SURESH GYAN VIHAR UNIVERSITY
Accredited by NAAC with 'A+' Grade

Bachelor of Science

(B.Sc.)

LINEAR ALGEBRA

Semester-iv

Author- Dr. Manvendra Mishra

SURESH GYAN VIHAR UNIVERSITY
Centre for Distance and Online Education
Mahal, Jagatpura, Jaipur-302017

Published by:

S. B. Prakashan Pvt. Ltd.

WZ-6, Lajwanti Garden, New Delhi: 110046

Tel.: (011) 28520627 | Ph.: 9205476295

Email: info@sbprakashan.com | Web.: www.sbprakashan.com

Designed & Graphic by : S. B. Prakashan Pvt. Ltd.

Printed at :

CONTENTS

MODULE

ONE

SYSTEMS OF LINEAR EQUATIONS AND MATRICES

1.1 Matrices and Matrix operations

A **matrix** is a rectangular array of numbers. The numbers in the array are called the **entries** in the matrix. The **size** of a matrix is described in terms of the number of rows (horizontal lines) and columns (vertical lines) it contains.

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 4 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 0 & -3 \end{bmatrix} \quad \begin{bmatrix} e & \pi & -\sqrt{2} \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \begin{bmatrix} 4 \end{bmatrix}$$ are some examples.

The first matrix in example has three rows and two columns, so its size is 3 by 2 (written 3×2). In a size description, the first number always denotes the number of rows, and the second denotes the number of columns. The remaining matrices in example have sizes $1 \times 4,\ 3 \times 3,\ 2 \times 1$, and 1×1, respectively.

A matrix with only one row, such as the second in Example, is called a row vector (or a row matrix), and a matrix with only one column, such as the fourth in that example, is called a column vector (or a column matrix). The fifth matrix in that example is both a row vector and a column vector.

In general, a matrix with m rows and n columns is written as

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

A matrix A with n rows and n columns is called a **square matrix** of order n, and the entries $a_{11}, a_{22}, ..., a_{nn}$ in are said to be on the **main diagonal** of A.

Two matrices are defined to be **equal** if they have the same size and their corresponding entries are equal.

If A and B are matrices of the same size, then the sum A + B is the matrix obtained by adding the entries of B to the corresponding entries of A, and the difference A - B is the matrix obtained by subtracting the entries of B from the corresponding entries of A. Matrices of different sizes cannot be added or subtracted.

If A is any matrix and c is any scalar, then the product cA is the matrix obtained by multiplying each entry of the matrix A by c. The matrix cA is said to be a scalar multiple of A.

If A is an m $\times$ r matrix and B is an r $\times$ n matrix, then the product AB is the m $\times$ n matrix whose entries are determined as follows: To find the entry in row i and column j of AB, single out row i from the matrix A and column j from the matrix B. Multiply the corresponding entries from the row and column together, and then add up the resulting products.

If $A_1, A_2, \ldots, A_r$ are matrices of the same size, and if $c_1, c_2, \ldots, c_r$ are scalars, then an expression of the form

$c_1 A_1 + c_2 A_2 + \ldots + c_r A_r$ is called a **linear combination** of $A_1, A_2, \ldots, A_r$ with coefficients $c_1, c_2, \ldots, c_r$.

If A is an m $\times$ n matrix, and if x is an n $\times$ 1 column vector, then the product Ax can be expressed as a linear combination of the column vectors of A in which the coefficients are the entries of x.

The matrix product can be written as the following linear combination of column vectors, for example:

$\begin{bmatrix} -1 & 3 & 2 \\ 1 & 2 & -3 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$ can be written as

$$2 \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} - 1 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 2 \\ -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}.$$

Matrix multiplication has an important application to systems of linear equations. Consider a system of m linear equations in n unknowns:

$$a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = b_1$$

$$a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n = b_2$$

$$\ldots \qquad \ldots \qquad \ldots \qquad \ldots$$

$$a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n = b_m.$$

The above system of equations can be expressed as a single matrix equation $AX = B$ (3), where

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

. The matrix A in this equation is called the **coefficient** matrix of the system, the matrix X in this equation is called the **variable** matrix of the system, the matrix B in this equation is called the **constant** matrix of the system.

If A is any m × n matrix, then the **transpose** of A, denoted by A^T , is defined to be the n × m matrix that results by interchanging the rows and columns of A; that is, the first column of A^T is the first row of A, the second column of A^T is the second row of A, and so forth.

For example: If $A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \\ 5 & 6 \end{bmatrix}$, then $A^T = \begin{bmatrix} 2 & 1 & 5 \\ 3 & 4 & 6 \end{bmatrix}$

If A is a square matrix, then the **trace** of A, denoted by **tr(A)**, is defined to be the sum of the entries on the main diagonal of A. *The trace of A is undefined if A is not a square matrix.*

Examples: Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, $B = \begin{bmatrix} -1 & 3 & 2 \\ 1 & 2 & -3 \\ 2 & 1 & -2 \end{bmatrix}$,
then tr(A)=$a_{11} + a_{22} + a_{33}$ and tr(B)=(-1)+2+(-2)=-1.

Determine whether the statements given below are true or false.

(a) The matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ has no main diagonal.

(b) An m $\times$ n matrix has m column vectors and n row vectors.

(c) If A and B are 2 $\times$ 2 matrices, then AB = BA.

(d) The ith row vector of a matrix product AB can be computed by multiplying A by the ith row vector of B.

(e) For every matrix A, it is true that $(A^T)^T = A$.

(f) If A and B are square matrices of the same order, then tr(AB) = tr(A)tr(B) (g) If A and B are square matrices of the same order, then $(AB)^T = A^T B^T$.

(h) For every square matrix A, it is true that $tr(A^T) = tr(A)$.

(i) If A is a 6 $\times$ 4 matrix and B is an m $\times$ n matrix such that $B^T A^T$ is a 2 $\times$ 6 matrix, then m = 4 and n = 2.

(j) If A is an n $\times$ n matrix and c is a scalar,

then tr(cA) = c tr(A).

(k) If A, B, and C are matrices of the same size such that

A - C = B - C, then A = B.

