3k_1 + 5k_2 + k_3 &= 0 \\ -k_1 + 4k_2 + 3k_3 &= 0 \\ 2k_1 + k_2 + 2k_3 &= 0 \end{aligned}$$

$$\begin{aligned}
& \begin{bmatrix} 1 & 2 & -1 \\ 3 & 5 & 1 \\ -1 & 4 & 3 \\ 2 & 1 & 2 \end{bmatrix} \rightarrow \\
& \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 0 & 6 & 2 \\ 0 & -3 & 4 \end{bmatrix} \rightarrow \\
& \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 0 & 0 & 26 \\ 0 & 0 & -8 \end{bmatrix} \rightarrow \\
& \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \\ 0 & 0 & -8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

Thus  $k_1 = k_2 = k_3 = 0$  and so,  $S$  is linearly independent.

- $V = \mathbb{R}^3$ ,  $S = \{(1, 0, 0), (0, 2, 0), (0, 0, 3), (6, 6, 6)\}$ . Then

$$(6, 6, 6) = k_1(1, 0, 0) + k_2(0, 2, 0) + k_3(0, 0, 3)$$

$$(6, 6, 6) = 6(1, 0, 0) + 3(0, 2, 0) + 2(0, 0, 3)$$

$$(0, 0, 0) = 6(1, 0, 0) + 3(0, 2, 0) + 2(0, 0, 3) - 1(6, 6, 6)$$

Thus  $S$  is linearly dependent.

Given  $S = \{v_1, v_2, \dots, v_r\}$ , a subset of  $r$  distinct vectors chosen from  $\mathbb{R}^n$  from the matrix whose  $j^{\text{th}}$  column in  $v_j$ ,  $1 \leq j \leq r$ . Use Gaussian elimination to obtain the row echelon form of this matrix and consider the path of the leading 1's. If they form a strict diagonal that goes from one end of the matrix to the other,  $S$  is linearly independent. If not,  $S$  is linearly dependent.

**Theorem 3.6.** Let  $S = \{v_1, v_2, \dots, v_r\}$  be a set of  $r$  distinct vectors chosen from vector space  $V$ . Then  $S$  is linearly dependent if and only if we can find some vector in  $S$  that can be expressed as a linear combination of the other vectors in  $S$ .

*Proof.* (Forward) Since  $S$  is linearly dependent, we can find a set of scalars,  $k_1, k_2, \dots, k_r$ , at least one of which is not 0, such that

$$k_1v_1 + k_2v_2 + \dots + k_rv_r = 0$$

To be definite, assume  $k_1 \neq 0$ . Then

$$k_1v_1 = -k_2v_2 - k_3v_3 - \dots - k_rv_r$$

Dividing by  $k_1$ ,

$$v_1 = \left(\frac{-k_2}{k_1}\right)v_2 + \left(\frac{-k_3}{k_1}\right)v_3 + \cdots + \left(\frac{-k_r}{k_1}\right)v_r$$

$$v_1 = c_2v_2 + c_3v_3 + \cdots + c_rv_r$$

(Backward) To be definite, assume  $v_1$  can be expressed as a linear combination of  $v_2, v_3, \dots, v_r$ . That is, assume we can find scalars  $k_2, k_3, \dots, k_r$  such that

$$v_1 = k_2v_2 + k_3v_3 + \cdots + k_rv_r$$

Then

$$0 = (-1)v_1 + k_2v_2 + k_3v_3 + \cdots + k_rv_r$$

Since  $-1 \neq 0$ ,  $S$  is linearly dependent.  $\square$

**Theorem 3.7.** Let  $V = \mathbb{R}^n$  and let  $S = \{v_1, v_2, \dots, v_r\}$  be a subset of  $r$  distinct vectors chosen from  $V$ . If  $r > n$ ,  $S$  is linearly dependent.

Let  $C_{x-1}$  be the set of all functions of  $x$  that have  $x-1$  continuous derivatives. Under the conventional operators defined for functions, it is easy to verify that this is a vector space. Choose  $f_1, f_2, \dots, f_n$  to be  $n$  distinct functions in  $C_{x-1}$  and let  $S = \{f_1(x), f_2(x), \dots, f_n(x)\}$ . Suppose  $S$  is linearly dependent. Then we can find a set of scalars  $k_1, k_2, \dots, k_n$ , at least one of which is not 0, such that

$$k_1f_1f(x) + k_2f_2(x) + \cdots + k_nf_n(x) = 0$$

Then,

$$k_1f'_1f(x) + k_2f'_2(x) + \cdots + k_nf'_n(x) = 0$$

$$k_1f''_1f(x) + k_2f''_2(x) + \cdots + k_nf''_n(x) = 0$$

all the way down to

$$k_1f_1^{n-1}f(x) + k_2f_2^{n-1}(x) + \cdots + k_nf_n^{n-1}(x) = 0$$

In matrix form, this becomes

$$\begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{n-1}(x) & f_2^{n-1}(x) & \cdots & f_n^{n-1}(x) \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Since at least one of the  $k$ 's is not 0, this homogeneous system has nontrivial solutions. By the Equivalence Theorem, the determinant of this matrix is 0. This determinant is called the Wronskian of  $f_1, f_2, \dots, f_n$  and is denoted by  $W(x)$ . So if  $S = \{f_1, f_2, \dots, f_n\}$  is a linearly dependent set, then  $W(x) = 0$  for all  $x$ . Contrapositively, if  $W(x) \neq 0$  for some  $x$ , then  $S$  is linearly independent.

**Example 3.11.** Show that  $S = \{\cos x, \sin x, \sin^2 x\}$  is a linearly independent set.

$$\begin{aligned} W(x) &= \det \begin{pmatrix} \cos x & \sin x & \sin^2 x \\ -\sin x & \cos x & 2 \sin x \cos x \\ -\cos x & -\sin x & 2(\cos^2 x - \sin^2 x) \end{pmatrix} \\ &= \det \begin{pmatrix} \cos x & \sin x & \sin^2 x \\ -\sin x & \cos x & 2 \sin x \cos x \\ 0 & 0 & 2 \cos^2 x - \sin^2 x \end{pmatrix} \\ &= 2 \cos^2 x - \sin^2 x (\cos^2 x + \sin^2 x) = 2 \cos^2 x - \sin^2 x \end{aligned}$$

If  $x = 0$ ,  $W(0) = 2 \neq 0$ . Thus  $S$  is linearly independent.

The converse of our original result is false. That is, if  $W(x) = 0$ , for all  $x$ , we cannot conclude that the set in question will be linearly dependent.

**Example 3.12.** Let  $S = \{1, \cos 2x, \cos^2 x\}$ . Determine if this set is linearly independent or linearly dependent.

$$\begin{aligned} W(x) &= \det \begin{pmatrix} 1 & \cos 2x & \cos^2 x \\ 0 & -2 \sin 2x & -\sin 2x \\ 0 & -4 \cos 2x & -2 \cos 2x \end{pmatrix} \\ &= 4 \sin 2x (\cos^2 x - \sin^2 x) - 4 \sin 2x \cos 2x \\ &= 4 \sin 2x \cos 2x - 4 \sin 2x \cos 2x = 0 \end{aligned}$$

$S$  is linearly dependent.

### 3.4 Basis and Dimension

**Definition 3.8.** Let  $V$  be a vector space. A subset  $S = \{v_1, v_2, \dots, v_r\}$  of  $V$  is called a basis for  $V$  if and only if both of the following are true:

- $S$  spans  $V$
- $S$  is linearly independent

**Example 3.13.** Let  $V = \mathbb{R}^3$  and let  $S = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ . Is  $S$  a basis for  $\mathbb{R}^3$ ?

$$(a, b, c) = a(1, 0, 0) + b(0, 1, 0) + c(0, 0, 1)$$

This shows that  $S$  spans  $\mathbb{R}^3$ . Now consider

$$k_1(1, 0, 0) + k_2(0, 1, 0) + k_3(0, 0, 1) = (0, 0, 0)$$

$$(k_1, k_2, k_3) = (0, 0, 0)$$

$k_1 = k_2 = k_3 = 0$ . This shows that  $S$  is linearly independent. Thus,  $S$  is a basis for  $\mathbb{R}^3$  (standard basis for  $\mathbb{R}^3$ ).

In general, if  $V = \mathbb{R}^n$ , then  $S = \{e_1, e_2, \dots, e_n\}$  is a standard basis for  $\mathbb{R}^n$  where  $e_i = (0, 0, \dots, 1, \dots, 0)$ ,  $1 \leq i \leq n$ , and the  $i^{\text{th}}$  position is a 1.

**Example 3.14.** Let  $V = \mathbb{R}^3$  and  $S = \{v_1, v_2, v_3\}$  where  $v_1 = (1, 1, 1)$ ,  $v_2 = (2, 2, 0)$ , and  $v_3 = (3, 0, 0)$ . Is  $S$  a basis for  $\mathbb{R}^3$ ?

Consider the matrix  $A$  by placing  $v_j$  in the  $j^{\text{th}}$  column,  $1 \leq j \leq 3$ .

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$\det(A) = -6 \neq 0$ . By the Equivalence Theorem,  $AX = 0$  has only the trivial solution. Therefore,  $S$  is linearly independent. Also by the Equivalence Theorem,  $AX = B$  is solvable for every  $B$  and so,  $S$  spans  $\mathbb{R}^3$ . Thus,  $S$  is a basis for  $\mathbb{R}^3$ .

In summary, if we have  $n$  vectors chosen from  $\mathbb{R}^n$  and we compute the determinant of the matrix whose  $j^{\text{th}}$  column is  $v_j$ , then if the determinant is not equal to 0, the set of vectors is a basis for  $\mathbb{R}^n$ .

**Example 3.15.** Let  $V = M_{22}$  and let  $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ .

$$k_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + k_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + k_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + k_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} k_1 & k_2 \\ k_3 & k_4 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow k_1 = k_2 = k_3 = k_4 = 0 \rightarrow \text{linearly independence}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow a = b = c = d = 1$$

$S$  spans  $M_{22}$ . Thus  $S$  is a basis for  $M_{22}$  (standard basis).

**Example 3.16.** Let  $V = P_n$  and let  $S = \{1, x, x^2, \dots, x^n\}$ .

$$W(x) = \begin{bmatrix} 1 & x & x^2 & \cdots & x^n \\ 0 & 1 & 2x & \cdots & nx^{n-1} \\ 0 & 0 & 2 & \cdots & n(n-1)x^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n! \end{bmatrix} = \det(A) \neq 0 \text{ for all } x$$

Thus this is linearly independent.

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n = a_01 + a_1x + a_2x^2 + \cdots + a_nx^n$$

Thus  $S$  spans  $P_n$ . Therefore  $S$  is a basis for  $P_n$  (standard basis).

**Example 3.17.** Let  $V = \{0\}$ . The only possible basis is  $S = \{0\}$ . Thus it is linearly dependent since it contains the zero vector. Therefore this is not a basis for  $V$ .

The zero space is the only vector space without a basis.

**Definition 3.9.** A vector space  $V$  is called finite-dimensional if and only if it is either the zero space or it has a basis containing finitely many vectors in  $V$ . Therefore,  $\mathbb{R}^n, M_{mn}, P_n, \{0\}$  are all finite dimensional vector spaces.

Consider the selection of all polynomial functions in the single variable  $x$  (independent of degree). Call this collection  $P$  and define addition and scalar multiplication in the usual manner. Claim:  $P$  is not finite-dimensional. Suppose to the contrary, that  $P$  is finite-dimensional. If so, we can find a basis for  $P$  that contains finitely many vectors in it

$$\{P_1, P_2, \dots, P_k\}$$

Let  $d_j$  be the degree of  $P_j$ ,  $1 \leq j \leq k$ . Let  $D$  be the largest of the  $d_j$ 's. Then consider  $x^{D+1}$ . Since this polynomial cannot be expressed as a linear combination of  $P_1, P_2, \dots, P_k$  (its degree is too high), these vectors do not span  $P$ . Hence it will not be a basis for  $P$ . Thus,  $P$  is not finite-dimensional (such vectors are called infinite-dimensional).

**Theorem 3.8.** Let  $V$  be a nonzero finite-dimensional vector space and let  $S = \{v_1, v_2, \dots, v_m\}$  be a basis for  $V$ . Let  $S' = \{w_1, w_2, \dots, w_n\}$  be a subset of  $n$  vectors chosen from  $V$ . If  $n \geq m$ , then  $S'$  is linearly dependent.

*Proof.* Consider  $k_1 w_1 + k_2 w_2 + \dots + k_n w_n = 0$  and show that we can satisfy this equation where at least one of the  $k$ 's is not zero. Since  $S$  is a basis for  $V$ ,  $S$  spans  $V$ . This means that each vector in  $S'$  can be expressed as a linear combination of the vectors in  $S$ .

$$\begin{aligned} w_1 &= a_{11}v_1 + a_{12}v_2 + \dots + a_{1m}v_m \\ w_2 &= a_{21}v_1 + a_{22}v_2 + \dots + a_{2m}v_m \\ &\vdots \\ w_n &= a_{n1}v_1 + a_{n2}v_2 + \dots + a_{nm}v_m \end{aligned}$$

$$\begin{aligned} k_1(a_{11}v_1 + a_{12}v_2 + \dots + a_{1m}v_m) &+ k_2(a_{21}v_1 + a_{22}v_2 + \dots + a_{2m}v_m) \\ &+ \dots + k_n(a_{n1}v_1 + a_{n2}v_2 + \dots + a_{nm}v_m) = 0 \end{aligned}$$

Rearrangement on the left side yields:

$$\begin{aligned} (a_{11}k_1 + a_{12}k_2 + \dots + a_{1n}k_n)v_1 &+ (a_{21}k_1 + a_{22}k_2 \\ &+ \dots + a_{2n}k_n)v_2 + \dots + (a_{m1}k_1 + a_{m2}k_2 + \dots + a_{mn}k_n)v_m = 0 \end{aligned}$$

Since  $S$  is a basis for  $V$ ,  $S$  is linearly independent. So then

$$\begin{aligned} a_{11}k_1 + a_{12}k_2 + \dots + a_{1n}k_n &= 0 \\ a_{21}k_1 + a_{22}k_2 + \dots + a_{2n}k_n &= 0 \\ &\vdots \\ a_{m1}k_1 + a_{m2}k_2 + \dots + a_{mn}k_n &= 0 \end{aligned}$$

This is a homogeneous linear system with more variables ( $n$ ) than equations ( $m$ ). Such a system has nontrivial solutions. That is, we can find a set of scalars  $k_1, k_2, \dots, k_n$ , at least one of which is not zero, that satisfies this requirement. Conclude:  $S'$  is linearly dependent. Contrapositively, if  $S'$  is linearly independent, then  $n \leq m$ . But in another way, if a basis for  $V$  contains  $m$  vectors in it, any linearly independent subset of vectors chosen from  $V$  cannot contain more than  $m$  vectors in it.  $\square$

**Theorem 3.9.** Let  $V$  be a nonzero finite-dimensional vector space and let  $S = \{v_1, v_2, \dots, v_m\}$  and  $S' = \{w_1, w_2, \dots, w_n\}$  be 2 basis for  $V$ . Then  $m = n$ .

*Proof.* Since  $S'$  is a basis for  $V$ ,  $S'$  is linearly independent. By the contraposition of the theorem,  $n \leq m$ . Since  $S$  is a basis for  $V$ ,  $S$  is linearly independent. Using the same logic,  $m \leq n$ . Then  $m = n$ .  $\square$

Consequently, any basis for  $\mathbb{R}^n$  contains  $n$  vectors. Every basis for  $M_{mn}$  contains  $mn$  vectors. Every basis for  $P_n$  contains  $n + 1$  vectors. The number of vectors in a basis of a nonzero finite-dimensional vector space is invariant.

Let  $V$  be a finite-dimensional vector space. Then

- if  $v = \{0\}$ , the dimension is equal to 0
- if  $v \neq \{0\}$ , the dimension is equal to the number of vectors in any basis for  $V$

Dimensions:

- $\dim(\mathbb{R}^n) = n$
- $\dim(M_{mn}) = mn$
- $\dim(P_n) = n + 1$

**Example 3.18.** Let  $V$  be the vector space consisting of all ordered triples that satisfy the homogeneous system: 
$$\begin{matrix} x_1 + x_2 - x_3 = 0 \\ x_1 + 2x_2 + 3x_3 = 0 \end{matrix}$$
 The solution set of this system has already been shown to be a vector space under ordinary operations. Call it  $V$ . Find  $\dim(V)$ .

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 4 \end{bmatrix}$$

If  $\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = V$ , then

$$\begin{matrix} x_1 - 5x_3 = 0 \\ x_2 + 4x_3 = 0 \end{matrix} \rightarrow x_1 = 5x_3, x_2 = -4x_3, x_3 = x_3 \rightarrow \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 5x_3 \\ -4x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}$$

$\begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}$  spans  $V$ . This set is also linearly independent since it is a nonzero singleton. Thus  $S = \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}$  is a basis for  $V$ . So  $\dim(V) = 1$ .

In the following theorem, we assume  $V$  is a nonzero finite-dimensional vector space with  $\dim(V) = n$ .

**Theorem 3.10.** Let  $S = \{v_1, v_2, \dots, v_n\}$  be a subset of  $n$  distinct vectors chosen from  $V$ . If  $S$  is linearly independent, then  $S$  is a basis for  $V$ .

*Proof.* To establish that  $S$  is a basis for  $V$ , we need to show that  $S$  spans  $V$ . To do so, we must show that every vector in  $V$  is expressible as a linear combination of  $v_1, v_2, \dots, v_n$ . Choose  $v \in V$ . If  $v = v_i$  for some  $i$  between 1 and  $n$ , then the result is trivially true.

$$v = v_i = 0v_1 + 0v_2 + \dots + 1v_i + \dots + 0v_n$$

In what follows, assume  $v \neq v_i$  for some  $1 \leq i \leq n$ . Suppose  $v$  cannot be expressed as a linear combination of  $v_1, v_2, \dots, v_n$ . That means that  $v \notin \text{span}\{v_1, v_2, \dots, v_n\}$ . It follows that  $\{v_1, v_2, \dots, v_n\}$  is linearly independent. This is a contradiction since every subset of  $V$  containing more than  $n$  vectors must be linearly dependent. By contradiction,  $v$  is expressible as a linear combination of  $v_1, v_2, \dots, v_n$ . Thus  $S$  spans  $V$  and so  $S$  is a basis for  $V$ .  $\square$

**Theorem 3.11.** Let  $S = \{v_1, v_2, \dots, v_n\}$  be a subset of  $n$  distinct vectors chosen from  $V$ . If  $S$  spans  $V$ , then  $S$  is a basis for  $V$ .

*Proof.* We need to show that  $S$  is linearly independent. We can assume, without loss of generality, that  $n > 1$ . Assume, to the contrary, that  $S$  is linearly dependent. Then we can find some vector in  $S$  that can be expressed as a linear combination of the others. To be definite, assume  $v_n$  to be so expressible. Then  $\{v_1, v_2, \dots, v_{n-1}\}$  will continue to span  $V$ . This set cannot be linearly independent, for if it were, it would be a basis for  $V$  with fewer than  $n$  vectors thus contradicting  $\dim(V) = n$ . So  $\{v_1, v_2, \dots, v_n\}$  is linearly dependent. Continue until eventually we obtain a linearly dependent set consisting of a single singleton vector. This is impossible. Thus  $S$  is linearly independent and so  $S$  is a basis for  $V$ .  $\square$

Summary: if  $\dim(V) = n$  and if  $S = \{v_1, v_2, \dots, v_n\}$  is a subset of  $V$  containing  $n$  vectors

$$S \text{ is linearly independent} \iff S \text{ spans } V$$

**Theorem 3.12.** Let  $V$  be a nonzero finite-dimensional vector space with  $\dim(V) = n$ . Let  $S = \{v_1, v_2, \dots, v_r\}$  be a linearly independent subset of  $r$  vectors chosen from  $V$ , where  $r < n$ . Then we can always find  $n - r$  vectors in  $V$ , say  $v_{r+1}, v_{r+2}, \dots, v_n$ , such that  $\{v_1, v_2, \dots, v_r, v_{r+1}, v_{r+2}, \dots, v_n\}$  is a basis for  $V$ . (expanding to a basis)

*Proof.*  $\{v_1, v_2, \dots, v_r\}$  does not span  $V$ . for if it does, it would be a basis for  $V$  that contains fewer than  $n$  vectors, which contradicts  $\dim(V) = n$ . This implies that we can find more vectors in  $V$ , say  $v_{r+1}$  that cannot be expressed as a linear combination of  $\{v_1, v_2, \dots, v_r\}$ . That is,  $v_{r+1} \notin \text{span}\{v_1, v_2, \dots, v_r\}$ . Thus  $\{v_1, v_2, \dots, v_r, v_{r+1}\}$  is a linearly independent set. If  $r+1 = n$ , we're done. If  $r+1 < n$ , find  $v_{r+2}$  such that  $v_{r+2} \notin \text{span}\{v_1, v_2, \dots, v_r, v_{r+1}\}$ . So  $\{v_1, v_2, \dots, v_r, v_{r+1}, v_{r+2}\}$  is linearly independent. If  $r+2 = n$ , we're done. If not, continue. After  $n-r$  steps, we obtain a linearly independent set of  $n$  vectors which contains the original set as part of it. This is an expanded basis for  $V$ .  $\square$



**Theorem 3.13.** Let  $\{v_1, v_2, \dots, v_r\}$  be a spanning set of  $r$  vectors chosen from  $V$ , where  $r > n$ . Then we can always find  $r - n$  vectors in this set that can be deleted from it in order to create a basis for  $V$ . (contracting to a basis)

*Proof.*  $\{v_1, v_2, \dots, v_r\}$  is necessarily linearly dependent since if it were linearly independent, it would be a basis for  $V$  that contains more than  $n$  vectors, violating  $\dim(V) = n$ . This means that some vector in this set can be removed and the remaining  $r - 1$  vectors will continue to span  $V$ . To be definite, assume  $v_r$  can be removed. Then  $\{v_1, v_2, \dots, v_{r-1}\}$  will span  $V$ . If  $r - 1 = n$ , we're done. If  $r - 1 > n$ , continue the process. After  $r - n$  steps, we will obtain a subset of  $\{v_1, v_2, \dots, v_n\}$  that contains  $n$  vectors and that spans  $V$ . This must be a basis for  $V$ .  $\square$

**Theorem 3.14.** Let  $W$  be a subspace of  $V$ . Then  $\dim(W) \leq n$  if and only if  $W \equiv V$ .

*Proof.* Suppose  $\dim(W) > n$ . This implies that we can find a basis for  $W$  that contains more than  $n$  vectors in it. This set is linearly independent. These vectors are all in  $V$ , since  $W$  is part of  $V$ . But this is impossible since any linearly independent set of vectors in  $V$  cannot contain more than  $n$  vectors in it. By contradiction,  $\dim(W) \leq n$ . It remains to show that  $\dim(W) = n$  if and only if  $W \equiv V$ .

(Backward) If  $W = V$ , then  $\dim(W) = \dim(V) = n$ .

(Forward) Suppose  $\dim(W) = n$ . Let  $\{w_1, w_2, \dots, w_n\}$  be a basis for  $W$ . To show  $W = V$ , we need to show that  $W$  is a subset of  $V$  and  $V$  is a subset of  $W$ .  $W$  is a subspace of  $V$ . So  $W \subseteq V$ . Choose  $v \in V$ . If we can show  $v \in W$ , then we can conclude that  $V \subseteq W$ . Suppose  $v \notin W$ , then  $v \notin \text{span}\{w_1, w_2, \dots, w_n\}$ . But  $\{w_1, w_2, \dots, w_n\}$  is a linearly independent set. Then  $\{w_1, w_2, \dots, w_n, v\}$  is a linearly independent set, which has  $n + 1$  vectors in it. Since each vector in this set is a vector in  $V$ , this contradicts  $\dim(V) = n$ . Therefore  $v \in W$ . Thus,  $W \subseteq V$  and  $V \subseteq W$  and so  $W = V$ .  $\square$

**Theorem 3.15.** Let  $V$  be a nonzero finite-dimensional vector space with  $\dim(V) = n$  and let  $S = \{v_1, v_2, \dots, v_n\}$  be a basis for  $V$ . Then every vector in  $V$  is uniquely expressible as a linear combination of  $v_1, v_2, \dots, v_n$ .

*Proof.* To show uniqueness, we need to show that there is at least one way and at most one way to express every vector in  $V$  as a linear combination of  $v_1, v_2, \dots, v_n$ .

At least one: Choose  $v \in V$ . Since  $S$  spans  $V$ , we can express  $v$  as a linear combination of  $v_1, v_2, \dots, v_n$ .

At most one: Choose  $v \in V$ . Suppose we write  $v = k_1v_1 + k_2v_2 + \dots + k_nv_n$  and we can write  $v = l_1v_1 + l_2v_2 + \dots + l_nv_n$  where  $k_i \neq l_i$  for at least one  $i$ ,  $1 \leq i \leq n$ . Then

$$k_1v_1 + k_2v_2 + \dots + k_nv_n = l_1v_1 + l_2v_2 + \dots + l_nv_n$$

$$k_1v_1 - l_1v_1 + k_2v_2 - l_2v_2 + \dots + k_nv_n - l_nv_n = 0$$

$$(k_1 - l_1)v_1 + (k_2 - l_2)v_2 + \dots + (k_n - l_n)v_n = 0$$

Since  $S$  is basis for  $V$ ,  $S$  is linearly independent. So  $k_1 - l_1 = 0, k_2 - l_2 = 0, \dots, k_n - l_n = 0$ . Thus  $k_1 = l_1, k_2 = l_2, \dots, k_n = l_n$ . This contradicts  $k_i \neq l_i$  for some  $i$ ,  $1 \leq i \leq n$ . So there is at least one way to express  $v$  as a linear combination of  $v_1, v_2, \dots, v_n$ .  $\square$

Let  $V$  be a nonzero finite-dimensional vector space and let  $S = \{v_1, v_2, \dots, v_n\}$  be a basis for  $V$ . For any  $v \in V$ ,  $v = k_1v_1 + k_2v_2 + \dots + k_nv_n$ , where the  $k$ 's are uniquely determined:

- $k_1, k_2, \dots, k_n$  are called the coordinates of  $V$  with respect to  $S$
- $(V)_S = \{k_1, k_2, \dots, k_n\}$  is called the coordinate vector of  $V$  with respect to  $S$

**Example 3.19.** Let  $S = \{v_1, v_2, v_3\}$  where  $v_1 = (1, 1, 1)$ ,  $v_2 = (2, 2, 0)$ , and  $v_3 = (3, 0, 0)$ .

(a) Show  $S$  is a basis for  $\mathbb{R}^3$ . (b) Find  $(V)_S$ .

(a)

$$\det\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} = -6 \neq 0$$

$S$  is a basis for  $\mathbb{R}^3$ .

(b)

$$\begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 0 & 8 \\ 1 & 0 & 0 & 6 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 6 \\ 1 & 2 & 0 & 8 \\ 1 & 2 & 3 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 1 \\ 0 & 2 & 3 & -4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

$$(V)_S = (6, 1, -2)$$

$$6(1, 1, 1) + 1(2, 2, 0) - 2(3, 0, 0) = (2, 8, 6)$$

**Example 3.20.** Given  $S = \{1, 1+x, 1+x^2\}$ . (a) Show that  $S$  is a basis for  $P_n$ . (b) Find  $(V)_S$  if  $v = 1+x+x^2$ .

(a)

$$W(x) = \begin{bmatrix} 1 & 1+x & 1+x^2 \\ 0 & 1 & 2x \\ 0 & 0 & 2 \end{bmatrix} = 2 \neq 0$$

So  $S$  is linearly independent. Thus  $S$  is a basis for  $P_2$ .

$$(b) \quad 1+x+x^2 = -1(1) + 1(1+x) + 1(1+x^2)$$

$$(V)_S = (-1, 1, 1)$$

### 3.5 Row Space, Column Space and Nullspace of a Matrix

Consider  $m \times n$  matrix  $A$  shown below:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \vdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \vdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

**Definition 3.10.** The row vectors of  $A$ , represented by  $r_1, r_2, r_3, \dots, r_m$ , are the vectors:

$$r_1 = [a_{11} \quad a_{12} \quad a_{13} \quad \dots \quad a_{1n}]$$

$$r_2 = [a_{21} \quad a_{22} \quad a_{23} \quad \dots \quad a_{2n}]$$

$$r_3 = [a_{31} \quad a_{32} \quad a_{33} \quad \dots \quad a_{3n}]$$

$$\vdots$$

$$r_m = \begin{bmatrix} a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

There are  $m$  vectors, each of which is an element of  $\mathbb{R}^n$ .

**Definition 3.11.** The column vectors of  $A$ , represented by  $c_1, c_2, c_3, \dots, c_n$ , are the vectors:

$$c_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{bmatrix}$$

$$c_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{bmatrix}$$

$$c_3 = \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ \vdots \\ a_{m3} \end{bmatrix}$$

$$\vdots$$

$$c_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

There are  $n$  vectors, each of which is an element of  $\mathbb{R}^n$ .

Let  $A$  be a  $m \times n$  matrix:

1. The row space of  $A$ , written  $\text{row}(A)$  is the subspace of  $\mathbb{R}^n$  consisting of all linear combinations of the row vectors of  $A$ . That is,  $\text{row}(A) = \text{span} \{r_1, r_2, r_3, \dots, r_m\}$ .
2. The column space of  $A$ , written  $\text{col}(A)$  is the subspace of  $\mathbb{R}^m$  consisting of all linear combinations of the column vectors of  $A$ . That is,  $\text{col}(A) = \text{span} \{c_1, c_2, c_3, \dots, c_n\}$ .
3. The nullspace of  $A$ , written  $N(A)$ , is the subspace of  $\mathbb{R}^n$  consisting of all solutions of the homogeneous linear system with coefficient matrix  $A$ . That is,  $N(A) = \{x \in \mathbb{R}^n \mid AX = 0\}$

Note: in one of the previous examples,  $\begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}$  is a basis for  $N(A)$ .

When solved a homogeneous linear system, we found a basis for the null space of its coefficient matrix. This calculation was not affected by elementary row operations on the given matrix.

Suppose  $r_1, r_2, r_3, \dots, r_m$  are the row vectors of  $m \times n$  matrix  $A$  so that  $\text{row}(A) = \text{span} \{r_1, r_2, r_3, \dots, r_m\}$ . Perform an elementary row operation on  $A$  to obtain  $m \times n$  matrix  $A'$  where row vectors are  $r'_1, r'_2, r'_3, \dots, r'_m$ . Then each  $r'_i$  is a linear combination of  $r_1, r_2, r_3, \dots, r_m$ . Thus  $r'_1, r'_2, r'_3, \dots, r'_m \in \text{span} \{r_1, r_2, r_3, \dots, r_m\}$  or  $\text{Row}(A') \subseteq \text{Row}(A)$ . Since elementary row operations are reversible,  $\text{Row}(A) \subseteq \text{Row}(A')$ . Thus  $\text{Row}(A) = \text{Row}(A')$ . Consequently, if we perform elementary row operations on  $A$  to obtain a row echelon form of  $A$ , the row space of the result is identical to the row space of  $A$ .

**Example 3.21.** Find a basis for  $\text{Row}(A)$  if  $A = \begin{bmatrix} 1 & 1 & 0 & 2 & 0 & 1 \\ 1 & 2 & 0 & -1 & 4 & 3 \\ 3 & 5 & 0 & 1 & 13 & 5 \\ -2 & -3 & 0 & -1 & -4 & -4 \\ 5 & 9 & 0 & -2 & 16 & 13 \end{bmatrix}$ .

Row Echelon Form of  $A =$

$$\begin{bmatrix} 1 & 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & -3 & 4 & 2 \\ 0 & 0 & 0 & 1 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$k_1(1, 1, 0, 2, 0, 1) + k_2(0, 1, 0, -3, 4, 2) + k_3(0, 0, 0, 1, 5, -2) = (0, 0, 0, 0, 0, 0)$$

Thus  $\{(1, 1, 0, 2, 0, 1), (0, 1, 0, -3, 4, 2), (0, 0, 0, 1, 5, -2)\}$  spans  $\text{Row}(A)$ . The nonzero vectors in row echelon form of  $A$  will span  $\text{Row}(A)$ . They will also be linearly independent. Hence they form a basis for  $\text{Row}(A)$ .

Observe that for any matrix  $A$ ,  $\text{Col}(A) = \text{Row}(A^T)$ .

**Example 3.22.** Find a basis for  $\text{Col}(A)$  if  $A = \begin{bmatrix} 1 & 1 & 0 & 2 & 0 & 1 \\ 1 & 2 & 0 & -1 & 4 & 3 \\ 3 & 5 & 0 & 1 & 13 & 5 \\ -2 & -3 & 0 & -1 & -4 & -4 \\ 5 & 9 & 0 & -2 & 16 & 13 \end{bmatrix}$ .

$$A^T = \begin{bmatrix} 1 & 1 & 3 & -2 & 5 \\ 1 & 2 & 5 & -3 & 9 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & -1 & 1 & -1 & -2 \\ 0 & 4 & 13 & -4 & 16 \\ 1 & 3 & 5 & -4 & 13 \end{bmatrix}$$

So then the row echelon form of  $A^T = \begin{bmatrix} 1 & 1 & 3 & -2 & 5 \\ 0 & 1 & 2 & -1 & 4 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ . Thus  $S = \{(1, 1, 3, -2, 5), (0, 1, 2, -1, 4), (0, 0, 1, 0, 0), (0, 0, 0, 0, 0), (0, 0, 0, 0, 0), (0, 0, 0, 0, 0)\}$ .

is a basis for  $\text{Row}(A^T)$ .

Conclude:  $\left\{ \begin{bmatrix} 1 \\ 1 \\ 3 \\ -2 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\}$  is a basis for  $\text{Col}(A)$ .

**Example 3.23.** Find a basis for  $N(A)$  where  $A = \begin{bmatrix} 1 & 1 & 0 & 2 & 0 & 1 \\ 1 & 2 & 0 & -1 & 4 & 3 \\ 3 & 5 & 0 & 1 & 13 & 5 \\ -2 & -3 & 0 & -1 & -4 & -4 \\ 5 & 9 & 0 & -2 & 16 & 13 \end{bmatrix}$ . Row Ech-

elon form of  $A = \begin{bmatrix} 1 & 0 & 0 & 0 & -29 & 9 \\ 0 & 1 & 0 & 0 & 19 & -4 \\ 0 & 0 & 0 & 1 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ . If  $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} \in N(A)$ , then:

$$\begin{aligned} x_1 - 29x_5 + 9x_6 &= 0 & x_1 &= 29x_5 - 9x_6 \\ 19x_5 - 4x_6 &= 0 & \rightarrow x_2 &= -19x_5 + 4x_6 \\ x_4 + 5x_5 - 2x_6 &= 0 & x_4 &= -5x_5 + 2x_6 \end{aligned}$$

From this, we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 29x_5 & -9x_6 \\ -19x_5 & 2x_6 \\ x_3 \\ -5x_5 & 2x_6 \\ x_5 \\ x_6 \end{bmatrix} = x_3 \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} + x_5 \begin{bmatrix} 29 & -19 & 0 & -5 & 1 & 0 \end{bmatrix} + x_6 \begin{bmatrix} -9 & 4 & 0 & 2 & 0 & 1 \end{bmatrix}$$

Conclusion:  $\left\{ \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 29 & -19 & 0 & -5 & 1 & 0 \end{bmatrix}, \begin{bmatrix} -9 & 4 & 0 & 2 & 0 & 1 \end{bmatrix} \right\}$  is a basis for  $N(A)$ .

**Theorem 3.16.** The linear system  $AX = B$  is consistent if and only if  $B \in \text{Col}(A)$ .

*Proof.* (Forward) Suppose  $AX = B$  is consistent. So there exists scalars  $k_1, k_2, \dots, k_n$  such

that  $A \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = B$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = B \rightarrow$$

$$\begin{bmatrix} a_{11}k_1 + a_{12}k_2 + \dots + a_{1n}k_n \\ a_{21}k_1 + a_{22}k_2 + \dots + a_{2n}k_n \\ \vdots \\ a_{m1}k_1 + a_{m2}k_2 + \dots + a_{mn}k_n \end{bmatrix} = B \rightarrow k_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + k_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + k_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = B$$

This means

$$k_1 c_1 + k_2 c_2 + \dots + k_n c_n = B$$

This means that  $B$  is expressible as a linear combination of the column vectors of  $A$ . Thus  $B \in \text{Col}(A)$ .