(l) If A, B, and C are square matrices of the same order such that AC = BC, then A = B. (m) If AB + BA is defined, then A and B are square matrices of the same size.

(n) If B has a column of zeros, then so does AB if this product is defined.

(o) If B has a column of zeros, then so does BA if this product is defined.

Answers: (a) True (b) False (c) False (d) False (e) True (f) False (g) False (h) True (i) True (j) True (k) True (l) False (m)True (n) True (o) False

Algebraic Properties of Matrices

Properties of Matrix Arithmetic

Assuming that the sizes of the matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid.

(a) A + B = B + A [Commutative law for matrix addition]

(b) A + (B + C) = (A + B) + C [Associative law for matrix addition]

(c) A(BC) = (AB)C [Associative law for matrix multiplication]

(d) A(B + C) = AB + AC [Left distributive law]

(e) (B + C)A = BA + CA [Right distributive law]

(f) A(B - C) = AB - AC

(g) (B - C)A = BA - CA

(h) a(B + C) = aB + aC

(i) a(B - C) = aB - aC

(j) (a + b)C = aC + bC

(k) (a - b)C = aC - bC

(l) a(bC) = (ab)C

(m) a(BC) = (aB)C = B(aC)

Commutative law for multiplication will not hold always. The equality of AB and BA can fail for three possible reasons:

1. AB may be defined and BA may not (for example, if A is 2×3 and B is 3×4).

2. AB and BA may both be defined, but they may have different sizes (for example, if A is 2×3 and B is 3×2).

3. AB and BA may both be defined and have the same size, but the two products may be different.

A matrix whose entries are all zero is called a **zero matrix**.

Properties of Zero Matrices

If c is a scalar, and if the sizes of the matrices are such that the operations can be perfomed, then:

(a) $A + 0 = 0 + A = A$

(b) $A - 0 = A$

(c) $A - A = A + (-A) = 0$

(d) $0A = 0$

(e) If $cA = 0$, then $c = 0$ or $A = 0$.

Since we know that the commutative law of real arithmetic is not valid in matrix arithmetic, it should not be surprising that there are other rules that fail as well. For example, consider the following two laws of real arithmetic:

- If $ab = ac$ and $a \neq 0$, then $b = c$. [The cancellation law]
- If $ab = 0$, then at least one of the factors on the left is 0.

The next two examples show that these laws are not true in matrix arithmetic.

Consider the matrices

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}, \ B = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}, \ C = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix}.$$

Here, $AB = AC = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}$.

Although A $\neq$ 0, canceling A from both sides of the equation AB = AC would lead to the incorrect conclusion that B = C. Thus, the cancellation law does not hold, in general, for matrix multiplication (though there may be particular cases where it is true).

Consider the example of two matrices$A$ and B for which $AB = 0$, but $A \neq 0$ and $B \neq 0$:

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}, \ B = \begin{bmatrix} 3 & 7 \\ 0 & 0 \end{bmatrix}.$$

A square matrix with 1's on the main diagonal and zeros elsewhere is called an **identity matrix**. An identity matrix is denoted by the letter I.

For example: $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

If it is important to emphasize the size, we will write I_n for the n × n identity matrix.

If A is any m × n matrix, then $AI_n = A$ and $I_m A = A$.

1.2 Determinant

For every square matrix $A = [a_{ij}]$, we associate a number called **determinant** of A, denoted by $|A|$ or det(A).

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

If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $det(A) = a_{11}a_{22} - a_{21}a_{12}$.

If $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, then

$$det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

If A is a square matrix, then the **minor** of entry a_{ij} is denoted by M_{ij} and is defined to be the determinant of the submatrix that remains after the ith row and j th column are deleted from A. The number $(-1)^{i+j}M_{ij}$ is denoted by C_{ij} and is called the **cofactor** of entry a_{ij} .

Example. Consider the matrix $A = \begin{bmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{bmatrix}$.

The minor of $a_{11} = M_{11} = \begin{vmatrix} 5 & 6 \\ 4 & 8 \end{vmatrix} = 16.$

The cofactor of $a_{11} = C_{11} = (-1)^{1+1}M_{11} = M_{11} = 16.$

The minor of $a_{12} = M_{12} = \begin{vmatrix} 2 & 6 \\ 1 & 8 \end{vmatrix} = 10.$

The cofactor of $a_{12} = C_{12} = (-1)^{1+2}M_{12} = -M_{12} = -10.$

Similarly, we can find for all other entries of A.

If A is an n × n matrix, then the number obtained by multiplying the entries in any row or column of A by the corresponding cofactors and adding the resulting products is called the **determinant** of A, and the sums themselves are called **cofactor expansions** of A. That is,

$det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$, cofactor expansion along the jth column

$det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}$, cofactor expansion along the ith row.

Example. Find the determinant of the matrix $A = \begin{bmatrix} 3 & 1 & 0 \\ -2 & -4 & 3 \\ 5 & 4 & -2 \end{bmatrix}$ by cofactor expansion along the first row.

Solution. $det(A) = \begin{vmatrix} 3 & 1 & 0 \\ -2 & -4 & 3 \\ 5 & 4 & -2 \end{vmatrix} = 3 \begin{vmatrix} -4 & 3 \\ 4 & -2 \end{vmatrix} - 1 \begin{vmatrix} -2 & 3 \\ 5 & -2 \end{vmatrix} + 0 \begin{vmatrix} -2 & -4 \\ 5 & 4 \end{vmatrix}$
$= 3(-4) - (1)(-11) + 0 = -1.$

Theorem 1.1. Let A be a square matrix. If A has a row of zeros or a column of zeros, then det(A) = 0.

Proof. Since the determinant of A can be found by a cofactor expansion along any row or column, we can use the row or column of zeros. Thus, if we let C_1, C_2,...,C_n denote the cofactors of A along that row or column, then it follows from cofactor expansion that $\det(A) = 0 \cdot C_1 + 0 \cdot C_2 + \cdots + 0 \cdot C_n = 0$.

Theorem 1.2. Let A be a square matrix. Then $\det(A) = \det(A^T)$.