(Reverse) Logic is reversible □

### 3.6 Rank and Nullity

In the previous section, we examined the row space, column space and nullspace of a  $5 \times 6$  matrix  $X$ . It turned out that  $\dim(\text{Row}(A)) = 3$ ,  $\dim(\text{Col}(A)) = 3$ , and  $\dim(\text{N}(A)) = 3$ . It turns out to be true that for any matrix  $A$ ,  $\dim(\text{Row}(A)) = \dim(\text{Col}(A))$ . (proof omitted)

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

- the rank of  $A$ , written  $r(A)$ , is the dimension of  $\text{Row}(A)$ , or equivalently, the dimension of  $\text{Col}(A)$

$$r(A) = \dim(\text{Row}(A)) = \dim(\text{Col}(A))$$

- the nullity of  $A$ , written  $n(A)$ , is the dimension of  $\text{N}(A)$

$$n(A) = \dim(\text{N}(A))$$

in the previous example, the rank of  $A$  is 3 and the nullity of  $A$  is also 3.

Facts:

- Let  $A$  be an  $m \times n$  matrix

$$- r(A^T) = r(A)$$

$$r(A) = \dim(\text{Col}(A)) = \dim(\text{Row}(A^T)) = r(A^T)$$

$$- r(A) \leq \min(m, n)$$

$\text{Row}(A)$  is a subspace of  $\mathbb{R}^n$ . So  $\dim(\text{Row}(A)) \leq \dim(\mathbb{R}^n) \rightarrow r(A) \leq n$ .  $\text{Column}(A)$  is a subspace of  $\mathbb{R}^m$ . So  $\dim(\text{Col}(A)) \leq \dim(\mathbb{R}^m) \rightarrow r(A) \leq m$ . So  $r(A) \leq \min(m, n)$ .

- Let  $A$  be an  $n \times n$  matrix
  - $r(A)$  = the number of nonzero rows in the row echelon form of  $A$  = the number of leading 1's in the row echelon form of  $A$  = the number of leading variables in the linear system with coefficient matrix  $A$
  - $n(A)$  = the number of free variables in the linear system with coefficient matrix  $A$

**Theorem 3.17.** Dimension Theorem for Matrices: Let  $A$  be an  $m \times n$  matrix. Then

$$r(A) + n(A) = n$$

**Theorem 3.18.** Equivalence Theorem: Let  $A$  be an  $n \times n$  matrix, then the following statements are equivalent (all are true or all are false):

1.  $A$  is invertible
2.  $AX = 0 \rightarrow X = 0$
3.  $A \sim I_n$
4.  $AX = B$  is uniquely solvable for every  $B \in \mathbb{R}^n$
5.  $\det(A) \neq 0$
6.  $n(A) = 0$
7.  $r(A) = n$
8.  $\{r_1, r_2, \dots, r_n\}$  is a linearly independent set
9.  $\{c_1, c_2, \dots, c_n\}$  is a linearly independent set

Plan: (a) 5 proves 6, (b) 6 proves 7, (c) 7 proves 8, (d) 8 proves 9, (e) 9 proves 5

*Proof.* (a) Since  $AX = 0$  has  $X = 0$  as its only solution, the nullspace of  $A$  consists of the zero vector only. That is,  $n(A) = \{0\}$ . So

$$\dim(n(A)) = \dim(\{0\})$$

$$n(A) = 0$$

This is (6).

(b)  $n(A) = 0$ . By the Dimension Theorem for Matrices,  $r(A) + n(A) = n$ . Substitution then gives  $r(A) = n$ , which is (7).

(c)  $r(A) = n$ . So  $\dim(\text{Row}(A)) = n$ . By definition, the row vectors of  $A$ ,  $r_1, r_2, \dots, r_n$  span  $\text{Row}(A)$  ( $\text{Row}(A) = \text{span}\{r_1, r_2, \dots, r_n\}$ ). Since the row space has dimension  $n$  and there are  $n$  vectors in this spanning set,  $A$  is a basis for  $\text{Row}(A)$ , hence is linearly independent. So says (8).

(d)  $\{r_1, r_2, \dots, r_n\}$  is linearly independent (by hypothesis) and spans  $\text{Row}(A)$  (by definition). So  $\{r_1, r_2, \dots, r_n\}$  is a basis for  $\text{Row}(A)$ , which means  $\dim(\text{Row}(A)) = n$ . Since  $\dim(\text{Row}(A)) = \dim(\text{Col}(A))$ , it follows that  $\dim(\text{Col}(A)) = n$ . But  $\{c_1, c_2, \dots, c_n\}$  spans  $\text{Col}(A)$  (by definition). So  $\{c_1, c_2, \dots, c_n\}$  is a basis for  $\text{Col}(A)$ . Therefore  $\{c_1, c_2, \dots, c_n\}$  is a linearly independent set, which is (9).

(e) Since  $\{c_1, c_2, \dots, c_n\}$  is a linearly independent set, it follows that

$$k_1 c_1 + k_2 c_2 + \dots + k_n c_n = 0 \rightarrow k_1 = k_2 = \dots = k_n = 0$$

That is,

$$k_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + k_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + k_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = 0$$

So

$$\begin{bmatrix} a_{11}k_1 + a_{12}k_2 + \dots + a_{1n}k_n \\ a_{21}k_1 + a_{22}k_2 + \dots + a_{2n}k_n \\ \vdots \\ a_{m1}k_1 + a_{m2}k_2 + \dots + a_{mn}k_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \vdots & a_{mn} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

That is,  $AX = 0$  and so  $X = 0$ . This is (5). □

**Theorem 3.19.** Let  $AX = B$  be a non-homogeneous consistent linear system and let  $X_0$  be a solution of it ( $AX_0 = B$ ). Then  $\bar{X}$  is a solution of this system if and only if

$$\bar{X} = X_0 + k_1 v_1 + k_2 v_2 + \dots + k_n v_n$$

when  $\{v_1, v_2, \dots, v_n\}$  is a basis for  $N(A)$ .

*Proof.* (Forward) Suppose  $\bar{X}$  is a solution to  $AX = B$ , then  $A\bar{X} = B$ . By hypothesis,  $AX_0 = B$ . Substituting then gives:

$$\begin{aligned} A\bar{X} &= AX_0 \\ A\bar{X} - AX_0 &= 0 \\ A(\bar{X} - X_0) &= 0 \end{aligned}$$

That is,  $\bar{X} - X_0$  is in  $N(A)$ . Since  $\{v_1, v_2, \dots, v_n\}$  is a basis for  $N(A)$ , there exists constants  $k_1, k_2, \dots, k_n$  such that

$$\begin{aligned} \bar{X} - X_0 &= k_1 v_1 + k_2 v_2 + \dots + k_n v_n \\ \bar{X} &= X_0 + k_1 v_1 + k_2 v_2 + \dots + k_n v_n \end{aligned}$$



(Reverse) Suppose  $\bar{X} = X_0 + k_1v_1 + k_2v_2 + \cdots + k_nv_n$ . Then

$$\begin{aligned} A\bar{X} &= A(X_0 + k_1v_1 + k_2v_2 + \cdots + k_nv_n) \\ &= AX_0 + Ak_1v_1 + Ak_2v_2 + \cdots + Ak_nv_n \\ &= B + k_10 + k_20 + \cdots + 0k_n \\ &= B + 0 + 0 + \cdots + 0 = B \end{aligned}$$

So  $\bar{X}$  is a solution to  $AX = B$ . □

**Example 3.24.** Find the general solution to

$$\begin{aligned} x_2 + 2x_3 + 3x_4 + x_6 &= -1 \\ 2x_1 + 2x_2 + 4x_3 + 8x_4 + 12x_6 &= 2 \\ -4x_1 + 3x_2 + 6x_3 + 5x_4 + x_5 - 14x_6 &= -10 \end{aligned}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 2 - b - 5c \\ -1 - 2a - 3b - c \\ a \\ b \\ 1 - 3c \\ -c \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + a \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} -1 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + c \begin{bmatrix} -5 \\ -1 \\ 0 \\ 0 \\ -3 \\ -1 \end{bmatrix}$$

## 4 Eigenvalues and Eigenvectors

### 4.1 Eigenvalues and Eigenvectors

**Definition 4.1.** Let  $A$  be an  $n \times n$  matrix. A real constant  $\lambda$  is called an eigenvalue of  $A$  if and only if we can find a nonzero vector  $X \in \mathbb{R}^n$  such that  $AX = \lambda X$ . If such an  $X$  can be found, we call it an eigenvector of  $A$ , corresponding to  $\lambda$ .

Suppose that we wish to find eigenvalues of  $n \times n$  matrix  $A$ . If  $\lambda$  is one such, then  $AX = \lambda X$  for some  $X \neq 0$ .

$$\begin{aligned} AX &= \lambda X \\ \lambda X - AX &= 0 \\ \lambda I_n X - AX &= 0 \\ (\lambda I_n - A)X &= 0 \end{aligned}$$

This represents a homogeneous linear system with coefficient matrix  $\lambda I_n - A$  which we know has a nontrivial solution. By the Equivalence Theorem, it follows that  $\det(\lambda I_n - A) = 0$ . This is called the characteristic equation of  $A$ . The left hand side of this equation is called the characteristic polynomial of  $A$  which we will denote by  $p_A(\lambda)$ .

The real roots of the character equation of  $A$  are the eigenvalues of  $A$ .

**Example 4.1.** If  $A = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix}$ , find the eigenvalues of  $A$ . For each eigenvalues found, determine its corresponding eigenvector of  $A$ .

$$\lambda I_2 - A = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} \lambda - 4 & 1 \\ -2 & \lambda - 1 \end{bmatrix}$$

$$\det(\lambda I_2 - A) = (\lambda - 4)(\lambda - 1) + 2 = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3) = p^*(x)$$

For  $\lambda = 2$ :

$$\lambda I_2 - A = 2I_2 - A = \begin{bmatrix} -2 & 1 \\ -2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} -2 & 1 \\ 0 & 0 \end{bmatrix}$$

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  is an eigenvector of  $A$  corresponding to  $\lambda = 2$ . Then  $\begin{matrix} -2x_1 + x_2 = 0 \\ x_2 = 2x_1 \end{matrix}$ . Thus  $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 2x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ . Therefore one such eigenvector is  $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ .

$$AX = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 2X$$

For  $\lambda = 3$ :

$$\lambda I_2 - A = 3I_2 - A = \begin{bmatrix} -1 & 1 \\ -2 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & -1 \\ 0 & 0 \end{bmatrix}$$

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  is an eigenvector of  $A$  corresponding to  $\lambda = 3$ . Then  $\begin{matrix} -x_1 + x_2 = 0 \\ x_2 = x_1 \end{matrix}$ . Thus  $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ . Therefore one such eigenvector is  $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ .

**Example 4.2.** Find eigenvalues of  $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ .

$$\det(\lambda I_2 - A) = 0$$

$$\det\left(\begin{bmatrix} \lambda & -1 \\ 1 & \lambda \end{bmatrix}\right) = 0$$

$$\lambda^2 + 1 = 0$$

$$\lambda^2 = -1$$

No real solutions.

This example shows that not every square matrix will have real eigenvalues.

**Definition 4.2.** The set of all eigenvectors of an  $n \times n$  matrix  $A$  corresponding to eigenvalue  $\lambda$ , together with the zero vector, will form a vector space. This vector space is called the eigenspace of  $A$  corresponding to  $\lambda$  and represented by  $E_\lambda$ .

$$E_\lambda = N(\lambda I_n - A)$$

**Example 4.3.** In the previous example,  $\left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$  is a basis for  $E_{\lambda=2}$ . Similarly,  $\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$  is a basis for  $E_{\lambda=3}$ . Consequently,  $\dim(E_{\lambda=2}) = 1$  and  $\dim(E_{\lambda=3}) = 1$ .

**Example 4.4.** If  $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & -2 \\ -1 & 2 & 2 \end{bmatrix}$ , find the eigenvalues of  $A$ . Then find a basis for each eigenspace of  $A$ .

$$\begin{aligned} \det(\lambda I_3 - A) &= 0 \\ \det\left(\begin{bmatrix} \lambda - 1 & 0 & 0 \\ -2 & \lambda + 3 & 2 \\ 1 & -2 & \lambda - 2 \end{bmatrix}\right) &= 0 \\ (\lambda - 1)[(\lambda + 3)(\lambda - 2) + 4] &= 0 \\ (\lambda - 1)(\lambda^2 + \lambda - 2) &= 0 \\ (\lambda - 1)(\lambda - 1)(\lambda + 2) &= 0 \\ (\lambda - 1)^2(\lambda + 2) &= 0 \end{aligned}$$

There are 2 distinct such eigenvalues,  $\lambda = 1$  (multiplicity of 2) and  $\lambda = -2$ .

For  $\lambda = 1$ :

$$1I_3 - A = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 4 & 2 \\ 1 & -2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus  $E_{\lambda=1} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ .  $x_1 - 2x_2 - x_3 = 0$   
 $x_1 = 2x_2 + x_3$  Therefore

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_2 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} x_3 \\ 0 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

This shows that  $\left\{\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$  spans  $E_{\lambda=1}$ . Since this set has 2 elements in it, and there are

not multiples of one another, the set is also linearly independent. Thus  $\left\{\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$  is a

basis for  $E_{\lambda=1}$ . So  $\dim(E_{\lambda=1}) = 2$ .

For  $\lambda = -2$ :

$$-2I_3 - A = \begin{bmatrix} -3 & 0 & 0 \\ -2 & 1 & 2 \\ 1 & -2 & -4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & -2 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus  $E_{\lambda=-2} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ .  $x_1 = 0$  Therefore  
 $x_2 + 2x_3 = 0$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}$$

This shows that  $\left\{ \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \right\}$  spans  $E_{\lambda=-2}$ . So  $\dim(E_{\lambda=-2}) = 1$ .

Suppose  $A$  is an  $n \times n$  triangular matrix.

$$A = \begin{bmatrix} a_{11} & & & & \\ a_{21} & a_{22} & & & \\ a_{31} & a_{32} & a_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ a_{n1} & \dots & \dots & \dots & a_{nn} \end{bmatrix}$$

Then

$$\det(\lambda I_n - A) = 0$$

$$\det \left( \begin{bmatrix} \lambda - a_{11} & & & & \\ \lambda - a_{21} & \lambda - a_{22} & & & \\ \lambda - a_{31} & \lambda - a_{32} & \lambda - a_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ \lambda - a_{n1} & \dots & \dots & \dots & \lambda - a_{nn} \end{bmatrix} \right) = 0$$

$$(\lambda - a_{11})(\lambda - a_{22})(\lambda - a_{33}) \dots (\lambda - a_{nn}) = 0$$

$$\lambda = a_{11}, a_{22}, a_{33}, \dots, a_{nn}$$

## Miscellaneous Results Concerning Eigenvalues and Eigenvectors

1.

**Theorem 4.1.** Let  $A$  be an  $n \times n$  matrix. Then  $\lambda = 0$  is an eigenvalue of  $A$  if and only if  $\det(A) = 0$ .

*Proof.* (Forward) Since  $\lambda = 0$  is an eigenvalue of  $A$ , then it must be true that  $\det(\lambda I - A) = 0$ . This implies that  $\det(-A) = 0$ . Since  $\det(-A) = (-1)^n \det(A)$ , we get  $(-1)^n \det(A) = 0$ . So  $\det(A) = 0$ .

(Reverse) If  $\det(-A) = 0$ , then  $\det(0I_n - A) = 0$ . Then  $\lambda = 0$  is an eigenvalue of  $A$ .  $\square$

The contrapositive of this result tells us that  $\det(A) \neq 0$  if and only if  $\lambda = 0$  is not an eigenvalue of  $A$ . The right hand side of this becomes a tenth statement to the Equivalence Theorem.

2. If we know the eigenvalues of an  $n \times n$  matrix  $A$ , what can we predict about the eigenvalues of "relatives" of  $A$ ?

(a) If  $\lambda$  is an eigenvalue of  $A$ , then  $\lambda$  is an eigenvalue of  $A^T$ . (Put another way, the eigenvalue of  $A$  and  $A^T$  are identical.)

*Proof.* Suppose  $\lambda$  is an eigenvalue of  $A$ . This means that  $\det(\lambda I_n - A) = 0$ . Since  $\det(B) = \det(B^T)$  for any square matrix  $B$ , it follows that  $\det((\lambda I_n - A)^T) = 0$ . So  $\det((\lambda I_n)^T - A^T) = 0$ . Then  $\det(\lambda I_n - A^T) = 0$ . By definition,  $\lambda$  is an eigenvalue of  $A^T$ . Thus the eigenvalue of  $A$  and  $A^T$  are identical. But it is not true that their corresponding spaces are identical as well.  $\square$

(b) If  $\lambda$  is an eigenvalue of  $A$  and  $s$  is any constant, then  $\lambda - s$  is an eigenvalue of  $A - sI$ .

*Proof.* Since  $\lambda$  is an eigenvalue of  $A$ , we have  $AX = \lambda X$  for some  $X \neq 0$ . Then

$$(A - sI_n)X = AX - (sI_n)X$$

$$AX = sX$$

$$AX - sX = \lambda X - sX = (\lambda - s)X$$

 $\square$ 

Note: Eigenspaces are not altered.

(c) If  $A$  is an invertible  $n \times n$  matrix and if  $\lambda$  is an eigenvalue of  $A$ , then  $\frac{1}{\lambda}$  is an eigenvalue of  $A^{-1}$ .

*Proof.*

$$\begin{aligned} AX &= \lambda X \\ A^{-1}(AX) &= A^{-1}(\lambda X) \\ X &= \lambda(A^{-1}X) \end{aligned}$$

Since  $A$  is invertible,  $\lambda \neq 0$ . So

$$\begin{aligned} \frac{1}{\lambda}X &= A^{-1}X \\ A^{-1}X &= \frac{1}{\lambda}X \end{aligned}$$

Thus by definition,  $\frac{1}{\lambda}$  is an eigenvalue for  $A^{-1}$  (same  $X$ ). □

- (d) If  $\lambda$  is an eigenvalue of  $A$ , then  $\lambda^k$  is an eigenvalue of  $A^k$  for every positive integer  $k$ .

*Proof.* Since  $\lambda$  is an eigenvalue of  $A$ , there exists  $X \neq 0$  such that  $AX = \lambda X$ .

$$\begin{aligned} AX &= \lambda X \\ A(AX) &= A(\lambda X) \\ A^2X &= \lambda(AX) \\ &= \lambda(\lambda X) \\ &= \lambda^2X \end{aligned}$$

So  $\lambda^2$  is an eigenvalue of  $A^2$ . Continue the process, or more formally, use mathematical induction to establish the result stated. □

- (e) If  $\lambda$  is an eigenvalue of  $A$  and if  $p(y) = c_my^m + c_{m-1}y^{m-1} + \cdots + c_2y^2 + c_1y + c_0$  is a polynomial function, then  $p(\lambda)$  is an eigenvalue of  $p(A)$ .