Proof. Since transposing a matrix changes its columns to rows and its rows to columns, the cofactor expansion of A along any row is the same as the cofactor expansion of A^T along the corresponding column. Thus, both have the same determinant.

Theorem 1.3. Let A be an n × n matrix.

(a) If B is the matrix that results when a single row or single column of A is multiplied by a scalar k, then

$\det(B) = k \det(A)$.

(b) If B is the matrix that results when two rows or two columns of A are interchanged, then $\det(B) = -\det(A)$.

(c) If B is the matrix that results when a multiple of one row of A is added to another or when a multiple of one column is added to another, then $\det(B) = \det(A)$.

Theorem 1.4. If A is a square matrix with two proportional rows or two proportional columns, then $\det(A) = 0$.

Example matrices having two proportional rows or columns;

thus, each has a determinant of zero.

$$\begin{bmatrix} -1 & 4 \\ -2 & 8 \end{bmatrix}, \quad \begin{bmatrix} 1 & -2 & 7 \\ -4 & 8 & 5 \\ 2 & -4 & 3 \end{bmatrix}$$

1.3 Inverse of a matrix

A square matrix whose determinant is equal to zero is called a **singular matrix**.

A square matrix whose determinant is not equal to zero is called a **non-singular matrix**.

Consider $A = \begin{bmatrix} 5 & 10 \\ 3 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 4 \\ 5 & 8 \end{bmatrix}$. Now, $\det(A)$ or $|A| = 0$ and $\det(B)$ or $|B| \neq 0$. So A is singular and B is non-singular.

If A is a square matrix, and if a matrix B of the same size can be found such that AB = BA = I , then A is said to be **invertible (or nonsingular)** and B is called an **inverse** of A. If no such matrix B can be found, then A is said to be **singular**.

The relationship AB = BA = I is not changed by interchanging

A and B, so if A is invertible and B is an inverse of A, then it is also true that B is invertible, and A is an inverse of B. Thus, when AB = BA = I we say that A and B are inverses of one another.

Consider the matrices $A = \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}$. A and B are inverses of each other. One can check it by $AB = BA = I$.

Theorem 1.5 If B and C are both inverses of the matrix A, then B = C.

Proof. Since B is an inverse of A, we have BA = I. Multiplying both sides on the right by C gives (BA)C = IC = C. But it is also true that (BA)C = B(AC) = BI = B, so C = B.

Remark 1: If A is invertible, then its inverse will be denoted by the symbol A^{-1}. Then, $AA^{-1} = I = A^{-1}A$.

Formula for finding the inverse:

Let A be a square matrix. Then $A^{-1} = \dfrac{adj A}{det(A)}$.

Theorem 1.6. If A and B are invertible matrices with the same size, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. We can establish the invertibility and obtain the stated formula at the same time by showing that $(AB)(B^{-1}A^{-1}) = (B^{-1}A^{-1})(AB) = I$. But $(AB)(B^{-1}A^{-1}) =$

$A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$ and similarly, $(B^{-1}A^{-1})(AB) = I$.

Remark 2: A product of any number of invertible matrices is invertible, and the inverse of the product is the product of the inverses in the reverse order.

Some properties:

If A is invertible and n is a nonnegative integer, then:

$A^0 = I$ and $A^n = AA \ldots A[n factors]$

$A^{-n} = (A^{-1})n = A^{-1}A^{-1} \ldots A^{-1}[n factors]$

$A^rA^s = A^{r+s}$ and $(A^r)^s = A^{rs}$

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

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

kA is invertible for any nonzero scalar k, and $(kA)^{-1} = k^{-1}A^{-1}$.

Properties of the transpose:

If the sizes of the matrices are such that the stated operations can be performed, then:

$(A^T)^T = A$

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

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

$(kA)^T = kA^T$

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

Remark 3: The transpose of a product of any number of matrices is the product of the transposes in the reverse order.

Theorem 1.7. If A is an invertible matrix, then A^T is also invertible and $(A^T)^{-1} = (A^{-1})^T$.

Proof. We can establish the invertibility and obtain the formula at the same time by showing that $A^T(A^{-1})^T = (A^{-1})^T A^T = I$

But from the facts that $(AB)^T = B^T A^T$ and $I^T = I$, we have

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

$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$ which completes the proof.

Elementary row operations:

1. Multiply a row by a nonzero constant c.
2. Interchange two rows.
3. Add a constant c times one row to another.

It should be evident that if we let B be the matrix that results from A by performing one of the operations in this list, then the matrix A can be recovered from B by performing the corresponding operation in the following list:

1. Multiply the same row by 1/c.

2. Interchange the same two rows.

3. If B resulted by adding c times row r_i of A to row r_j , then add -c times r_j to r_i.

It follows that if B is obtained from A by performing a sequence of elementary row operations, then there is a second sequence of elementary row operations, which when applied to B recovers A.

Matrices A and B are said to be **row equivalent** if either (hence each) can be obtained from the other by a sequence of elementary row operations.

A matrix E is called an **elementary** matrix if it can be obtained from an identity matrix by performing a single elementary row operation.

Consider $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Multiplying the second row of I_2 by -3 we get an elementary matrix $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$.

Interchaning second and fourth rows of I_4 yields an elementary matrix $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

Theorem 1.8.*Row Operations by Matrix Multiplication*: If the elementary matrix E results from performing a certain row operation on I_m and if A is an m $\times$ n matrix, then the product EA is the matrix that results when this same row operation is performed on A.

Theorem 1.9. Every elementary matrix is invertible, and the inverse is also an elementary matrix.

Proof. If E is an elementary matrix, then E results by performing some row operation on I . Let E_0 be the matrix that results when the inverse of this operation is performed on I . Applying the theorem 1.8. and using the fact that inverse row operations cancel the effect of each other, it follows that $E_0E = I$ and $EE_0 = I$. Thus, the elementary matrix E_0 is the inverse of E.

1.4 Diagonal and triangular matrices

A square matrix in which all the entries off the main diagonal are zero is called a **diagonal matrix**.

Example. $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$, $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

A general n $\times$ n diagonal matrix D can be written as

$$\begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_n \end{bmatrix} \qquad (1).$$

A diagonal matrix is invertible if and only if all of its diagonal entries are nonzero; in this case the inverse of (1) is

$$\begin{bmatrix} 1/d_1 & 0 & \dots & 0 \\ 0 & 1/d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1/d_n \end{bmatrix} = D^{-1} \qquad (2).$$

If D is the diagonal matrix (1) and k is a positive integer, then

$$\begin{bmatrix} d_1^k & 0 & \dots & 0 \\ 0 & d_2^k & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_n^k \end{bmatrix} = D^k \qquad (3).$$

Example. If $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, then

$A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/3 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$, $A^5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -243 & 0 \\ 0 & 0 & 32 \end{bmatrix}$, $A^{-5} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/243 & 0 \\ 0 & 0 & 1/32 \end{bmatrix}$

A square matrix in which all the entries above the main diagonal are zero is called **lower triangular**, and a square matrix in which all the entries below the main diagonal are zero is called **upper triangular**. A matrix that is either upper triangular or lower triangular is called **triangular**.

Observe that diagonal matrices are both upper triangular and lower triangular since they have zeros below and above the main diagonal. Observe also that a square matrix in row echelon form is upper triangular since it has zeros below the main diagonal.

Theorem 1.10.(a) The transpose of a lower triangular matrix is upper triangular, and the transpose of an upper triangular matrix is lower triangular.

(b) The product of lower triangular matrices is lower triangular, and the product of upper triangular matrices is upper triangular.

(c) A triangular matrix is invertible if and only if its diagonal entries are all nonzero.

(d) The inverse of an invertible lower triangular matrix is lower triangular, and the inverse of an invertible upper triangular matrix is upper triangular.

Example. Consider the upper triangular matrices

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 0 & 2 & 4 \\ 0 & 0 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

It follows from part (c) of the above theorem that the matrix A is invertible but the matrix B is not. Moreover, the theorem also tells us that A^{-1}, AB, and BA must be upper triangular.

Also, $A^{-1} = \begin{bmatrix} 1 & -3/2 & 7/5 \\ 0 & 1/2 & -2/5 \\ 0 & 0 & 1/5 \end{bmatrix}$, $AB = \begin{bmatrix} 3 & -2 & -2 \\ 0 & 0 & 2 \\ 0 & 0 & 5 \end{bmatrix}$, $BA = \begin{bmatrix} 3 & 5 & -1 \\ 0 & 0 & -5 \\ 0 & 0 & 5 \end{bmatrix}$

Remark. If A is an n $\times$ n triangular matrix (upper triangular, lower triangular, or diagonal), then $det(A)$ is the product of the entries on the main diagonal of the matrix; that is, $det(A) = a_{11}a_{22}\dots a_{nn}$.

A square matrix A is said to be **symmetric** if A $=$ A^T .

Example. $\begin{bmatrix} 1 & 4 & 5 \\ 4 & -3 & 0 \\ 5 & 0 & 7 \end{bmatrix}$, $\begin{bmatrix} 7 & -3 \\ -3 & -5 \end{bmatrix}$

All diagonal matrices are symmetric.

Algebraic properties of symmetric matrices

If A and B are symmetric matrices with the same size, and if k is any scalar, then:

(a) A^T is symmetric.

(b) A + B and A - B are symmetric.

(c) kA is symmetric.

Theorem 1.11. The product of two symmetric matrices is symmetric if and only if the matrices commute.

Proof. Let A and B be symmetric matrices with the same size. Then it follows from the result $(AB)^T = B^T A^T$ and the symmetry of A and B that

$(AB)^T = B^T A^T = BA.$

Thus, $(AB)^T = AB$ if and only if $AB = BA$, that is, if and only if A and B commute.

In general, a symmetric matrix need not be invertible. For example, a diagonal matrix with a zero on the main diagonal is symmetric but not invertible. But if a symmetric matrix happens to be invertible, then its inverse must also be symmetric.

Theorem 1.12. If A is an invertible symmetric matrix, then A^{-1} is symmetric.

Proof Assume that A is symmetric and invertible. From the facts that

$(A^T)^{-1} = (A^{-1})^T$ and $A = A^T$, we have

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

which proves that A^{-1} is symmetric.

Remark. Matrix products of the form AA^T and A^TA arise in a variety of applications. If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix, so the products AA^T and A^TA are both square matrices—the matrix AA^T has size $m \times m$, and the matrix A^TA has size $n \times n$. Such products are always symmetric since $(AA^T)^T = (A^T)^TA^T = AA^T$ and $(A^TA)^T = A^T(A^T)^T = A^TA$.

Example. One can verify that AA^T and A^TA are symmetric for the matrix

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

Theorem 1.13. If A is an invertible matrix, then AA^T and A^TA are also invertible.

Proof Since A is invertible, so is A^T by Theorem 1.7. Thus AA^T and A^TA are invertible, since they are the products of invertible matrices.

1.5 Introduction to system of linear equations

In two dimensions a line in a rectangular xy-coordinate system can be represented by an equation of the form $ax + by = c(a, b \quad not \ \ both \ \ 0)$ and in three dimensions a plane in a rectangular xyz-coordinate system can be represented by an equation of the form $ax + by + cz = d(a, b, c \quad not \ \ all \ \ 0)$. These are examples of "linear equations," the first being a linear equation in the variables x and y and the second a linear equation in the variables x, y, and z.

More generally, we define a **linear equation** in the n variables $x_1, x_2, \ldots, x_n$ to be one that can be expressed in the form $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$ (1) where $a_1, a_2, \ldots, a_n$ and b are constants, and the $a's$ are not all zero. In the special cases where $n = 2$ or $n = 3$, we will often use variables without subscripts and write linear equations as $a_1x + a_2y = b(a_1, a_2 \quad not \ \ both \ \ 0)$ $a_1x + a_2y + a_3z = b(a_1, a_2, a_3 \quad not \quad all \quad 0)$. In the special case where $b = 0$, Equation (1) has the form $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$ which is called a **homogeneous**

linear equation in the variables $x_1, x_2, \ldots, x_n$.