*Proof.* Since  $\lambda$  is an eigenvalue of  $A$ , we know  $AX = \lambda X$  for some  $X \neq 0$ . From the previous part, we know that  $A^kX = \lambda^kX$ .

$$\begin{aligned} p(A) &= c_mA^m + c_{m-1}A^{m-1} + \cdots + c_2A^2 + c_1A + c_0I_n \\ (p(A))X &= (c_mA^m + c_{m-1}A^{m-1} + \cdots + c_2A^2 + c_1A + c_0I_n)X \\ &= c_mA^mX + c_{m-1}A^{m-1}X + \cdots + c_2A^2X + c_1AX + c_0I_nX \\ &= c_m\lambda^mX + c_{m-1}\lambda^{m-1}X + \cdots + c_2\lambda^2X + c_1\lambda X + c_0X \\ &= (c_m\lambda^m + c_{m-1}\lambda^{m-1} + \cdots + c_2\lambda^2 + c_1\lambda + c_0)X \\ &= p(\lambda)X \end{aligned}$$

□

3.

**Theorem 4.2.** Cayley-Hamilton's Theorem: If  $A$  is an  $n \times n$  matrix and  $p^*(x)$  is a characteristic polynomial, then  $p^*(A) = 0$ . (Put another way, every square matrix satisfies its own characteristic equation).

The proof of this theorem is a straight forward algebraic exercise of  $n = 2$ . (Worse but still doable for  $n = 3$ .) The general proof is more complicated and will be omitted. In general, if  $A$  is an  $n \times n$  matrix, we can use this theorem to compute  $A^k$  as a linear combination of  $A^{k-1}, A^{k-2}, \dots, A^2, A, I_n$ .

**Example 4.5.** Let  $A = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix}$ . Find the eigenvalues of  $A^2$ . The eigenvalues of  $A$  is 2, 3.

$$A^2 = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 14 & -5 \\ 10 & -1 \end{bmatrix}$$

$$\det\left(\begin{bmatrix} \lambda - 14 & 5 \\ -10 & \lambda + 1 \end{bmatrix}\right) = 0 \rightarrow (\lambda - 14)(\lambda + 1) + 50 = 0$$

$$\lambda^2 - 13\lambda + 36 = 0$$

$$(\lambda - 4)(\lambda - 9) = 0$$

Thus the eigenvalues of  $A^2$  are 4 and 9, which is  $2^2$  and  $3^2$  respectively.

**Example 4.6.** Find the eigenvalues of  $p(A)$  if  $p(y) = 3y^2 + 5y + 10$  using the matrix from the previous example.

$$p(A) = 3A^2 + 5A + 10$$

$$p(A) = 3 \begin{bmatrix} 14 & -5 \\ 10 & -1 \end{bmatrix} + 5 \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} + 10 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 72 & -20 \\ 40 & 12 \end{bmatrix}$$

Since  $\lambda = 2, 3$  for  $A$ ,  $p(2) = 32$  and  $p(3) = 52$  are both eigenvalues for  $p(A)$ .

**Example 4.7.** Verify Cayley-Hamilton's Theorem for the previous matrix.

$$p^*(x) = (\lambda - 2)(\lambda - 3) = x^2 - 5x + 6$$

$$p^*(A) = A^2 - 5A + 6I_n = \begin{bmatrix} 14 & -5 \\ 10 & -1 \end{bmatrix} - 5 \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} + 6 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = 0$$

**Example 4.8.** If  $B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & -2 \\ 1 & 2 & 2 \end{bmatrix}$ , find  $B^7$ .

We have already found that

$$p^*(\lambda) = (\lambda - 1)^2(\lambda + 2) = \lambda^3 - 3\lambda + 2$$

$$B^3 - 3B + 2I_3 = 0$$

$$B^3 = 3B - 2I_3$$

$$B^6 = (B^3)^2$$

$$= (3B - 2I_3)^2$$

$$= 9B^2 - 12B + 4I_3$$

$$B^7 = BB^6$$

$$= 9B^3 - 12B^2 + 4B$$

$$= 9(3B - 2I_3) - 12B^2 + 4B$$

$$= -12B^2 + 31B - 18I_3$$



## 4.2 Diagonalizability

Let  $A = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix}$ . Its eigenvalues are  $\lambda = 2, 3$ . The bases are: for  $E_{\lambda=2} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$  and for  $E_{\lambda=3} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ . Let  $P = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$  where  $P$  invertible ( $\det(P) = -1$ ). Then  $P^{-1} = \begin{bmatrix} 1 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$ . Now:

$$P^{-1}AP = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 6 & -3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = D$$

We obtained a diagonal matrix which displays the eigenvalues of  $A$  along its main diagonal in the same order that the basis vectors for the eigenspaces were inserted as the columns of  $P$ . We say in this case that  $A$  is diagonalizable and that  $P$  is an invertible matrix that diagonalizes  $A$ .

$$P^{-1}AP = D \rightarrow P(P^{-1}AP)P^{-1} = PDP^{-1} \rightarrow A = PDP^{-1}$$

$$A^k = \underbrace{AAA \dots A}_{k \text{ copies}} \rightarrow A^k = (PDP^{-1})(PDP^{-1})(PDP^{-1}) \dots (PDP^{-1}) = PD^kP^{-1}$$

**Definition 4.3.** Let  $A$  be an  $n \times n$  matrix. We say that  $A$  is diagonalizable if and only if we can find an invertible matrix  $P$  such that  $PDP^{-1} = D$  where  $D$  is a diagonal matrix. We then say that  $P$  diagonalizes  $A$  and that  $A$  is similar to  $D$ .

How can we determine if a given  $n \times n$  matrix is diagonalizable?

In order for  $n \times n$  matrix to have a chance of being diagonalizable, it must possess  $n$  real eigenvalues, counting multiplicatives. Once this hurdle is cleared, find a basis for each eigenspace of the matrix. Suppose  $\lambda_1, \lambda_2, \dots, \lambda_k$  are the distinct eigenvalues of  $A$  and that  $m_1, m_2, \dots, m_k$  are their respective multiplicatives.

Note:  $\sum_{i=1}^k m_i = n$

Let  $d_i = \dim(E_{\lambda_i})$  for  $1 \leq i \leq k$ . Then  $A$  is diagonalizable. We can define an invertible  $n \times n$  matrix  $P$  by placing the eigenvectors in the basis for the eigenspaces as columns. Then  $D = P^{-1}AP$  will be the diagonal matrix that displays the eigenvalues of  $A$  along its main diagonal where the order of presentation corresponds to the choice of column vectors placed into  $P$ .

If  $d_i < m_i$  for some  $i$ , then  $A$  is not diagonalizable. If we can ascertain that  $A$  is diagonalizable, then we can compute  $A^k$  without using Cayley-Hamilton's Theorem.

**Example 4.9.** Let  $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & -2 \\ -1 & 2 & 2 \end{bmatrix}$ . We already shown that  $\lambda = 1$  (multiplicity of 2)

and  $\lambda = -2$  as the eigenvalues of  $A$ . Also we found  $\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$  to be the basis for  $E_{\lambda=1}$

$(\dim(E_{\lambda=1}) = 2)$  and  $\left\{ \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \right\}$  to be the basis for  $E_{\lambda=-2}$  ( $\dim(E_{\lambda=-2}) = 1$ ). Thus  $A$  is diagonalizable.

$$P = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix} \text{ diagonalizes } A$$

$$P^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{4}{3} \\ \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \end{bmatrix}$$

$$P^{-1}AP = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{4}{3} \\ \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & -2 \\ -1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = D$$

$$A^{10} = PD^{10}P^{-1} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -2 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1024 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{4}{3} \\ \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3} \end{bmatrix}$$

**Example 4.10.** Let  $B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & -2 \\ 1 & 2 & 2 \end{bmatrix}$  Is  $B$  diagonalizable?

The matrix  $B$  has 3 real eigenvalues,  $\lambda = 1$  (multiplicity of 2) and  $\lambda = -2$ . But a base for  $E_{\lambda=1}$  was  $\left\{ \begin{bmatrix} 0 \\ -\frac{1}{2} \\ 1 \end{bmatrix} \right\}$  and that's it. So  $\dim(E_{\lambda=1}) = 1$ . But  $1 < 2$ . So  $B$  is not diagonalizable.

Under a special circumstance, we have a guarantee that  $A$  is diagonalizable. Let  $A$  be a  $n \times n$  matrix with  $n$  distinct eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Then  $m_1 = m_2 = \dots = m_n = 1$ . Since each eigenspace of  $A$  has dimension  $\geq 1$ , and since each dimension is less than or equal to the corresponding multiplicity, we get  $d_i \geq 1$  and  $d_i \leq m_i = 1$ . So  $d_1 = d_2 = \dots = d_n = 1$ . Since  $d_i = m_i$ , where  $1 \leq i \leq n$ ,  $A$  is diagonalizable.

Caution: If  $A$  has  $n$  real eigenvalues that are not distinct, we may still be able to diagonalize  $A$ .

**Example 4.11.** A certain product is made by two competing companies, X and Y, that completely controls the market. Each year, X retains  $\frac{1}{3}$  of its customers while the other  $\frac{2}{3}$  shifts to Y. Each year, Y retains  $\frac{1}{2}$  of its customers while the other  $\frac{1}{2}$  shifts to X. Predict the distribution of the market in the long run.

Construct a  $2 \times 2$  matrix  $A$  as follows:

- $a_{11}$  = the probability that a customer of X remains a customer of X a year later
- $a_{21}$  = the probability that a customer of X becomes a customer of Y a year later
- $a_{12}$  = the probability that a customer of Y becomes a customer of X a year later
- $a_{22}$  = the probability that a customer of Y remains a customer of Y a year later

$$A = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{2} \end{bmatrix}$$

Observations:

- all entries in this square matrix are numbers between 0 and 1 inclusive
- the sum of the entries in each column of  $A$  is equal to 1

Such a matrix is called stochastic. Analysis of a stochastic matrix is called a Markov process. These arise when percentages and/or probabilities need to be predicted.

Let  $\begin{bmatrix} a \\ b \end{bmatrix}$  represent the actual distribution of the market. In this example,  $a = \frac{3}{4}$  and  $b = \frac{1}{4}$ .

$$A \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{1}{3}a + \frac{1}{2}b \\ \frac{2}{3}a + \frac{1}{2}b \end{bmatrix}$$

This indicates the distribution of the market after one year.

$$A \begin{bmatrix} \frac{1}{3}a + \frac{1}{2}b \\ \frac{2}{3}a + \frac{1}{2}b \end{bmatrix} = A(A \begin{bmatrix} a \\ b \end{bmatrix}) = A^2 \begin{bmatrix} a \\ b \end{bmatrix}$$

This gives the distribution of the market after 2 years. This suggests that  $A^k \begin{bmatrix} a \\ b \end{bmatrix}$  will give us the distribution of the market after  $k$  years.

$$\det(\lambda I_2 - A) = 0 \rightarrow \det\left(\begin{bmatrix} \lambda - \frac{1}{3} & -\frac{1}{2} \\ -\frac{2}{3} & \lambda - \frac{1}{2} \end{bmatrix}\right)$$

$$\begin{aligned} (\lambda - \frac{1}{3})(\lambda - \frac{1}{2}) - \frac{1}{3} &= 0 \\ (6\lambda + 1)(\lambda - 1) &= 0 \\ \lambda &= -\frac{1}{6}, 1 \end{aligned}$$

For  $E_{\lambda=-\frac{1}{6}}$ :

$$\begin{aligned} \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ -\frac{2}{3} & -\frac{2}{3} \end{bmatrix} &\rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in E_{\lambda=-\frac{1}{6}} &\rightarrow \begin{aligned} x_1 + x_2 &= 0 \\ x_2 &= -x_1 \end{aligned} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ -x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned}$$

$\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$  is a basis for  $E_{\lambda=-\frac{1}{6}}$ .

For  $E_{\lambda=1}$ :

$$\begin{bmatrix} \frac{2}{3} & 1\frac{1}{2} \\ -\frac{2}{3} & \frac{1}{2} \end{bmatrix} \rightarrow \begin{bmatrix} \frac{2}{3} & -\frac{1}{2} \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in E_{\lambda=1} \rightarrow \begin{array}{l} x_1 - \frac{3}{4}x_2 = 0 \\ x_1 = \frac{3}{4}x_2 \end{array}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{4}x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} \frac{3}{4} \\ 1 \end{bmatrix} = \frac{x_2}{4} \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$\left\{ \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right\}$  is a basis for  $E_{\lambda=1}$ .

$P = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}$  diagonalizes  $A$ . Therefore  $P^{-1} = \frac{1}{7} \begin{bmatrix} 4 & -3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{7} & -\frac{3}{7} \\ -\frac{1}{7} & \frac{1}{7} \end{bmatrix}$ .

$$P^{-1}AP = D = \begin{bmatrix} -\frac{1}{6} & 0 \\ 0 & 1 \end{bmatrix}$$

$$A = PDP^{-1} \rightarrow A^k = PD^kP^{-1}$$

$$\begin{aligned} A^k \begin{bmatrix} a \\ b \end{bmatrix} &= PD^kP^{-1} \begin{bmatrix} a \\ b \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} (\frac{1}{6})^k & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{4}{7} & -\frac{3}{7} \\ -\frac{1}{7} & \frac{1}{7} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \\ &= \frac{1}{7} \begin{bmatrix} a(4(-\frac{1}{6})^k + 3) + b(-3(-\frac{1}{6})^k + 3) \\ a(-4(-\frac{1}{6})^k + 4) + b(3(-\frac{1}{6})^k + 4) \end{bmatrix} \end{aligned}$$

As  $k \rightarrow \infty$

$$A^k \begin{bmatrix} a \\ b \end{bmatrix} \rightarrow \frac{1}{7} \begin{bmatrix} 3a + 3b \\ 4a + 4b \end{bmatrix} = \begin{bmatrix} \frac{3}{7} \\ \frac{4}{7} \end{bmatrix}$$

$\begin{bmatrix} \frac{3}{7} \\ \frac{4}{7} \end{bmatrix}$  is called the stable state of the matrix.

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{3}{7} \\ \frac{4}{7} \end{bmatrix} = \begin{bmatrix} \frac{3}{7} \\ \frac{4}{7} \end{bmatrix}$$

After 5 years, X's portion of the market differs from  $\frac{3}{7}$  by 0.00013.

**Theorem 4.3.** Let  $A$  be a stochastic matrix. Then  $\lambda = 1$  is always an eigenvalue of  $A$ . Moreover, the stable state of the corresponding Markov process is the element in  $E_{\lambda=1}$  with the property that the sum of the components is equal to 1.

*Proof.* Recall that the eigenvalues of  $A$  are identical to the eigenvalues of  $A^T$ . To show that  $\lambda = 1$  is always an eigenvalue of  $A$ , we will demonstrate that  $\lambda = 1$  is always an eigenvalue

of  $A^T$ .

$$A^T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$A^T \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

So  $\lambda = 1$  is an eigenvalue of  $A^T$ , hence of  $A$ . If  $\left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right\}$  is a basis for  $E_{\lambda=1}$  for matrix  $A$ ,

then any element in  $E_{\lambda=1}$  is of the form:

$$\begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{bmatrix} \rightarrow \begin{cases} kx_1 + kx_2 + \cdots + kx_n = 1 \\ k(x_1 + x_2 + \cdots + x_n) = 1 \end{cases} \quad k = \frac{1}{x_1 + x_2 + \cdots + x_n}$$

□

### 4.3 Systems of Differential Equations

$y' = ky$  has a general solution:  $y = ck^{kx}$  where  $c$  is the value of  $y$  corresponding to  $x = 0$ .

System of Differential Equations:  $Y' = AY$

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \\ &\vdots \\ y_n' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n \end{aligned}$$

or in matrix form:

$$\begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

**Example 4.12.** Consider the system:  $y_1' = 2y_1$  and  $y_2' = 3y_2$  where  $y_1(0) = 4$  and  $y_2(0) = 6$ . Find  $y_1$  and  $y_2$ .

$$\begin{aligned} y_1 &= c_1 e^{2x} \rightarrow y_1(0) = c_1 e^{2 \cdot 0} \rightarrow c_1 = 4 \\ y_2 &= c_2 e^{3x} \rightarrow y_2(0) = c_2 e^{3 \cdot 0} \rightarrow c_2 = 6 \end{aligned}$$

$$y_1 = 4e^{2x}, y_2 = 6e^{6x} \rightarrow \begin{aligned} y_1' &= 2y_1 + 0y_2 \\ y_2' &= 0y_1 + 3y_2 \end{aligned}$$

$$Y' = AY$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}' = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$A$  is diagonal.

**Example 4.13.** Solve the system:  $y_1' = 4y_1 - y_2$ ,  $y_2' = 2y_1 + y_2$  given that  $y_1(0) = 3$  and  $y_2(0) = -1$ .

$$Y' = AY$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}' = \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

From earlier considerations, we know that  $A$  is diagonalizable. It has eigenvalues of 2 and 3 and  $P = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$  diagonalizes  $A$ .