A finite set of linear equations is called a **system of linear equations** or, more briefly, a **linear system**. The variables are called **unknowns**. A general linear system of m equations in the n unknowns $x_1, x_2, \ldots, x_n$ can be written as

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

The above system of equations can be expressed as a single matrix equation $AX = B$ (3), where

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Any column vector X which satisfies the matrix equation (3) is called the **solution** of the system. That is X is the column vector consisting of the set of values of $x_1, x_2, \ldots, x_n$ which satisfy simultaneously the m equations in the system (2). When $B \neq 0$, the system is said to be **non-homogeneous**. When $B = 0$, the system is said to be **homogeneous**.

A linear system may have exactly one solution, an infinite number of solutions or no solution at all. Systems that have one or more solutions are called **consistent system** and systems that do not have a solution are called **inconsistent systems**.

1.5.1 Linear system in two and three unknowns

Linear systems in two unknowns arise in connection with intersections of lines. For example, consider the linear system $a_1x + b_1y = c_1, \quad a_2x + b_2y = c_2$ in which the graphs of the equations are lines in the xy-plane. Each solution (x, y) of this system corresponds to a point of intersection of the lines, so there are three possibilities:

1. The lines may be parallel and distinct, in which case there is no intersection and consequently no solution.

2. The lines may intersect at only one point, in which case the system has exactly one solution.

3. The lines may coincide, in which case there are infinitely

many points of intersection (the points on the common line) and consequently infinitely many solutions.

Thus, a consistent linear system of two equations in two unknowns has either one solution or infinitely many solutions—there are no other possibilities. The same is true for a linear system of three equations in three unknowns

$a_1x + b_1y + c_1z = d_1, \quad a_2x + b_2y + c_2z = d_2, \quad a_3x + b_3y + c_3z = d_3$

in which the graphs of the equations are planes. The solutions of the system, if any, correspond to points where all three planes intersect, so again we see that there are only three possibilities—no solutions, one solution, or infinitely many solutions.

Every system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.

Example 1. Solve the linear system

$x - y = 1$

$2x + y = 6.$

Solution. We can eliminate x from the second equation by using the first equation. This gives the system

$x - y = 1$

$3y = 4.$

From the second equation we obtain $y = 4/3$ and on substituting this value in the first equation we obtain $x = 1 + y = 7/3$. Thus, the system has the unique solution $x = 7/3, y = 4/3$.

Geometrically, this means that the lines represented by the equations in the system intersect at the single point $(7/3, 4/3)$.

Example 2. Solve the linear system

$x + y = 4$

$3x + 3y = 6.$

Solution. We can eliminate x from the second equation by using the first equation. This yields the simplified system

$x + y = 4$

$0 = -6.$

The second equation is contradictory, so the given system has no solution.

Geometrically, this means that the lines corresponding to the equations in the original system are parallel and distinct.

Example 3. Solve the linear system

$4x - 2y = 1$

$16x - 8y = 4.$

Solution. We can eliminate x from the second equation by using the first equation. This yields the system

$4x - 2y = 1$

$0 = 0.$

The second equation does not impose any restrictions on x and y and hence can be omitted. Thus, the solutions of the system are those values of x and y that satisfy the single equation

$4x - 2y = 1.$

Geometrically, this means the lines corresponding to the two equations in the original system coincide.

One way to describe the solution set is to solve this equation for x in terms of y to obtain $x = (1 + 2y)/4$, i.e; $x = \frac{1}{4} + \frac{1}{2}y$ and then assign an arbitrary value t **(called a parameter)** to y. Thus, we obtain $x = \frac{1}{4} + \frac{1}{2}t, \quad y = t$. This gives infinite number of solutions for the given system.

Example 4. Solve the linear system

$x - y + 2z = 5$

$2x - 2y + 4z = 10$

$3x - 3y + 6z = 15$

Solution. This system can be solved by inspection, since the second and third equations are multiples of the first. Geometrically, this means that the three planes coincide and that those values of x, y, and z that satisfy the equation $x - y + 2z = 5$ automatically satisfy all three equations. Thus, it is enough to find the solutions of $x - y + 2z = 5$.

We can do this by first solving this equation for x in terms of y and z, then assigning arbitrary values r and s (parameters) to y and z respectively, and then expressing the solution by the three parametric equations

$x = 5 + r - 2s, \quad y = r, \quad z = s.$

Specific solutions can be obtained by choosing numerical values for the parameters r and s. For example, taking $r = 1$ and $s = 0$ yields the solution $(6, 1, 0)$.

1.5.2 Non-homogeneous system

Consider the following m equations in n unknowns.

$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1$

$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2$

$\ldots \qquad \ldots \qquad \ldots \qquad \ldots$

$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m.$

The above system of equations can be expressed as a single matrix equation $AX = B$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

The matrix $[AB]$, obtained by placing the column matrix B to the right of the matrix A is called the **augmented matrix** for the system.

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

The algebraic operations to perform to solve the system are:

1. Multiply an equation through by a nonzero constant.
2. Interchange two equations.
3. Add a constant times one equation to another.

Correspondingly these are the following operations to perform to solve the system on the rows of the augmented matrix:

1. Multiply a row through by a nonzero constant.

2. Interchange two rows.

3. Add a constant times one row to another.

These are called **elementary row operations** on a matrix.

Row reduced echelon form

By a sequence of elementary row operations, the augumented matrix of a linear system can be transformed to a matrix that have the following properties:

1. If a row does not consist entirely of zeros, then the first nonzero number in the row is a 1. We call this a leading 1.

2. If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.

3. In any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row.

4. Each column that contains a leading 1 has zeros everywhere else in that column.

A matrix that has the first three properties is said to be in **row reduced echelon form**.

Algorithm for reducing to row echelon form

Step 1. Locate the leftmost column that does not consist entirely of zeros.

Step 2. Interchange the top row with another row, if necessary, to bring a nonzero entry to the top of the column found in Step 1.

Step 3. If the entry that is now at the top of the column found in Step 1 is a, multiply the first row by 1/a in order to introduce a leading 1.

Step 4. Add suitable multiples of the top row to the rows below so that all entries below the leading 1 become zeros.

Step 5. Now cover the top row in the matrix and begin again with Step 1 applied to the submatrix that remains. Continue in this way until the entire matrix is in row echelon form.

Step 6. Beginning with the last nonzero row and working upward, add suitable multiples of each row to the rows above to introduce zeros above the leading 1's.

Example 5. Reduce the matrix $A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & -2 \\ 4 & -3 & -1 \\ 2 & 4 & 2 \end{bmatrix}$ to echelon form.

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

$R_2 \to R_2 - 3R_1, \quad R_3 \to R_3 - 4R_1, \quad R_4 \to R_4 - 2R_1$

$$\cong \begin{bmatrix} 1 & 2 & 1 \\ 0 & -5 & -5 \\ 0 & -11 & -5 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_2 \to \frac{-1}{5} R_2$

$$\cong \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & -11 & -5 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_3 \to R_3 + 11R_2$

$$\cong \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_3 \to \frac{1}{6} R_3$

$$\cong \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

1.6 Gauss Elimination method for solving a system of linear equations

Example 6. Solve the below system by Gauss elimination method.

$x + 2y + z = 2$

$3x + y - 2z = 1$

$4x - 3y - z = 3$

$2x + 4y + 2z = 4.$

Solution. Here $A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & -2 \\ 4 & -3 & -1 \\ 2 & 4 & 2 \end{bmatrix}, B = \begin{bmatrix} 2 \\ 1 \\ 3 \\ 4 \end{bmatrix}.$

So augmented matrix $[AB] = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 3 & 1 & -2 & 1 \\ 4 & -3 & -1 & 3 \\ 2 & 4 & 2 & 4 \end{bmatrix}$.

Reduce the $[AB]$ to row echelon form.

That gives $[AB] = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. (do as in example 5.)

The corresponding system of equations in matrix form is

$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & -2 \\ 4 & -3 & -1 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}$, which is equivalent to the equations $x + 2y + z = 2, y + z = 1, z = 1$.

By back substitution, we get the solution as $x = 1, y = 0, z = 1$.

Example 7. Solve the below system by Gauss elimination method.

$-2x_3 + 7x_5 = 12$

$2x_1 + 4x_2 - 10x_3 + 6x_4 + 12x_5 = 28$

$2x_1 + 4x_2 - 5x_3 + 6x_4 - 5x_5 = -1.$

Solution. Here the augmented matrix is

$$[AB] = \begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix} \cong \begin{bmatrix} 2 & 4 & -10 & 6 & 12 & 28 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix} \cong$$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix} \cong \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & -2 & 0 & 7 & 12 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix} \cong$$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix} \cong \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 5 & 0 & -17 & -29 \end{bmatrix} \cong$$

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 0 & 0 & 1/2 & 1 \end{bmatrix} \cong \begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -7/2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix},$$

which is now in echelon form.

Corresponding matrix form is

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 \\ 0 & 0 & 1 & 0 & -7/2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 14 \\ -6 \\ 2 \end{bmatrix}$$ which is equivalent to

$$x_1 + 2x_2 - 5x_3 + 3x_4 + 6x_5 = 14$$

$x_3 - \frac{7}{2}x_5 = -6$

$x_5 = 2$. By back substitution, $x_5 = 2, x_3 = 1$.

Assigning $x_2 = r, x_4 = s$, the solution is

$x_1 = 7 - 2r - 3s, x_2 = r, x_3 = 1, x_4 = s, x_5 = 2$.

Example 8. Solve the below system by Gauss Jordan elimination method.

$x_1 + 3x_2 - 2x_3 + 2x_5 = 0$

$2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 = -1$

$5x_3 + 10x_4 + 15x_6 = 5$

$2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 = 6$

Solution. Here the augmented matrix is

$$[AB] = \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 2 & 6 & 0 & 8 & 4 & 18 & 6 \end{bmatrix} \cong \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & -2 & 0 & -3 & -1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 0 & 0 & 4 & 8 & 0 & 18 & 6 \end{bmatrix} \cong$$

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 5 & 10 & 0 & 15 & 5 \\ 0 & 0 & 4 & 8 & 0 & 18 & 6 \end{bmatrix} \cong \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \end{bmatrix} \cong$$

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cong \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cong$$

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cong \begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$, which is now in echelon form.

Corresponding matrix form is

$$\begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1/3 \\ 0 \end{bmatrix}$$ which is equivalent to

$x_1 + 3x_2 + 4x_4 + 2x_5 = 0$

$x_3 + 2x_4 = 0$

$x_6 = 1/3.$

Solving , we obtain $x_1 = -3x_2 - 4x_4 - 2x_5 x_3 = -2x_4 x_6 = 1/3$

Finally, by assigning the free variables x_2, x_4, x_5 arbitrary values r, s, t, respectively, we get

$x_1 = -3r - 4s - 2t, x_2 = r, x_3 = -2s, x_4 = s, x_5 = t, x_6 = 1.$

Example 9. Solve the below system by Gauss Jordan elimination method.

$x_1 + 2x_2 - 3x_3 - 4x_4 = 6$

$x_1 + 3x_2 + x_3 - 2x_4 = 4$

$2x_1 + 5x_2 - 2x_3 - 5x_4 = 10$

Solution. Here the augmented matrix is

$[AB] = \begin{bmatrix} 1 & 2 & -3 & -4 & 6 \\ 1 & 3 & 1 & -2 & 4 \\ 2 & 5 & -2 & -5 & 10 \end{bmatrix}$. The row reduced echelon form is

$[AB] \cong \begin{bmatrix} 1 & 0 & -11 & 0 & 10 \\ 0 & 1 & 4 & 0 & -2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$.

Corresponding matrix form is

$\begin{bmatrix} 1 & 0 & -11 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 10 \\ -2 \\ 0 \end{bmatrix}$ which is equivalent to

$x_1 - 11x_3 = 10$

$x_2 + 4x_3 = -2$

$x_4 = 0$.

Assinging $x_3 = t$, we get the general solution as

$x_1 = 11t + 10, x_2 = -4t - 2, x_3 = t, x_4 = 0$.