Let  $Y = PU$  where  $U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ , then  $Y' = PU'$ . Substituting into  $Y' = AY$ , we get  $PU' = AP U$ . Then  $U' = (P^{-1}AP)U = DU$ .

$$U_1 = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} U \rightarrow \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}' = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$u_1 = c_1 e^{2x}, u_2 = c_2 e^{3x}$$

Therefore

$$Y = PU \rightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} c_1 e^{2x} \\ c_2 e^{3x} \end{bmatrix} \rightarrow \begin{aligned} y_1 &= c_1 e^{2x} + c_2 e^{3x} \\ y_2 &= 2c_1 e^{2x} + c_2 e^{3x} \end{aligned}$$

$$y_1(0) = 3 = c_1 + c_2$$

$$y_2(0) = 4 = 2c_1 + c_2$$

$$c_1 = -4, c_2 = 7$$

Plug back into  $y_1, y_2$ :

$$y_1 = 4e^{2x} + 7e^{3x}, y_2 = -8e^{2x} + 7e^{3x}$$

**Example 4.14.** Solve the differential equation:  $y'' + y' - 20y = 0$  given that  $y''(0) = 10$ ,  $y'(0) = 13$ .

Define  $y_1 = y$ ,  $y_2 = y'$ . Therefore

$$y_1' = y' = y_2 \rightarrow y_1' = 0y_1 + y_2$$

$$y_2' = y'' = 20y - y' = 20y_1 - y_2 \rightarrow y_2' = 20y_1 - y_2$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}' = \begin{bmatrix} 0 & 1 \\ 20 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad A \text{ does not diagonalize}$$

Eigenvalue of  $A$ :

$$\det\left(\begin{bmatrix} \lambda & -1 \\ -20 & \lambda + 1 \end{bmatrix}\right) = 0 \rightarrow \begin{aligned} \lambda(\lambda + 1) - 20 &= 0 \\ \lambda^2 + \lambda - 20 &= 0 \\ (\lambda - 4)(\lambda + 5) &= 0 \\ \lambda &= 4, -5 \end{aligned}$$

For  $E_{\lambda=4}$ :

$$\begin{aligned} \begin{bmatrix} 4 & -1 \\ 20 & 5 \end{bmatrix} &\rightarrow \begin{bmatrix} 4 & -1 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in E_{\lambda=4} &\rightarrow \begin{aligned} 4x_1 - x_2 &= 0 \\ x_2 &= 4x_1 \end{aligned} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ 4x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} \end{aligned}$$

Base:  $\left\{ \begin{bmatrix} 4 \\ 1 \end{bmatrix} \right\}$

For  $E_{\lambda=-5}$ :

$$\begin{aligned} \begin{bmatrix} -5 & -1 \\ -20 & -4 \end{bmatrix} &\rightarrow \begin{bmatrix} -5 & -1 \\ 0 & 0 \end{bmatrix} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in E_{\lambda=-5} &\rightarrow \begin{aligned} -5x_1 + x_2 &= 0 \\ x_2 &= 5x_1 \end{aligned} \\ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= \begin{bmatrix} x_1 \\ 5x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 5 \end{bmatrix} \end{aligned}$$

Base:  $\left\{ \begin{bmatrix} 1 \\ -5 \end{bmatrix} \right\}$

$P = \begin{bmatrix} 1 & 1 \\ 4 & -5 \end{bmatrix}$  diagonalizes  $A$  and  $P^{-1}AP = D = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}$

$$Y = PU \rightarrow PU' = APU \rightarrow U' = (P^{-1}AP)U \rightarrow \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}' = D \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{aligned} u_1' &= 4u_1 & u_1 &= c_1 e^{4x} \\ u_2' &= -5u_2 & u_2 &= c_2 e^{-5x} \end{aligned}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 4 & -5 \end{bmatrix} \begin{bmatrix} c_1 e^{4x} \\ c_2 e^{-5x} \end{bmatrix} \rightarrow \begin{aligned} y_1 &= c_1 e^{4x} + c_2 e^{-5x} \\ y_2 &= 4c_1 e^{4x} - 5c_2 e^{-5x} \end{aligned}$$

$$\begin{aligned} y(0) &= 10 & c_1 + c_2 &= 10 & c_1 &= 7 \\ y'(0) &= 13 & 4c_1 - 5c_2 &= 13 & c_2 &= 3 \end{aligned}$$

Solution:  $y = 7e^{4x} + 3e^{-5x}$

## 5 Inner Product Spaces

### 5.1 Inner Product

**Definition 5.1.** Let  $V$  be a vector space. The inner product on  $V$  is a function that assigns to each pair of vectors  $u, v \in V$  a unique real number, written  $\langle u, v \rangle$  such that the following properties hold:

1. Symmetry Axiom:  $\langle u, v \rangle = \langle v, u \rangle$  for all  $u, v \in V$
2. Additivity Axiom:  $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$  for all  $u, v, w \in V$
3. Homogeneity Axiom:  $\langle ku, v \rangle = k\langle u, v \rangle$  for all  $u, v \in V$
4. Positivity Axiom:  $\langle v, v \rangle \geq 0$  for all  $v \in V$  where equality holds if and only if  $v = 0$

**Example 5.1.** Let  $V = \mathbb{R}^n$ . Let  $\alpha_1, \alpha_2, \dots, \alpha_n$  be  $n$  possible scalars for  $u = (u_1, u_2, \dots, u_n)$  and  $v = (v_1, v_2, \dots, v_n)$ . Define  $\langle u, v \rangle = \alpha_1 u_1 v_1 + \alpha_2 u_2 v_2 + \dots + \alpha_n u_n v_n$ .

1.  $\langle u, v \rangle = \langle v, u \rangle$  ✓
2.  $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$  ✓
3.  $\langle ku, v \rangle = k\langle u, v \rangle$  ✓
4.  $\langle v, v \rangle = \alpha_1 v_1 v_1 + \alpha_2 v_2 v_2 + \dots + \alpha_n v_n v_n = \alpha_1 v_1^2 + \alpha_2 v_2^2 + \dots + \alpha_n v_n^2 \geq 0$

*Proof.*

$$\begin{aligned}
 \langle u, v \rangle &= 0 \\
 \alpha_1 v_1^2 + \alpha_2 v_2^2 + \dots + \alpha_n v_n^2 &= 0 \\
 \alpha_i v_i^2 &= 0 \text{ for } 1 \leq i \leq n \\
 v_i^2 &= 0 \text{ for } 1 \leq i \leq n \\
 v_i &= 0 \text{ for } 1 \leq i \leq n \\
 v &= (0, 0, \dots, 0) = 0 \\
 \langle v, v \rangle &= \alpha_1 0^2 + \alpha_2 0^2 + \dots + \alpha_n 0^2 = 0
 \end{aligned}$$

□

If one or none of the scalars is equal to 0, this is not an inner product on  $\mathbb{R}^n$  since the second part of axiom IV does not necessarily hold. This is called a weighed Euclidean inner product on  $\mathbb{R}^n$ . The  $\alpha$ 's are called the weights.

This example shows that there can be infinitely many products on an inner product space. Special Case:  $\alpha_1 = \alpha_2 = \dots = \alpha_n = 1 \rightarrow u_1 v_1 + u_2 v_2 + \dots + u_n v_n$  is called the Euclidean inner product on  $\mathbb{R}^n$ . In  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , this is simply the dot product.

Let  $A$  be an  $n \times n$  matrix and define the trace of  $A$  as  $\text{tr}(A) = \sum_{i=1}^n a_{ii}$   
Properties of the Trace:



1.  $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
2.  $\text{tr}(kA) = k\text{tr}(A)$
3.  $\text{tr}(A^T) = \text{tr}(A)$

**Example 5.2.** Let  $V = M_{n \times n}$ . For  $A, B \in V$ , define  $\langle A, B \rangle = \text{tr}(A^T B)$ .

1.  $\langle A, B \rangle \text{tr}(A^T B) = \text{tr}((A^T B)^T) = \text{tr}(B^T A) = \langle B, A \rangle$
2.  $\langle A + B, C \rangle = \text{tr}((A + B)^T C) = \text{tr}((A^T + B^T)C) = \text{tr}(A^T C + B^T C) = \text{tr}(A^T C) + \text{tr}(B^T C) = \langle A, C \rangle + \langle B, C \rangle$
3.  $\langle kA, B \rangle = \text{tr}((kA)^T B) = \text{tr}(kA^T B) = k\text{tr}(A^T B) = k\langle A, B \rangle$
4. It is easy to show that  $\text{tr}(A^T A)$  is equal to the sum of the squares of the entries of  $A$ .  
 $\langle A, A \rangle = \text{tr}(A^T A) = \text{sum of squares of the entries of } A \geq 0$ , where equality holds if and only if  $A = 0$ .

**Theorem 5.1.** Let  $V$  be an inner product space with inner product  $\langle \rangle$ , then

1.  $\langle 0, v \rangle = 0$  for all  $v \in V$
2.  $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$  (right additivity)
3.  $\langle u, kv \rangle = k\langle u, v \rangle$  (right homogeneity)
4.  $\langle u - v, w \rangle = \langle u, w \rangle - \langle v, w \rangle$  (subtractivity)
5.  $\langle u, v - w \rangle = \langle u, v \rangle - \langle u, w \rangle$  (right subtractivity)

*Proof.* 1. Choose  $k = 0$  in axiom III.  $\langle 0u, v \rangle = 0\langle u, v \rangle \rightarrow \langle 0, v \rangle = 0$

$$2. \langle u, v + w \rangle = \langle v + w, u \rangle = \langle v + u, w + u \rangle = \langle u, v \rangle + \langle u, w \rangle$$

3. Similar to (3)

$$4. \langle u - v, w \rangle = \langle u + (-1)v, w \rangle = \langle u, w \rangle + \langle (-1)v, w \rangle = \langle u, w \rangle - \langle v, w \rangle$$

$$5. \langle u, v - w \rangle = \langle u, v \rangle - \langle u, w \rangle \text{ Similar to (4)}$$

□

**Theorem 5.2.** Cauchy-Schwarz Inequality: Let  $V$  be an inner product space with inner product  $\langle \rangle$ , then for any vectors  $u, v \in V$ ,

$$\langle u, v \rangle^2 = \langle u, u \rangle \langle v, v \rangle$$

*Proof.* Case 1:  $u = 0$ . LHS becomes  $\langle 0, v \rangle^2 = 0^2 = 0$ . RHS becomes  $\langle 0, 0 \rangle \langle v, v \rangle = 0 \langle v, v \rangle = 0$ . Thus  $0 \leq 0$ .

Case 2:  $u \neq 0$ . For any real number  $x$ , consider  $\langle xu + v, xu + v \rangle$ . This is  $\geq 0$  by axiom IV. Thus

$$\begin{aligned}\langle xu + v, xu + v \rangle &= \langle xu, xu + v \rangle + \langle v, xu + v \rangle \\ &= \langle xu, xu \rangle + \langle xu, v \rangle + \langle x, xu \rangle + \langle v, v \rangle \\ &= x^2 \langle u, u \rangle + x \langle u, v \rangle + x \langle v, u \rangle + \langle v, v \rangle\end{aligned}$$

Then  $ax^2 + bx + c \geq 0$ . Consider the graph of  $y = ax^2 + bx + c$ . Since  $u \neq 0$ ,  $a = \langle u, u \rangle \geq 0$ . So this graph is a parabola that opens upward. Moreover, since  $y \geq 0$ , this parabola cannot drop below the x axis.  $ax^2 + bx + c = 0$  has either no real solutions or exactly one solution. This means that either  $b^2 - 4ac < 0$  or  $b^2 - 4ac = 0$ . That is,  $(2\langle u, v \rangle)^2 - 4\langle u, u \rangle \langle v, v \rangle \leq 0$ . Thus,  $\langle u, v \rangle^2 - 2\langle u, v \rangle \langle v, v \rangle \leq 0$   $\square$

**Example 5.3.** If  $V = \mathbb{R}^n$  and  $\langle \rangle$  is Euclidean, then for  $u = (u_1, u_2, \dots, u_n)$  and  $v = (v_1, v_2, \dots, v_n)$ , Cauchy's Inequality states:

$$(u_1v_1 + u_2v_2 + \dots + u_nv_n)^2 \leq (u_1^2 + u_2^2 + \dots + u_n^2)(v_1^2 + v_2^2 + \dots + v_n^2)$$

**Example 5.4.** Let  $V = \mathbb{P}_n$ . For  $f, g \in V$ , define  $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ . It is a straightforward exercise to show that this is an inner product on  $V$ . By Cauchy's inequality,

$$\left(\int_0^1 f(x)g(x)dx\right)^2 \leq \left(\int_0^1 f^2(x)dx\right)\left(\int_0^1 g^2(x)dx\right)$$

## 5.2 Length, Distance and Angle for Inner Product Spaces (Orthogonality)

**Definition 5.2.** Let  $V$  be an inner product space with inner product  $\langle \rangle$ , then if  $u \in V$ , then the norm of  $u$  (or the length of  $u$ ), written  $\|u\|$ , is given by

$$\|u\| = \langle u, u \rangle^{\frac{1}{2}}$$

**Definition 5.3.** Let  $V$  be an inner product space with inner product  $\langle \rangle$ , then if  $u, v \in V$ , then the distance between  $u$  and  $v$ , written  $d(u, v)$ , is given by

$$d(u, v) = \|u - v\|$$

**Example 5.5.** Let  $V = \mathbb{R}^2$  and let  $\langle \rangle$  be Euclidean. Let  $u = \langle u_1, u_2 \rangle$  and  $v = \langle v_1, v_2 \rangle$ . Then

$$\|u\| = \sqrt{u_1^2 + u_2^2}$$

$$d(u, v) = \|u - v\| = \|(u_1 - v_1, u_2 - v_2)\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$$

**Example 5.6.** Let  $V = \mathbb{P}_2$  and let  $f, g \in V$ . Then  $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ . Choose  $f = f(x) = x$  and  $g = g(x) = x^2$ . Then

$$\|f\| = \langle f, f \rangle^{\frac{1}{2}} = \sqrt{\int_0^1 (x)^2 dx} = \sqrt{\left.\frac{x^3}{3}\right|_0^1} = \sqrt{\frac{1}{3}} = \frac{1}{\sqrt{3}}$$

$$\|g\| = \langle g, g \rangle^{\frac{1}{2}} = \sqrt{\int_0^1 (x^2)^2 dx} = \sqrt{\left. \frac{x^5}{5} \right|_0^1} = \sqrt{\frac{1}{5}} = \frac{1}{\sqrt{5}}$$

So  $f$  is greater than  $g$ .

$$\langle f, g \rangle = \int_0^1 (x)(x^2) dx = \int_0^1 x^3 dx = \left. \frac{x^4}{4} \right|_0^1 = \frac{1}{4}$$

Is Cauchy-Schwarz Theorem satisfied?

$$\langle f, g \rangle^2 \leq \langle f, f \rangle \langle g, g \rangle$$

$$\left(\frac{1}{4}\right)^2 \leq \left(\frac{1}{3}\right)\left(\frac{1}{5}\right)$$

$$\frac{1}{16} \leq \frac{1}{15} \checkmark$$

$$\begin{aligned} d(f, g) &= \|f - g\| \\ &= \sqrt{\int_0^1 (x - x^2)^2 dx} \\ &= \sqrt{\int_0^1 (x^2 - 2x^3 + x^4) dx} \\ &= \sqrt{\left[ \frac{x^3}{3} - \frac{x^4}{2} + \frac{x^5}{5} \right]_0^1} \\ &= \sqrt{\frac{1}{3} - \frac{1}{2} + \frac{1}{5}} \\ &= \sqrt{\frac{1}{30}} \end{aligned}$$

Note:

$$\begin{aligned} \langle u, v \rangle^2 &\leq \langle u, u \rangle \langle v, v \rangle \\ &= \|u\|^2 \|v\|^2 \\ &\leq (\|u\| \|v\|)^2 \\ |\langle u, v \rangle| &\leq \|u\| \|v\| \\ |\langle u, v \rangle| &\leq \|u\| \|v\| \end{aligned}$$

**Theorem 5.3.** Let  $V$  be an inner product space with inner product  $\langle \cdot \rangle$ . For all  $u, v, w \in V$  and  $k$  scalar,

1.  $\|u\| \geq 0$ , where equality holds if and only if  $u = 0$
2.  $d(u, v) \geq 0$ , where equality holds if and only if  $u = v$
3.  $\|u + v\| \leq \|u\| + \|v\|$  (Triangle Inequality)
4.  $d(u, v) \leq d(u, w) + d(w, v)$  (Triangle Inequality)

$$5. \|ku\| = |k|\|u\|$$

$$6. d(v, u) = d(u, v)$$

*Proof.* 1. Since  $\|u\| = \sqrt{\langle u, u \rangle}$ , a principle square root,  $\|u\| \geq 0$ . If  $\|u\| = 0$ , then  $\langle u, u \rangle = 0$  and thus  $u = 0$  (by axiom 4)

$$2. d(u, v) = \|u, v\| \geq 0 \text{ where equality holds if and only if } u - v = 0, \text{ i.e, if and only if } u = v.$$

3. In the proof of Cauchy-Schwarz Inequality, we found  $\langle xu + v, xu + v \rangle = x^2\langle u, u \rangle + 2x\langle u, v \rangle + \langle v, v \rangle$ . Choosing  $x = 1$ ,  $\langle u + v, u + v \rangle = \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle$  or  $\|u + v\|^2 = \|u\|^2 + 2\langle u, v \rangle + \|v\|^2 \geq \|u\|^2 + 2|\langle u, v \rangle| + \|v\|^2$ . By the Cauchy-Schwarz Inequality,  $|\langle u, v \rangle| \leq \|u\|\|v\| \geq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$ . Thus  $\|u + v\|^2 \leq (\|u\| + \|v\|)^2$  or  $\|u + v\| \leq \|u\| + \|v\|$ .

$$4. d(u, v) = \|u - v\| = \|(u - w) - (w - v)\| \leq \|u - w\| + \|w - v\| = d(u, w) + d(w, v)$$

$$5. \|ku\|^2 = \langle ku, ku \rangle = k^2\langle u, u \rangle = k^2\|u\|^2 = (|k|\|u\|)^2. \text{ Thus } \|ku\| = |k|\|u\|.$$

$$6. d(v, u) = \|v - u\| = \|-1(u - v)\| = |-1|\|u - v\| = \|u - v\| = d(u, v)$$

□

Suppose  $u, v$  are two nonzero vectors chosen from inner product space  $V$  with inner product  $\langle \cdot \rangle$ . From the alternative version of the Cauchy-Schwarz Inequality, we know  $|\langle u, v \rangle| \leq \|u\|\|v\|$ . Dividing by the positive quantity  $\|u\|\|v\|$  on both sides, we get  $\frac{|\langle u, v \rangle|}{\|u\|\|v\|} \leq 1$ . Equivalently,  $|\frac{\langle u, v \rangle}{\|u\|\|v\|}| \leq 1$ . Therefore

$$-1 \leq \frac{\langle u, v \rangle}{\|u\|\|v\|} \leq 1$$

**Definition 5.4.** Let  $u, v$  be two nonzero vectors chosen from inner product space  $V$  with inner product  $\langle \cdot \rangle$ . Then the angle between  $u$  and  $v$ , which we represent by  $\theta$ , is the unique angle between 0 and  $\pi$  with the property that

$$\cos(\theta) = \frac{\langle u, v \rangle}{\|u\|\|v\|}$$

or

$$\theta = \arccos\left(\frac{\langle u, v \rangle}{\|u\|\|v\|}\right)$$

**Example 5.7.** Let  $V = P_2$ . Let  $f(x) = x$  and  $g(x) = x^2$ . Find the angle between  $f$  and  $g$  where  $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$ .

$$\langle f, g \rangle = \frac{1}{4}, \|f\| = \frac{1}{\sqrt{3}}, \|g\| = \frac{1}{\sqrt{5}}$$

$$\cos(\theta) = \frac{\langle f, g \rangle}{\|f\|\|g\|} = \frac{\sqrt{15}}{4}$$

$$\theta = \arccos\left(\frac{\sqrt{15}}{4}\right) = 0.25 \text{ radians or } 14.5^\circ$$

**Example 5.8.** Let  $V = M_{2 \times 2}$  and  $\langle A, B \rangle = \text{tr}(A^T B)$ . If  $A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}$  and  $B = \begin{bmatrix} -1 & 2 \\ 6 & 2 \end{bmatrix}$ , find the angle between  $A$  and  $B$ .

$$\|A\| = \langle A, A \rangle^{\frac{1}{2}} = \sqrt{\text{tr}(A^T A)} = \sqrt{(2)^2 + (1)^2 + (-1)^2 + (3)^2} = \sqrt{15}$$

$$\|B\| = \langle B, B \rangle^{\frac{1}{2}} = \sqrt{\text{tr}(B^T B)} = \sqrt{(-1)^2 + (2)^2 + (6)^2 + (2)^2} = \sqrt{45}$$

$$\langle A, B \rangle = \text{tr}(A^T B) = \text{tr}\left(\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 6 & 2 \end{bmatrix}\right) = \text{tr}\left(\begin{bmatrix} -8 & x \\ x & 8 \end{bmatrix}\right) = 0$$

$$\cos(\theta) = \frac{\langle A, B \rangle}{\|A\|\|B\|} = \frac{0}{\sqrt{15}\sqrt{45}} = 0 \rightarrow \theta = \arccos(0) = \frac{\pi}{2}$$

**Definition 5.5.** Let  $V$  be an inner product space with inner product  $\langle \cdot, \cdot \rangle$ . Two nonzero vectors  $u, v \in V$  are said to be orthogonal, written  $u \perp v$ , if and only if  $\theta$ , the angle between  $u$  and  $v$ , is equal to  $\frac{\pi}{2}$ .

In the previous example,  $A \perp B$ .

Note: We can rearrange  $\cos(\theta) = \frac{\langle u, v \rangle}{\|u\|\|v\|}$  as follows:  $\langle u, v \rangle = \|u\|\|v\| \cos(\theta)$ . Then the following is equivalent:

$$\langle u, v \rangle = 0 \iff \cos(\theta) = 0 \iff \theta = \frac{\pi}{2} \iff u \perp v$$

**Example 5.9.** If  $V = \mathbb{R}^2$  and  $u = (1, 3)$  and  $v = (6, -2)$ , can we claim that  $u \perp v$ ?

1. If the inner product is Euclidean, the answer is YES since  $(1)(6) + (3)(-2) = 0$
2. if the inner product is weighed Euclidean inner product with  $\alpha_1 = 2$  and  $\alpha_2 = 3$ , then the answer is NO since  $(2)(1)(5) + (3)(3)(-2) \neq 0$

This example shows that orthogonality is dependent upon the inner product being used.

**Theorem 5.4.** Generalized Pythagorean Theorem: Let  $V$  be an inner product space with inner product  $\langle \cdot, \cdot \rangle$ . Let  $u, v$  be two nonzero vectors chosen from  $V$ . Then

$$u \perp v \iff \|u + v\|^2 = \|u\|^2 + \|v\|^2$$

*Proof.*  $u \perp v$  implies  $\langle u, v \rangle = 0$  Then:

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle \\ &= \|u\|^2 + 2\langle u, v \rangle + \|v\|^2 \\ &= \|u\|^2 + \|v\|^2 \end{aligned}$$

□

### 5.3 Orthonormality, The Gram-Schmidt Process

**Example 5.10.** Suppose  $V = \mathbb{R}^3$  and we choose the Euclidean inner product on  $V$ . Define  $S = \{v_1, v_2, v_3\}$  where  $v_1 = (1, -1, 1)$ ,  $v_2 = (0, 1, 1)$  and  $v_3 = (2, 1, -1)$ .

$$\langle v_1, v_2 \rangle = (1)(0) + (-1)(1) + (1)(1) = 0 \rightarrow v_1 \perp v_2$$

$$\langle v_1, v_3 \rangle = (1)(2) + (-1)(1) + (1)(-1) = 0 \rightarrow v_1 \perp v_3$$

$$\langle v_2, v_3 \rangle = (0)(2) + (1)(1) + (1)(-1) = 0 \rightarrow v_2 \perp v_3$$

**Definition 5.6.** Let  $V$  be an inner product space with inner product  $\langle \cdot, \cdot \rangle$ . A set of vectors  $\{v_1, v_2, \dots, v_n\}$ , none of which equal to 0, is called orthogonal if and only if  $\langle v_i, v_j \rangle = 0$  for all  $1 \leq i, j \leq n$ ,  $i \neq j$ .

In the previous example,  $S$  is an orthogonal set of vectors chosen from  $\mathbb{R}^3$  with respect to the Euclidean inner product on  $\mathbb{R}^3$ .

**Theorem 5.5.** Let  $V$  be an inner product space with inner product  $\langle \cdot, \cdot \rangle$  and let  $S = \{v_1, v_2, \dots, v_n\}$  be an orthogonal set of vectors chosen from  $V$ . Then  $S$  is linearly independent.

*Proof.* Consider  $k_1v_1 + k_2v_2 + \dots + k_nv_n = 0$ . To show  $S$  is linearly independent, we have to show  $k_1 = k_2 = \dots = k_n = 0$ . Choose  $v_i \in S$  and consider  $\langle 0, v_i \rangle = 0$ . So then  $\langle k_1v_1 + k_2v_2 + \dots + k_iv_i + \dots + k_nv_n, v_i \rangle = 0$ . This yields  $\langle k_1v_1, v_i \rangle + \langle k_2v_2, v_i \rangle + \dots + \langle k_iv_i, v_i \rangle + \dots + \langle k_nv_n, v_i \rangle = 0$ . Thus  $k_1\langle v_1, v_i \rangle + k_2\langle v_2, v_i \rangle + \dots + k_i\langle v_i, v_i \rangle + \dots + k_n\langle v_n, v_i \rangle = 0$ . Since  $S$  is an orthogonal set,  $\langle v_j, v_i \rangle = 0$  for all  $i \neq j$ . This then becomes  $k_i\langle v_i, v_i \rangle = 0$ . Since  $v_i \neq 0$ ,  $\langle v_i, v_i \rangle \neq 0$  and so  $k_i = 0$ . This is independent of the choice of  $i$ . So because  $k_1 = k_2 = \dots = k_n = 0$ ,  $S$  is linearly independent.  $\square$

In the previous example,  $S = \{(1, -1, 1), (0, 1, 1), (2, 1, -1)\}$  was shown to be an orthogonal set with respect to the Euclidean inner product on  $\mathbb{R}^3$ . Based on the previous theorem,  $S$  is also linearly independent. But  $S$  contains 3 vectors and  $\dim(\mathbb{R}^3) = 3$ . So  $S$  is also a basis for  $\mathbb{R}^3$ . Such a basis is called an orthogonal basis.

**Definition 5.7.** Suppose that  $V$  is an inner product space with inner product  $\langle \cdot, \cdot \rangle$  and suppose  $v \in V$  is a nonzero vector.

$$\left\| \frac{1}{\|v\|} v \right\| = \left| \frac{1}{\|v\|} \right| \|v\| = \frac{1}{\|v\|} \|v\| = 1$$

This product has yielded a vector of length 1 and it is called normalization.

**Example 5.11.** Normalize the following vectors:

$$v_1 = (1, -1, 1) \rightarrow \|v_1\| = \sqrt{(1)^2 + (-1)^2 + (1)^2} = \sqrt{3} \rightarrow w_1 = \frac{v_1}{\|v_1\|} = \left( \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right)$$

$$v_2 = (0, 1, 1) \rightarrow \|v_2\| = \sqrt{(0)^2 + (1)^2 + (1)^2} = \sqrt{2} \rightarrow w_2 = \frac{v_2}{\|v_2\|} = \left( 0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$$

$$v_3 = (2, 1, -1) \rightarrow \|v_3\| = \sqrt{(2)^2 + (1)^2 + (-1)^2} = \sqrt{6} \rightarrow w_3 = \frac{v_3}{\|v_3\|} = \left(\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)$$

It is easy to verify that  $\{w_1, w_2, w_3\}$  is an orthogonal set. The vectors in this set all have length of 1 (such vectors are called unit vectors).

**Definition 5.8.** Let  $V$  be an inner product space with inner product  $\langle \cdot, \cdot \rangle$ . A set of nonzero vectors chosen from  $V$  is said to be orthonormal if and only if the set consists of unit vectors which are pairwise orthogonal.

Accordingly, in the previous example,  $\{w_1, w_2, w_3\}$  is an orthonormal set of vectors chosen from  $\mathbb{R}^3$  with respect to the Euclidean inner product, where  $w_1 = \frac{v_1}{\|v_1\|} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$ ,  $w_2 = \frac{v_2}{\|v_2\|} = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$  and  $w_3 = \frac{v_3}{\|v_3\|} = \left(\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right)$ .

Note: Orthonormal  $\rightarrow$  orthogonal  $\rightarrow$  linearly independent

So  $\{w_1, w_2, w_3\}$  is linearly independent, and hence form a basis on  $\mathbb{R}^3$ . Such a basis is called an orthonormal basis.

Using the orthonormal basis leads to pleasant computations.

**Theorem 5.6.** Let  $\{w_1, w_2, \dots, w_n\}$  be an orthonormal basis for inner product space  $V$  with inner product  $\langle \cdot, \cdot \rangle$ . For any  $u \in V$ ,

$$(u)_S = (\langle u, w_1 \rangle, \langle u, w_2 \rangle, \dots, \langle u, w_n \rangle)$$

*Proof.* Since  $S$  is a basis for  $V$ , the vector  $u \in V$  is uniquely expressible as a linear combination of  $w_1, w_2, \dots, w_n$ . That is,

$$u = k_1 w_1 + k_2 w_2 + \dots + k_n w_n$$

where the  $k$ 's are weighed scalars. This means that  $(u)_S = (k_1, k_2, \dots, k_n)$ . To prove the theorem, we need to show that  $k_i = \langle u, w_i \rangle$  where  $1 \leq i \leq n$ .

$$\begin{aligned} \langle u, w_i \rangle &= \langle k_1 w_1 + k_2 w_2 + \dots + k_n w_n, w_i \rangle \\ &= k_1 \langle w_1, w_i \rangle + k_2 \langle w_2, w_i \rangle \\ &\quad + \dots + k_i \langle w_i, w_i \rangle + \dots + k_n \langle w_n, w_i \rangle \\ &= k_i \langle w_i, w_i \rangle = k_i \end{aligned}$$

□

**Example 5.12.** In the previous example, find  $(u)_S$  where

$$S = \left\{ \left(\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}\right) \right\}$$

and  $u = (1, 1, 1)$ .

$$\begin{aligned} \langle u, w_1 \rangle &= (1)\left(\frac{1}{\sqrt{3}}\right) + (1)\left(-\frac{1}{\sqrt{3}}\right) + (1)\left(\frac{1}{\sqrt{3}}\right) = \frac{1}{\sqrt{3}} \\ \langle u, w_2 \rangle &= (1)(0) + (1)\left(\frac{1}{\sqrt{2}}\right) + (1)\left(\frac{1}{\sqrt{2}}\right) = \frac{2}{\sqrt{2}} \end{aligned}$$

$$\begin{aligned}\langle u, w_3 \rangle &= (1)\left(\frac{2}{\sqrt{6}}\right) + (1)\left(\frac{1}{\sqrt{6}}\right) + (1)\left(-\frac{1}{\sqrt{6}}\right) = \frac{2}{\sqrt{6}} \\ (u)_3 &= \left(\frac{1}{\sqrt{3}}, \frac{2}{\sqrt{2}}, \frac{2}{\sqrt{6}}\right)\end{aligned}$$

If  $V$  is an inner product space with inner product  $\langle \rangle$  and  $S$  is an orthonormal basis for  $V$ , say  $S = \{w_1, w_2, \dots, w_n\}$ , then for vectors  $u, v \in V$ ,  $(u)_S = (k_1, k_2, \dots, k_n)$  and  $(v)_S = (l_1, l_2, \dots, l_n)$ , we can show the following to be true:

1.  $\|u\| = \sqrt{k_1^2 + k_2^2 + \dots + k_n^2}$
2.  $d(u, v) = \sqrt{(k_1 - l_1)^2 + (k_2 - l_2)^2 + \dots + (k_n - l_n)^2}$
3.  $\langle u, v \rangle = k_1 l_1 + k_2 l_2 + \dots + k_n l_n$

Problem: Let  $V$  be a nonzero finite-dimensional inner product space with inner product  $\langle \rangle$ . If  $\{v_1, v_2, \dots, v_n\}$  is a basis for  $V$ , how can we obtain an orthonormal basis for  $V$ ? The process for doing this is called the Gram-Schmidt Orthonormalization process. There are two steps:

1. Obtain an orthogonal basis for  $V$
2. From the result of Stage 1, obtain an orthonormal basis for  $V$

Stage 1:

1. Define  $v_1 = u_1$
2. Define  $v_2 = u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1$

Claim:  $v_2 \neq 0$ . If  $v_2 = 0$ , then  $u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 = 0$ . So  $u_2 = \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1$  or  $u_2 = k v_1$ . This is impossible since  $u_1 + u_2$  are part of a basis for  $V$  and thus form a linearly independent set. This proves the claim

$$\begin{aligned}\langle u_2, v_1 \rangle &= \left\langle u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1, v_1 \right\rangle \\ &= \langle u_2, v_1 \rangle - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} \langle v_1, v_1 \rangle \\ &= \langle u_2, v_1 \rangle - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} \|v_1\|^2 = 0\end{aligned}$$

Thus  $v_2 \perp v_1$

3. Define  $v_3 = u_3 - \frac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2$ .  
We can show that  $v_3 \neq 0$ .

4. Furthermore, it is true that  $\langle v_3, v_1 \rangle = 0$  and  $\langle v_3, v_2 \rangle = 0$ . Thus  $v_3 \perp v_1, v_3 \perp v_2$ . Since  $v_2 \perp v_1$ , we now have  $\{v_1, v_2, v_3\}$  to be an orthogonal set.



5. After  $n$  steps, we will obtain  $\{v_1, v_2, \dots, v_n\}$  to be an orthogonal set of vectors chosen from  $V$ . Since orthogonality implies linear independence, we will have a linearly independent set of  $n$  vectors in a space of dimension  $n$ . Hence  $\{v_1, v_2, \dots, v_n\}$  is an orthogonal basis for  $V$ .

Stage 2:

1. Define  $w_1 = \frac{v_1}{\|v_1\|}$ ,  $w_2 = \frac{v_2}{\|v_2\|}$ ,  $\dots$ ,  $w_n = \frac{v_n}{\|v_n\|}$
2. Then  $\{w_1, w_2, \dots, w_n\}$  will be an orthonormal basis for  $V$ .

**Example 5.13.** Let  $u_1 = (3, 0, 0, 0)$ ,  $u_2 = (0, 1, 2, 1)$ ,  $u_3 = (0, -1, 3, 2)$ . Find the orthonormal basis of  $V$  with respect to the Euclidean inner product on  $V$ .

It is easy to show that  $\{u_1, u_2, u_3\}$  is a linearly independent set of vectors chosen from  $\mathbb{R}^4$ . Define  $V = \text{span}\{u_1, u_2, u_3\}$ . Then  $\{u_1, u_2, u_3\}$  is a basis for  $V$ . Thus

$$\begin{aligned} v_1 &= (3, 0, 0, 0) \\ v_2 &= (0, 1, 2, 1) - 0v_1 = (0, 1, 2, 1) \\ v_3 &= (0, -1, 3, 2) - 0v_1 - \frac{7}{6}(0, 1, 2, 1) = (0, -\frac{13}{6}, \frac{4}{6}, \frac{5}{6}) \end{aligned}$$

$\{v_1, v_2, v_3\}$  is an orthogonal basis for  $V$ .

$$\begin{aligned} w_1 &= \frac{v_1}{\|v_1\|} = \frac{1}{3}(3, 0, 0, 0) = (1, 0, 0, 0) \\ w_2 &= \frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{6}}(0, 1, 2, 1) = (0, \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}) \\ w_3 &= \frac{v_3}{\|v_3\|} = \frac{1}{\frac{\sqrt{210}}{6}}(0, -\frac{13}{6}, \frac{4}{6}, \frac{5}{6}) = (0, -\frac{13}{\sqrt{210}}, \frac{4}{\sqrt{210}}, \frac{5}{\sqrt{210}}) \end{aligned}$$

$\{w_1, w_2, w_3\}$  is an orthonormal basis for  $V$ .

**Example 5.14.** Let  $V = P_2$  and let  $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$  for  $f, g \in V$ . We already know that  $\{1, x, x^2\}$  is basis for  $V$ . Find an orthonormal basis for  $V$  with respect to the given inner product.

$$\begin{aligned} u_1 &= 1, u_2 = x, u_3 = x^2 \\ v_1 &= u_1 = 1 \\ v_2 &= u_2 - \frac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 = x - \frac{\int_0^1 x dx}{\int_0^1 dx} \cdot 1 = x - \frac{\frac{1}{2}}{1} \cdot 1 = x - \frac{1}{2} \\ v_3 &= u_3 - \frac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2 = x^2 - \frac{\int_0^1 x^2 dx}{\int_0^1 dx} \cdot 1 - \frac{\int_0^1 (x^3 - \frac{1}{2}x^2) dx}{\int_0^1 (x - \frac{1}{2})^2 dx} \cdot (x - \frac{1}{2}) = x^2 - x + \frac{1}{6} \\ w_1 &= \frac{v_1}{\|v_1\|} = \frac{1}{1} = 1 \\ w_2 &= \frac{v_2}{\|v_2\|} = \frac{x - \frac{1}{2}}{\frac{1}{2\sqrt{3}}} = 2\sqrt{3}(x - \frac{1}{2}) = \sqrt{3}(2x - 1) \\ w_3 &= \frac{v_3}{\|v_3\|} = \frac{x^2 - x + \frac{1}{6}}{\frac{1}{6\sqrt{5}}} = 6\sqrt{5}(x^2 - x + \frac{1}{6}) = \sqrt{5}(6x^2 - 6x + 1) \end{aligned}$$

## 6 Linear Transformations

### 6.1 Linear Transformations

**Definition 6.1.** Let  $V$  and  $W$  be two vector spaces. A function  $T : V \rightarrow W$  is called a linear transformation from  $V$  into  $W$  if and only if  $T(v_1 + v_2) = T(v_1) + T(v_2)$  and  $T(kv_1) = kT(v_1)$ , for all  $v_1, v_2 \in V$  and  $k = \text{scalar}$ . That is,  $T$  is a function which preserves addition and scalar multiplication.