Example 10. Suppose that the matrices below are augmented matrices for linear systems in the unknowns x_1, x_2, x_3, x_4. Discuss the existence and uniqueness of solutions to the corresponding linear systems.

$$A = \begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

1.7 Homogeneous system of linear equations

A system of linear equations is said to be **homogeneous** if the constant terms are all zero; that is, the system has the form

$$
\begin{array}{llll}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0 \\
\cdots \quad \cdots \quad \cdots \quad \cdots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0.
\end{array} \qquad (2)
$$

Here $B = 0$, so the matrix A and the augmented matrix $[AB]$ are the same. Every homogeneous system of linear equations is consistent because all such systems have $x_1 = 0, x_2 = 0, ..., x_n = 0$ as a solution. This solution is called the **trivial** solution; if there are other solutions, they are called **nontrivial** solutions.

Because a homogeneous linear system always has the trivial solution, there are only two possibilities for its solutions:

- The system has only the trivial solution.
- The system has infinitely many solutions in addition to the trivial solution.

Theorem 1.14.Free Variable Theorem for Homogeneous Systems If a homogeneous linear system has n unknowns, and if the reduced row echelon form of its augmented matrix has r nonzero rows, then the system has n - r free variables.

Theorem. A homogeneous linear system with more unknowns than equations has infinitely many solutions.

Proof. If a homogeneous linear system has m equations in n unknowns, and if m < n, then it must also be true that r < n. This being the case, the theorem implies that there is at least one free variable, and this implies that the system has infinitely many solutions.

Example 11. Solve the below system .

$x_1 + 3x_2 - 2x_3 + 2x_5 = 0$

$2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 = 0$

$5x_3 + 10x_4 + 15x_6 = 0$

$2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 = 0$

Solution. Here the augmented matrix is

$[AB] = \begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 & 0 \\ 0 & 0 & 5 & 10 & 0 & 15 & 0 \\ 2 & 6 & 0 & 8 & 4 & 18 & 0 \end{bmatrix}$. Continuing as in example 8., we get the echelon form of this matrix as $\begin{bmatrix} 1 & 3 & 0 & 4 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. Corresponding system of equations is

$x_1 + 3x_2 + 4x_4 + 2x_5 = 0$

$x_3 + 2x_4 = 0$

$x_6 = 0.$

Solving , we obtain $x_1 = -3x_2 - 4x_4 - 2x_5 x_3 = -2x_4 x_6 = 0$. Finally, by assigning the free variables x_2, x_4, x_5 arbitrary values r, s, t, respectively, we get $x_1 = -3r - 4s - 2t, x_2 = r, x_3 = -2s, x_4 = s, x_5 = t, x_6 = 0$.

Example 12. Solve the below system .

$x_1 - 2x_2 + x_3 - x_4 = 0$

$x_1 + x_2 - 2x_3 + 3x_4 = 0$

$4x_1 + x_2 - 5x_3 + 8x_4 = 0$

$5x_1 - 7x_2 + 2x_3 - x_4 = 0$

Solution. After reduucing to echelon form the corresponding system of equations in the matrix form will be

$$\begin{bmatrix} 1 & 0 & -1 & 5/3 \\ 0 & 1 & -1 & 4/3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving this and by substituting $x_3 = r, x_4 = s$, we get infinite number of solutions.

The general solution is $x_1 = r - \frac{5}{3}s, x_2 = r - \frac{4}{3}s, x_3 = r, x_4 = s.$

1.8 More on linear systems

Theorem 1.15. A system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.

Proof. If Ax = b is a system of linear equations, exactly one of the following is true:

(a) the system has no solutions, (b) the system has exactly one solution, or (c) the system has more than one solution. The proof will be complete if we can show that the system has infinitely many solutions in case (c).

Assume that $Ax = b$ has more than one solution, and let $x_0 = x_1 - x_2$, where x_1 and x_2 are any two distinct solutions. Because x_1 and x_2 are distinct, the matrix x_0 is nonzero.

Moreover, $Ax_0 = A(x_1 - x_2) = Ax_1 - Ax_2 = b - b = 0$.

If we now let k be any scalar, then

$A(x_1 + kx_0) = Ax_1 + A(kx_0) = Ax_1 + k(Ax_0) = b + k0 = b + 0 = b$.

But this says that $x_1 + kx_0$ is a solution of $Ax = b$. Since x_0 is nonzero and there are infinitely many choices for k, the system $Ax = b$ has infinitely many solutions.

Solving Linear Systems by Matrix Inversion

Theorem 1.16. If A is an invertible $n \times n$ matrix, then for each $n \times 1$ matrix b, the system of equations $Ax = b$ has exactly one solution, namely, $x = A^{-1}b$.

Proof. Since $A(A^{-1}b) = b$, it follows that $x = A^{-1}b$ is a solution of $Ax = b$.

To show that this is the only solution, we will assume that x_0 is an arbitrary solution and then show that x_0 must be the solution $A^{-1}b$.

If x_0 is any solution of $Ax = b$, then $Ax_0 = b$. Multiplying both sides of this equation by A^{-1}, we obtain $x_0 = A^{-1}b$.

Theorem 1.17. *Equivalent Statements*

If A is an n × n matrix, then the following statements are equivalent:

(a) A is invertible.

(b) $Ax = 0$ has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

Proof.

(a) ⇒ (b) Assume A is invertible and let x_0 be any solution of $Ax = 0$. Multiplying both sides of this equation by the matrix A^{-1} gives $A^{-1}(Ax_0) = A^{-1}0$, or $(A^{-1}A)x_0 = 0$, or $Ix_0 = 0$, or $x_0 = 0$. Thus, $Ax = 0$ has only the trivial solution.

(b) ⇒ (c) Let $Ax = 0$ be the matrix form of the system

$$\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = 0. \end{array} \qquad (1)$$

and assume that the system has only the trivial solution. If we solve by Gauss–Jordan elimination, then the system of equations corresponding to the reduced row echelon form of the augmented matrix will be

$$\begin{aligned} x_1 \quad\quad\quad\quad &= 0 \\ x_2 \quad\quad\quad &= 0 \\ \ddots \quad\quad & \qquad\qquad\qquad\qquad (2) \\ x_n &= 0. \end{aligned}$$

Thus the augmented matrix $\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 \end{bmatrix}$ for (1) can be reduced to the augmented matrix $\begin{bmatrix} 1 & a_{12} & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$ for (2) by a sequence of elementary row operations. If we disregard the last column (all zeros) in each of these matrices, we can conclude that the reduced row echelon form of A is I_n.