**Definition 6.2.** If  $T : V \rightarrow W$  is a linear transformation from  $V$  into  $W$ , then  $V$  is called the domain space of  $T$  and  $W$  is called the image space of  $T$ .

**Example 6.1.** 1. Let  $T : P_n \rightarrow P_{n-1}$  be defined by  $T(f) = f'$  for all  $f \in P_n$ . Choose  $f, g \in P_n$ .

$$T(f + g) = (f + g)' = f' + g' = T(f) + T(g)$$

$$T(kf) = (kf)' = k(f') = kT(f)$$

So  $T$  is a linear transformation from  $P_n$  into  $P_{n-1}$ . It is called the differential transformation.

2. Let  $T : M_{m \times n} \rightarrow M_{n \times m}$  be defined by  $T(A) = A^T$  for all  $A \in M_{m \times n}$ .

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

$$T(kA) = (kA)^T = k(A^T) = kT(A)$$

So  $T$  is a linear transformation from  $M_{m \times n}$  into  $M_{n \times m}$ . It is called the transposition transformation.

3. Let  $T : M_{n \times n} \rightarrow \mathbb{R}^1$  be defined by  $T(A) = \det(A)$ .

$$T(A + B) = \det(A + B) \neq \det(A) + \det(B) \neq T(A) + T(B)$$

Since  $T$  fails to preserve addition, it is not a linear transformation.

4. Let  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be defined by  $T((x_1, x_2, x_3)) = (x_1, x_1 + x_2, x_1 + x_2 + x_3)$

$$\begin{aligned} T((x_1, x_2, x_3) + (y_1, y_2, y_3)) &= (x_1 + y_1, x_1 + y_1 + x_2 + y_2, x_1 + y_1 + x_2 + y_2 + x_3 + y_3) \\ &= (x_1, x_1 + x_2, x_1 + x_2 + x_3) + (y_1, y_1 + y_2, y_1 + y_2 + y_3) \\ &= T((x_1, x_2, x_3)) + T((y_1, y_2, y_3)) \end{aligned}$$

$$\begin{aligned} T(k(x_1, x_2, x_3)) &= T((kx_1, kx_2, kx_3)) \\ &= (kx_1, kx_1 + kx_2, kx_1 + kx_2 + kx_3) \\ &= k(x_1, x_1 + x_2, x_1 + x_2 + x_3) \\ &= kT((x_1, x_2, x_3)) \end{aligned}$$

So  $T$  is a linear transformation from  $\mathbb{R}^3$  into  $\mathbb{R}^3$ . It is also a linear operator from  $\mathbb{R}^3$  into itself.

5. Let  $V$  be a vector space and let  $T : V \rightarrow \{0\}$  be defined by  $T(v) = 0$  for all  $v \in V$ .

$$T(v_1 + v_2) = 0 = 0 + 0 = T(v_1) + T(v_2)$$

$$T(kv_1) = 0 = k0 = kT(v_1)$$

So  $T$  is a linear transformation from  $V$  into  $\{0\}$ . It is called the zero transformation.

6. Let  $V$  be a vector space and let  $T : V \rightarrow V$  be defined by  $T(v) = v$  for all  $v \in V$ .

$$T(v_1 + v_2) = v_1 + v_2 = T(v_1) + T(v_2)$$

$$T(kv_1) = kv_1 = kT(v_1)$$

So  $T$  is a linear operator on  $V$ . It is called the identity operator on  $V$ .

7. Let  $V$  be a nonzero finite-dimensional vector space of dimension  $n$  and let  $B = \{v_1, v_2, \dots, v_n\}$  be a basis for  $V$ . For  $v \in V$ , we can write

$$v = k_1v_1 + k_2v_2 + \dots + k_nv_n$$

where the  $k$ 's are unique. Define  $T : V \rightarrow \mathbb{R}^n$  by  $T(v) = (v)_B$ . ( $T(v) = (k_1, k_2, \dots, k_n)$ )

$$v = k_1v_1 + k_2v_2 + \dots + k_nv_n$$

$$v' = k'_1v_1 + k'_2v_2 + \dots + k'_nv_n$$

$$v + v' = (k_1 + k'_1)v_1 + (k_2 + k'_2)v_2 + \dots + (k_n + k'_n)v_n$$

$$\begin{aligned} T(v + v') &= (k_1 + k'_1, k_2 + k'_2, \dots, k_n + k'_n) \\ &= (k_1, k_2, \dots, k_n) + (k'_1, k'_2, \dots, k'_n) = T(v) + T(v') \end{aligned}$$

$$kv = kk_1v_1 + kk_2v_2 + \dots + kk_nv_n$$

$$T(kv) = (kk_1, kk_2, \dots, kk_n) = k(k_1, k_2, \dots, k_n) = kT(v)$$

So  $T$  is a linear transformation from  $V$  into  $\mathbb{R}^n$ .

8. Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be defined by  $T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$  where  $A$  is a given  $m \times n$  matrix.

That is,  $T(X) = AX$  for all  $X \in \mathbb{R}^n$ .

$$T(X + X') = A(X + X') = AX + AX' = T(X) + T(X')$$

$$T(kX) = A(kX) = k(AX) = kT(X)$$

So  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear transformation from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ . We say that  $T$  is reduced by the matrix  $A$  or that  $A$  induces  $T$ .

9. It is easy to verify that  $C$ , the set of everywhere continuous functions, is a vector space under the standard operations of addition and scalar multiplication. Define  $T : C \rightarrow \mathbb{R}$  as follows:  $T(f) = \int_0^1 f(t)dt$  where  $f \in C$ .

$$T(f+g) = \int_0^1 (f+g)(t)dt = \int_0^1 (f(t)+g(t))dt = \int_0^1 f(t)dt + \int_0^1 g(t)dt = T(f) + T(g)$$

$$T(kf) = \int_0^1 (kf)(t)dt = \int_0^1 f(t)dt = k \int_0^1 f(t)dt = kT(f)$$

So  $T$  is a linear transformation from  $C$  into  $\mathbb{R}$ .

**Definition 6.3.** A linear transformation  $T : V \rightarrow V$  from  $V$  into itself is called linear operator on  $V$ .

In the previous example, the linear transformation in (4) is a linear operator on  $\mathbb{R}^3$ .

**Theorem 6.1.** Let  $T : V \rightarrow W$  be a linear transformation from  $V$  into  $W$ . Then

1.  $T(0_v) = 0_w$
2.  $T(v_1 - v_2) = T(v_1) - T(v_2)$  for all  $v_1, v_2 \in V$
3.  $T(k_1v_1 + k_2v_2 + \cdots + k_nv_n) = k_1T(v_1) + k_2T(v_2) + \cdots + k_nT(v_n)$

*Proof.* 1.  $T(0_v) = T(0v) = 0T(v) = 0_w$

$$2. T(v_1 - v_2) = T(v_1 + (-1)v_2) = T(v_1) + (-1)T(v_2) = T(v_1) - T(v_2)$$

3. Using the properties of linear transformations or mathematical induction, we get this result. □

Consequence of (3): Suppose  $\{v_1, v_2, \dots, v_n\}$  is a basis for  $V$ . If we know the image under  $T$  of these basic vectors, then we can define the image under  $T$  of any vector in  $V$ .

**Example 6.2.** If  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is a linear transformation from  $\mathbb{R}^3$  into  $\mathbb{R}^3$  for which  $T((1, 1, 0)) = (2, 3)$ ,  $T((1, 0, 1)) = (-1, 2)$  and  $T((0, 1, 1)) = (3, 1)$ , find  $T((x, y, z))$ .

It is easy to show that  $\{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$  is a basis for  $\mathbb{R}^3$ .

$$(x, y, z) = k_1(1, 1, 0) + k_2(1, 0, 1) + k_3(0, 1, 1)$$

$$\begin{bmatrix} 1 & 1 & 0 & x \\ 1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & x \\ 0 & -1 & 1 & y-x \\ 0 & 1 & 1 & z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & y \\ 0 & 1 & -1 & -y+x \\ 0 & 0 & 1 & \frac{z+y-x}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \frac{x+y-z}{2} \\ 0 & 1 & 0 & \frac{x-y+z}{2} \\ 0 & 0 & 1 & \frac{-x+y+z}{2} \end{bmatrix}$$

$$\begin{aligned} T((x, y, z)) &= T\left(\left(\frac{x+y-z}{2}\right)(1, 1, 0) + \left(\frac{x-y+z}{2}\right)(1, 0, 1) + \left(\frac{-x+y+z}{2}\right)(0, 1, 1)\right) \\ &= \left(\frac{x+y-z}{2}\right)T((1, 1, 0)) + \left(\frac{x-y+z}{2}\right)T((1, 0, 1)) + \left(\frac{-x+y+z}{2}\right)T((0, 1, 1)) \\ &= \left(\frac{x+y-z}{2}\right)(2, 3) + \left(\frac{x-y+z}{2}\right)(-1, 2) + \left(\frac{-x+y+z}{2}\right)(3, 1) = (-x+3y, 3x-z) \end{aligned}$$

**Definition 6.4.** Let  $T : V \rightarrow W$  be a linear transformation from  $V$  into  $W$ .

- the kernel of  $T$ , written  $\ker(T)$ , is the subset of  $V$  consisting of all vectors in  $V$  whose image under  $T$  is the zero vector in  $W$

$$\ker(T) = \{v \in V | T(v) = 0\}$$

If  $T$  is the differential transformation, its kernel consists of all constant functions.

- the range of  $T$ , written  $R(T)$ , is the subset of  $W$  consisting of all vectors in  $W$  which are images of at least one vector in  $V$

$$R(T) = \{w \in W | w = T(v) \text{ for some } v \in V\}$$

If  $T$  is the differential transformation, its image is all of  $P_{n-1}$ .

**Definition 6.5.** If  $T : M_{m \times n} \rightarrow M_{n \times m}$  where  $T(A) = A^T$  then  $\ker(T) = \{0\}$  and  $R(T) = M_{n \times m}$

**Definition 6.6.** Suppose  $A$  is an  $m \times n$  matrix and suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is the linear transformation induced by  $A$ .

That is,  $T(v) = Av$  for all  $v \in \mathbb{R}^n$ ,

$$\begin{aligned} v \in \ker(T) &\leftrightarrow Tv = 0 \\ &\leftrightarrow Av = 0 \\ &\leftrightarrow v \in N(A) \end{aligned}$$

$v$  is a solution of the homogeneous system with coefficient matrix  $A$ . Thus  $\ker(T) = N(A)$ .

$$\begin{aligned} w \in R(T) &\leftrightarrow T(v) = w \text{ for some } v \in \mathbb{R}^n \\ &\leftrightarrow Av = w \text{ for some } v \in \mathbb{R}^n \\ &\leftrightarrow Ax = w \text{ is a consistent linear system with coefficient matrix } A \\ &\leftrightarrow w \in \text{Col}(A) \\ R &= \text{Col}(A) \end{aligned}$$

**Theorem 6.2.** Let  $T : V \rightarrow W$  be a linear transformation from  $V$  into  $W$

- $\ker(T)$  is a subspace of  $V$
- $R(T)$  is a subspace of  $W$

*Proof.* •

$$\begin{aligned} v_1, v_2 &\in \ker(T) \\ T(v_1 + v_2) &= T(v_1) + T(v_2) \\ &= 0 + 0 = 0 \\ &\rightarrow v_1 + v_2 \in \ker(T) \end{aligned}$$

•

$$\begin{aligned}
T(kv_1) &= kT(v_1) \\
&= k0 \\
&= 0 \\
&\rightarrow kv_1 \in \ker(T) \text{ for some scalar } k
\end{aligned}$$

•

$$\begin{aligned}
w_1, w_2 &\in R(T) \\
w_1 \in R(T) &\rightarrow T(v_1) = w_1 \text{ for some } v_1 \in V \\
w_2 \in R(T) &\rightarrow T(v_2) = w_2 \text{ for some } v_2 \in V \\
w_1 + w_2 &= T(v_1) + T(v_2) = T(v_1 + v_2) \\
w_1 + w_2 &\in R(T)
\end{aligned}$$

•

$$\begin{aligned}
T(v_1) &= w_1 \\
kT(v_1) &= kw_1 \\
T(kw_1) &= kw_1 \\
kw_1 &\in R(T) \text{ for some scalar } k
\end{aligned}$$

□

Suppose  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is the linear transformation induced by  $m \times n$  matrix  $A$ . Then we have already observed that  $\ker(T) = N(A)$  and  $R(T) = Col(A)$ . So

$$\dim(\ker(T)) = \dim(N(A)) = n(A)$$

and

$$\dim(R(T)) = \dim(Col(A)) = r(A)$$

**Definition 6.7.** More generally, suppose  $T : V \rightarrow W$  is a linear transformation from  $V$  into  $W$ .

- if  $\ker(T)$  is finite-dimensional, then  $\dim(\ker(T))$  is called the nullity of  $T$ , written  $n(T)$
- if  $R(T)$  is finite-dimensional, then  $\dim(R(T))$  is called the rank of  $T$ , written  $r(T)$

**Example 6.3.** • For a matrix transformation,  $n(T) = n(A)$  and  $r(T) = r(A)$

- For a differential transformation,  $n(T) = 1$  and  $r(T) = n$
- For a transposition transformation,  $n(T) = 0$  and  $r(T) = nm$

**Example 6.4.** Let  $T : \mathbb{R}^4 \rightarrow \mathbb{R}^3$  be a linear transformation induced by  $A = \begin{bmatrix} 2 & 1 & 0 & -1 \\ -1 & 3 & 1 & 0 \\ 5 & -1 & -1 & -2 \end{bmatrix}$ .  
 $(T(v) = Av \text{ for all } v \in \mathbb{R}^n)$ . Find the rank and nullity of  $T$ .

$$w \in R(T) \rightarrow T(v) = w \text{ for some } v \in \mathbb{R}^4 \rightarrow Av = w \text{ for some } v \in \mathbb{R}^4$$

$$\begin{bmatrix} 2 & 1 & 0 & -1 \\ -1 & 3 & 1 & 0 \\ 5 & -1 & -1 & -2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 0 & -1 & w_1 \\ -1 & 3 & 1 & 0 & w_2 \\ 5 & -1 & -1 & -2 & w_3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -3 & -1 & 0 & -w_2 \\ 0 & 1 & \frac{2}{7} & -\frac{1}{7} & \frac{w_1+2w_2}{7} \\ 0 & 0 & 0 & 0 & w_3 - 2w_1 + w_2 \end{bmatrix}$$

$$w_3 - 2w_1 + w_2 = 0 \rightarrow w_3 = 2w_1 - w_2$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ 2w_1 - w_2 \end{bmatrix} = \begin{bmatrix} w_1 \\ 0 \\ 2w_1 \end{bmatrix} + \begin{bmatrix} 0 \\ w_2 \\ -w_2 \end{bmatrix} = w_1 \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} + w_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\} \text{ is a basis for } R(T) \text{ So } r(T) = 2.$$

$$v \in \ker(T) \rightarrow T(v) = 0 \rightarrow Av = 0$$

$$\begin{bmatrix} 2 & 1 & 0 & -1 \\ -1 & 3 & 1 & 0 \\ 5 & -1 & -1 & -2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Continuing the reduction process, the reduced row echelon form of  $A$  is:

$$\begin{bmatrix} 1 & 0 & -\frac{2}{7} & -\frac{3}{7} \\ 0 & 1 & \frac{2}{7} & -\frac{1}{7} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So

$$v_1 - \frac{1}{7}v_3 - \frac{3}{7}v_4 = 0, v_2 + \frac{2}{7}v_3 - \frac{1}{7}v_4 = 0$$

$$v_1 = \frac{1}{7}v_3 + \frac{3}{7}v_4, v_2 = -\frac{2}{7}v_3 - \frac{1}{7}v_4$$

Then

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{7}v_3 + \frac{3}{7}v_4 \\ -\frac{2}{7}v_3 - \frac{1}{7}v_4 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{7}v_3 \\ -\frac{2}{7}v_3 \\ v_3 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{3}{7}v_4 \\ \frac{1}{7}v_4 \\ 0 \\ v_4 \end{bmatrix} = v_3 \begin{bmatrix} \frac{1}{7} \\ -\frac{2}{7} \\ 1 \\ 0 \end{bmatrix} + v_4 \begin{bmatrix} \frac{3}{7} \\ \frac{1}{7} \\ 0 \\ 1 \end{bmatrix}$$

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 7 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \\ 7 \end{bmatrix} \right\} \text{ is a basis for } \ker(T). \text{ So } n(T) = 2.$$

Recall that if  $A$  is an  $m \times n$  matrix  $r(A) + n(A) = n$  (Dimension Theorem for Matrices).

**Theorem 6.3.** If  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is the linear transformation induced by  $A$ , then  $r(A) = r(T)$  and  $n(A) = n(T)$  and  $n = \dim(\mathbb{R}^n)$ . So

$$r(T) + n(T) = \dim(\mathbb{R}^n)$$

**Theorem 6.4.** Dimension Theorem for Linear Transformation: If  $T : V \rightarrow W$  is a linear transformation from  $V$  into  $W$ , where  $V$  is finite-dimensional, then

$$r(T) + n(T) = \dim(V)$$

## 6.2 Isomorphism

**Example 6.5.** If  $T : P_2 \rightarrow P_3$  is the differentiation transformation:

$$T(x^2) = 2x$$

$$T(x^2 + 1) = 2x$$

**Example 6.6.** If  $T : M_{m \times n} \rightarrow M_{n \times m}$  is the transposition transformation:

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

$$T(B) = B^T$$

If  $A^T = B^T$ ,

$$(A^T)^T = (B^T)^T \rightarrow A = B$$

**Definition 6.8.** A linear transformation  $T : V \rightarrow W$  is said to be one-to-one (1-1) if and only if for  $v_1, v_2$ , two distinct vectors in  $V$ , it follows that  $T(v_1) \neq T(v_2)$ . Contrapositively, if  $T(v_1) = T(v_2)$ , then  $v_1 = v_2$ .

The first example is not 1-1 while the second example is 1-1.

**Theorem 6.5.** Let  $T : V \rightarrow W$  be a linear transformation from  $V$  into  $W$ . Then  $T$  is 1-1 if and only if  $\ker(T) = \{0\}$

*Proof.* (Forward) Choose  $v \in \ker(T)$ . Then  $T(v) = 0$ . But we know  $T(0) = 0$ . So  $T(v) = T(0)$ . Since  $T$  is 1-1,  $v = 0$ . That is,  $\ker(T) = \{0\}$ .

(Backward) Suppose  $T(v_1) = T(v_2)$  for some  $v_1, v_2 \in V$ . Then

$$T(v_1) = T(v_2)$$

$$T(v_1) - T(v_2) = 0$$

$$T(v_1 - v_2) = 0$$

$$v_1 - v_2 \in \ker(T)$$

$$\text{Since } \ker(T) = \{0\}$$

$$v_1 - v_2 = 0$$

$$v_1 = v_2$$

So  $T$  is 1-1. □



Corollary:  $T$  is 1-1 if and only if  $n(T) = 0$ .

**Theorem 6.6.** Let  $A$  be an  $n \times n$  matrix and let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be the linear operator induced by  $A$ .

$$A \text{ is invertible} \leftrightarrow n(A) = 0 \leftrightarrow n(T) = 0 \leftrightarrow T \text{ is } 1 - 1$$

This adds an eleventh statement to the Equivalence Theorem:  $T$  is 1-1 where  $T$  is induced by  $A$

**Definition 6.9.** Let  $T : V \rightarrow W$  be a linear transformation from  $V$  into  $W$ . We say that  $T$  is onto if and only if  $R(T) = W$ .

**Example 6.7.** Let  $T : P_2 \rightarrow P_3$  be the linear transformation defined by  $T(p(x)) = xp(x)$  for all  $p \in P_2$ . Show that (1)  $T$  is 1-1 (2)  $T$  is not onto.

1. Suppose  $p \in \ker(T)$ . Then  $T(p) = 0$  which means that  $xp(x) = 0$  for all  $x$ . This implies that  $p(x) = 0$  for all  $x$ . Thus,  $T$  is 1-1.
2. Suppose  $q \in P_3$  and  $q \in R(T)$ . Then  $T(p) = q$  for some  $p \in P_2$ .

$$p = p(x) = ax^2 + bx + c$$

$$T(p) = q(x) = ax^3 + bx^2 + cx$$

There is no constant term present in  $q$ . So  $R(T) \neq P_3$ . ( $x^3 + x^2 + x + 1$  is not in  $R(T)$ .)  
So  $T$  is not onto.

Extra:  $n(T) = 0, r(T) = 3$  since  $\{x^3, x^2, x\}$  is a basis for  $R(T)$ .

$$r(T) + n(T) = \dim(P_3)$$

$$3 + 0 = 3$$

Suppose  $T : V \rightarrow W$  is a linear transformation from  $V$  into  $W$  where  $\dim(V) = \dim(W) = n$ . Then

$$\begin{aligned} T \text{ is } 1 - 1 &\leftrightarrow n(T) = 0 \\ &\leftrightarrow r(T) = \dim(V) \\ &\leftrightarrow r(T) = \dim(W) \\ &\leftrightarrow \dim(R(T)) = \dim(W) \\ &\leftrightarrow R(T) = W \\ &\leftrightarrow T \text{ is onto} \end{aligned}$$

So if  $V$  and  $W$  are the finite-dimensional vector spaces with equal dimensions,

$$T \text{ is } 1-1 \leftrightarrow T \text{ is onto}$$

**Theorem 6.7.** If  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is induced by the invertible  $n \times n$  matrix, then  $A$  is invertible  $\leftrightarrow T$  is 1-1. So  $T$  is 1-1  $\leftrightarrow T$  is onto. We now have statement twelve in the Equivalence Theorem:  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is onto, where  $T$  is induced by  $A$ .

**Theorem 6.8.** Let  $T : V \rightarrow W$  be a linear transformation from  $V$  into  $W$  where  $V$  and  $W$  are finite-dimensional and  $\dim(V) \neq \dim(W)$ . Then

1. if  $\dim W < \dim(V)$ , then  $T$  is not 1-1
2. if  $\dim(V) < \dim(W)$ , then  $T$  is not onto

*Proof.* 1. Suppose  $T$  is 1-1. Then  $n(T) = 0$ . By the Dimension Theorem for Linear Transformation,  $r(T) = \dim(V)$ . By hypothesis,  $\dim(W) < \dim(V)$ . Then  $\dim(W) < r(T)$ . That is,  $\dim(W) < \dim(R(T))$ . This is a contradiction since  $R(T)$  is a subspace of  $W$ . Thus  $\dim(R(T)) < \dim(W)$ . Conclude:  $T$  is not 1-1.

2. Suppose  $T$  is onto. Then  $R(T) = W$ . So  $r(T) = \dim(W)$ . By the Dimension Theorem for Linear Transformation,  $r(T) + n(T) = \dim(V)$ . Thus  $\dim(W) + \dim(T) = \dim(V)$ . Therefore  $n(T) = \dim(V) - \dim(W)$ . Hence  $n(T) < 0$ . This is impossible. So  $T$  is not onto.

□

**Definition 6.10.** A linear transformation  $T : V \rightarrow W$  is called an isomorphism if and only if  $T$  is both 1-1 and onto. If such an isomorphism can be found, we say that the vector spaces  $V$  and  $W$  are isomorphic to one another.

**Theorem 6.9.** Let  $V$  be a nonzero finite-dimensional vector space of dimension  $n$ . Then  $V$  is isomorphic to  $\mathbb{R}^n$ .

*Proof.* Let  $\{v_1, v_2, \dots, v_n\}$  be a basis for  $V$ . Call it  $B$ . For any  $v \in V$ , we can write

$$v = k_1 v_1 + k_2 v_2 + \dots + k_n v_n$$

where the  $k$ 's are unique. Define  $T : V \rightarrow \mathbb{R}^n$  by  $T(v) = (v)_B$ . That is,  $T(v) = (k_1, k_2, \dots, k_n)$ .

$T$  is 1-1:

$$\begin{aligned} v &\in \ker(T) \\ (k_1, k_2, \dots, k_n) &= (0, 0, \dots, 0) \\ k_1 = k_2 = \dots = k_n &= 0 \\ v &= 0 \end{aligned}$$

So  $\ker(T) = \{0\}$  which implies that  $T$  is 1-1.

Since  $v$  and  $\mathbb{R}^n$  have the same dimension and  $T$  is 1-1, it follows that  $T$  is also onto. So  $T$  is an isomorphism which mean that  $v$  and  $\mathbb{R}^n$  are isomorphic to one another . □

**Example 6.8.**  $P_3$  is isomorphic to  $\mathbb{R}^4$ .

$$T(a_0 + a_1x + a_2x^2 + a_3x^3) = (a_0, a_1, a_2, a_3)$$

$M_{2 \times 3}$  is isomorphic to  $\mathbb{R}^6$

$$T\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = (a, b, c, d, e, f)$$

**Theorem 6.10.** Let  $V$  and  $W$  be two nonzero finite-dimensional vector spaces and let  $T : V \rightarrow W$  be a linear transformation from  $V$  into  $W$ . Then  $T$  is an isomorphism if and only  $\dim(V) = \dim(W)$ .

*Proof.* (Forward) Suppose  $T$  is an isomorphism.  $T$  is 1-1 and onto.  $T$  is 1-1 thus  $\dim(W) \geq \dim(V)$ . Also,  $T$  is onto thus  $\dim(V) \geq \dim(W)$ . Therefore,  $\dim(V) = \dim(W)$ .

(Reverse) Suppose  $\dim(V) = \dim(W)$ . Call their common value  $n$ . From the previous result,  $V$  is isomorphic to  $\mathbb{R}^n$ . Conclude that  $V$  and  $W$  must be isomorphic to one another.  $\square$

**Example 6.9.**  $P_3$  and  $M_{2 \times 3}$  are isomorphic spaces since

$$T(a_0 + a_1x + a_2x^2 + a_3x^3) = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$$

### 6.3 Compositions and Inverse Transformations

**Definition 6.11.** Let  $T_1 : U \rightarrow V$  and  $T_2 : V \rightarrow W$  be two linear transformations. Then the composition of  $T_2$  with  $T_1$  is the mapping from  $U$  into  $W$  defined by

$$(T_2 \circ T_1)(u) = T_2(T_1(u))$$

for all  $u \in U$ .

**Theorem 6.11.**  $T_2 \circ T_1$  is a linear transformation from  $V$  into  $W$ .

*Proof.*

$$\begin{aligned} (T_2 \circ T_1)(u + u') &= T_2(T_1(u + u')) \\ &= T_2(T_1(u) + T_1(u')) \\ &= T_2(T_1(u)) + T_2(T_1(u')) \\ &= (T_2 \circ T_1)(u) + (T_2 \circ T_1)(u') \\ (T_2 \circ T_1)(ku) &= T_2(T_1(ku)) \\ &= T_2(kT_1(u)) \\ &= kT_2(T_1(u)) \\ &= k((T_2 \circ T_1)(u)) \end{aligned}$$

So  $T_2 \circ T_1$  is a linear transformation.  $\square$

**Example 6.10.** Suppose  $T_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be defined by  $T_1(x, y) = (2x + y, x - y)$  and  $T_2 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  be defined by  $T_2(x, y) = (x - 3y, 2x + y)$ .

$$\begin{aligned} (T_2 \circ T_1)(x, y) &= T_2(T_1(x, y)) \\ &= T_2(2x + y, x - y) \\ &= ((2x - y) - 3(x - y), 2(2x + y) + (x - y)) \\ &= (-x + 4y, 5x + y) \\ (T_1 \circ T_2)(x, y) &= T_1(T_2(x, y)) \\ &= T_1(x - 3y, 2x + y) \\ &= (2(x - 3y) + 2x + y, x - 3y - (2x + y)) \\ &= (4x - 5y, -x - 5y) \end{aligned}$$

Here,  $T_1 \circ T - 2 \neq T_2 \circ T_1$ .

In general, compositions of linear transformations is not commutative.

Suppose  $T : V \rightarrow W$  is a linear transformation from  $V$  into  $W$  that is 1-1. For each  $w \in R(T)$ , then there is exactly one  $v \in V$  such that  $T(v) = w$  (because  $T$  is 1-1). This enables us to define a mapping  $U : R(T) \rightarrow V$  as follows.

**Definition 6.12.**

$$U(w) = v \text{ if and only if } T(v) = w$$

Claim:  $U$  is a linear transformation from  $R(T)$  into  $V$ . Let  $w_1, w_2 \in R(T)$ . If  $w_1 \in R(T)$ , there is a unique  $v_1 \in V$  such that  $T(v_1) = w_1$ . In the same manner, if  $w_2 \in R(T)$ , there is a unique  $v_2 \in V$  such that  $T(v_2) = w_2$ . Then

$$w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2)$$

This is equivalent to saying that  $U(w_1 + w_2) = v_1 + v_2$ . But

$$T(v_1) = w_1 \leftrightarrow U(w_1) = v_1$$

and

$$T(v_2) = w_2 \leftrightarrow U(w_2) = v_2$$

Thus  $U(w_1 + w_2) = U(w_1) + U(w_2)$ . Now, let  $w_1 \in R(T)$ , thus  $T(v_1) = w_1$  for some  $v_1 \in V$ .

$$kw_1 = kT(v_1)$$

$$= T(kv_1)$$

$$U(w_1) = kv_1$$

$$U(kw_1) = kU(w_1)$$

Therefore:

$$U(w_1 + w_2) = U(w_1) + U(w_2)$$

$$U(kw_1) = kU(w_1)$$

Thus,  $U$  is a linear transformation from  $R(T)$  into  $V$ .

For any  $v \in V$ :

$$U(T(v))v \text{ and } (U \circ T)(v) = v$$

For any  $w \in W$ :

$$T(U(w)) = w \text{ and } (T \circ U)(w) = w$$

The two transformations  $T$  and  $U$  "undo" one another. We call such transformations inverse of one another. Notation: We write  $T^{-1}$  to represent the inverse of  $T$ .

For  $T$  1-1,  $T^{-1}$  exists:

$$T^{-1}(w) = v \leftrightarrow T(v) = w$$

$$T^{-1}(T(v)) = v \text{ for all } v \in V$$

$$T(T^{-1}(w)) = w \text{ for all } w \in R(T)$$

If  $T$  is an isomorphism, then  $T$  is also onto. So that  $R(T) = W$ . For such a  $T$ :

$$T^{-1}(T(v)) = v \text{ for all } v \in V$$

$$T(T^{-1}(w)) = w \text{ for all } w \in W$$

**Example 6.11.** Suppose  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is the linear operator on  $\mathbb{R}^3$  defined by  $T(v_1, v_2, v_3) = (v_1, v_1 + v_2, v_1 + v_2 + v_3)$ .  $T$  is an isomorphism. Find  $T^{-1}(w_1, w_2, w_3)$ . It suffices to find  $T^{-1}(1, 0, 0), T^{-1}(0, 1, 0), T^{-1}(0, 0, 1)$ .

$$T^{-1}(1, 0, 0) = (v_1, v_2, v_3)$$

$$T(v_1, v_2, v_3) = (1, 0, 0)$$

$$(v_1, v_1 + v_2, v_1 + v_2 + v_3) = (1, 0, 0)$$

$$v_1 = 1, v_2 = -1, v_3 = 0$$

$$T^{-1}(1, 0, 0) = (1, -1, 0)$$

$$T^{-1}(0, 1, 0) = T(v_1, v_2, v_3)$$

$$v_1 = 0, v_2 = 1, v_3 = -1$$

$$T^{-1}(0, 1, 0) = (0, 1, -1)$$

$$T^{-1}(0, 0, 1) = T(v_1, v_2, v_3)$$

$$v_1 = 0, v_2 = 0, v_3 = -1$$

$$T^{-1}(0, 0, 1) = (0, 0, 1)$$

$$\begin{aligned} T^{-1}(w_1, w_2, w_3) &= T^{-1}(w_1(1, 0, 0) + w_2(0, 1, 0) + w_3(0, 0, 1)) \\ &= w_1 T^{-1}(1, 0, 0) + w_2 T^{-1}(0, 1, 0) + w_3 T^{-1}(0, 0, 1) \\ &= w_1(1, -1, 0) + w_2(0, 1, -1) + w_3(0, 0, 1) \\ &= (w_1, w_2 - w_1, w_3 - w_2) \end{aligned}$$

**Theorem 6.12.** let  $A$  be an invertible  $n \times n$  matrix and let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be the linear operator on  $\mathbb{R}^n$  induced by  $A$  ( $T(v) = Av$  for all  $v \in \mathbb{R}^n$ .) Then  $T^{-1}$  exists, is a linear operator on  $\mathbb{R}^n$  and is induced by  $A^{-1}$ .

*Proof.*  $A$  is invertible, so  $T$  is 1-1 (by Equivalence Theorem). Therefore  $T^{-1}$  exists.  $T^{-1} : R(T) \rightarrow \mathbb{R}^n$ . Since  $T$  is a linear operator,  $T$  is also onto. So  $R(T) = \mathbb{R}^n$ . Therefore  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a linear operator on  $\mathbb{R}^n$ , For  $w \in \mathbb{R}^n$ , there is a unique  $v \in \mathbb{R}^n$  such that  $T(v) = w$ .

$$T^{-1}(T(v)) = T^{-1}(w)$$

$$v = T^{-1}(w)$$

$$Av = w$$

$$A^{-1}Av = A^{-1}w$$

$$v = A^{-1}w$$

So  $T^{-1}(w) = A^{-1}(w)$  which means that  $T^{-1}$  is induced by  $A$ . □

## 7 Chapter 4

### 7.1 Matrix Transformation

We already know that every  $m \times n$  matrix induces a linear transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  defined by  $T(v) = Av$ .

Question: If  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a linear transformation from  $\mathbb{R}^n$  into  $\mathbb{R}^m$ , can we always find an  $m \times n$  matrix  $A$  that induces  $T$ ? Yes.

Let  $\{e_1, e_2, \dots, e_n\}$  be the standard basis for  $\mathbb{R}^n$ .

$$e_j = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ 1 \\ \vdots \\ e_n \end{bmatrix}$$

where the 1 is on the  $j^{\text{th}}$  position ( $1 \leq j \leq n$ ). Let

$$T(e_1) = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, T(e_2) = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \dots, T(e_n) = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Choose  $v \in \mathbb{R}^n$  ( $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ ) Then we can write  $v = v_1 e_1 + v_2 e_2 + \dots + v_n e_n$ . So

$$\begin{aligned} T(v) &= T(v_1 e_1 + v_2 e_2 + \dots + v_n e_n) \\ &= v_1 T(e_1) + v_2 T(e_2) + \dots + v_n T(e_n) \\ &= v_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \dots + v_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \dots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \dots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \dots + a_{mn}v_n \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}}_A \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}}_v \end{aligned}$$

$T(v) = Av$ . So  $A$  induces  $T$ .

To find  $A$ , the matrix that induces  $T$ , place the image of the standard basis vector of  $\mathbb{R}^n$  as columns of  $A$ .

**Example 7.1.** If  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  is the linear transformation defined by  $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ 2x_1 - x_2 \\ -x_1 + 4x_2 \end{bmatrix}$ .

Find  $A$ , the matrix that induces  $T$ .

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

$$T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$$

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

**Definition 7.1.** We call  $A$  the standard matrix for  $T$  and represent it by  $[T]$ .

$$T(v) = [T]v$$

Alternative notation:  $T_A$ .

The following results can be shown to be true:

1. If  $T_1 : \mathbb{R}^n \rightarrow \mathbb{R}^k$ ,  $T_2 : \mathbb{R}^k \rightarrow \mathbb{R}^m$  are linear transformations, then

$$[T_2 \circ T_1] = [T_2][T_1]$$

2. If  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a 1-1 linear transformation, then

$$[T^{-1}] = [T]^{-1}$$

**Example 7.2.** If  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is the linear operator on  $\mathbb{R}^3$  defined by  $T\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}\right) =$

$\begin{bmatrix} v_1 \\ v_1 + v_2 \\ v_1 + v_2 + v_3 \end{bmatrix}$ . Find the rule that defines  $T^{-1}$ . (We already know that  $T$  is 1-1. So  $T$  has an inverse.)

$$[T] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$[T]^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = [T^{-1}]$$

$$T^{-1}\left(\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}\right) = [T^{-1}]\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 - w_1 \\ w_3 - w_2 \end{bmatrix}$$