(c) $\Rightarrow$ (d) Assume that the reduced row echelon form of A is I_n, so that A can be reduced to I_n by a finite sequence of elementary row operations. By theorem 1.8, each of these operations can be accomplished by multiplying on the left by an appropriate elementary matrix. Thus we can find elementary matrices E_1, E_2,...,E_k such that

$E_k \cdots E_2 E_1 A = I_n$ (3).

By theorem 1.9, E_1, E_2,...,E_k are invertible. Multiplying both sides of Equation (3) on the left successively by E_k^{-1} ,...,E_2^{-1} , E_1^{-1} we obtain

$A = E_1^{-1}\, E_2^{-1} \cdots E_k^{-1}\, I_n = E_1^{-1}\, E_2^{-1} \cdots E_k^{-1}$ (4).

By theorem 1.9, this equation expresses A as a product of elementary matrices.

(d)$\Rightarrow$ (a) If A is a product of elementary matrices, then from *the remark 2, properties and theorem 1.9*, the matrix A is a product of invertible matrices and hence is invertible.

A Method for Inverting Matrices

An application of Theorem 1.17.

Inversion Algorithm:

To find the inverse of an invertible matrix A, find a sequence of elementary row operations that reduces A to the identity and then perform that same sequence of operations on In to obtain A^{-1}.

Example: Find the inverse of $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$.

We want to reduce A to the identity matrix by row operations and simultaneously apply these operations to I to produce A^{-1}. To accomplish this we will adjoin the identity matrix to the right side of A, thereby producing a partitioned matrix of the form $[A|I]$. Then we will apply row operations to this matrix until the left side is reduced to I ; these operations will convert the right side to A^{-1}, so the final matrix will have the form $[I|A^{-1}]$. Now the computation:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{array}\right]$$

We added -2 times the first row to the second and -1 times the first row to the third.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{array}\right]$$

We added 2 times the second row to the third.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$$

We multiplied the third row by -1.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 0 & -14 & 6 & 3 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$$

We added 3 times the third row to the second and -3 times the third row to the first.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -40 & 16 & 9 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right]$$

We added -2 times the second row to the first.

Thus, $A^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}$.

Example. Consider the system of linear equations

$x_1 + 2x_2 + 3x_3 = 5$

$2x_1 + 5x_2 + 3x_3 = 3$

$x_1 + 8x_3 = 17$

In matrix form this system can be written as Ax = b, where

$$A = \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{array}\right], X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, B = \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix}$$

In the above example, we showed that A is invertible and

$$A^{-1} = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix}.$$

Thus, the solution of the system is

$$x = A^{-1}b = \begin{bmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 17 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}.$$

Hence, $x_1 = 1, x_2 = -1, x_3 = 2.$

Example: Show that the given below matrix is not invertible.

$$\begin{bmatrix} 1 & 6 & 4 \\ 2 & 4 & -1 \\ 1 & 2 & 5 \end{bmatrix}.$$

Proceeding

$$\left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 2 & 4 & -1 & 0 & 1 & 0 \\ 1 & 2 & 5 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 8 & 9 & 1 & 0 & 1 \end{array}\right]$$

We added -2 times the first row to the second and added the first row to the third.

$$\left[\begin{array}{ccc|ccc} 1 & 6 & 4 & 1 & 0 & 0 \\ 0 & -8 & -9 & -2 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{array}\right]$$

We added the second row to the third.

Since we have obtained a row of zeros on the left side, A is not invertible.

Example: Use Theorem 1.17. to determine whether the given homogeneous systems have nontrivial solutions.

(a) $x_1 + 2x_2 + 3x_3 = 0$

$2x_1 + 5x_2 + 3x_3 = 0$

$x_1 + 8x_3 = 0$

(b) $x_1 + 6x_2 + 4x_3 = 0$

$2x_1 + 4x_2 - x_3 = 0$

$-x_1 + 2x_2 + 5x_3 = 0$

Solution. From parts (a) and (b) of Theorem 1.17. a homogeneous linear system has only the trivial solution if and only if its coefficient matrix is invertible. From the above examples the coefficient matrix of system (a) is invertible and that of system (b) is not. Thus, system (a) has only the trivial solution while system (b) has nontrivial solutions.

Linear Systems with a Common Coefficient Matrix

One is concerned with solving a sequence of systems

$Ax = b_1, Ax = b_2, Ax = b_3, ..., Ax = b_k$

each of which has the same square coefficient matrix A. If A is invertible, then the solutions $x_1 = A^{-1}b_1, x_2 = A^{-1}b_2, x_3 = A^{-1}b_3, ..., x_k = A^{-1}b_k$

can be obtained with one matrix inversion and k matrix multiplications. An efficient way to do this is to form the partitioned matrix

$$\begin{bmatrix} A & | & b_1 & | & b_2 & | & \ldots & b_k \end{bmatrix} \qquad (1)$$

in which the coefficient matrix A is "augmented" by all k of the matrices $b_1, b_2, ..., b_k$, and then reduce (1) to reduced row echelon form by Gauss–Jordan elimination. In this way we can solve all k systems at once. This method has the added advantage that it applies even when A is not invertible.

Example. Solve the systems

(a) $x_1 + 2x_2 + 3x_3 = 4 \qquad 2x_1 + 5x_2 + 3x_3 = 5 \qquad x_1 + 8x_3 = 9$

(b) $x_1 + 2x_2 + 3x_3 = 1 \qquad 2x_1 + 5x_2 + 3x_3 = 6 \qquad x_1 + 8x_3 = -6$

Solution. The two systems have the same coefficient matrix. If we augment this coefficient matrix with the columns of constants on the right sides of these systems, we obtain

$$\left[\begin{array}{ccc|c|c} 1 & 2 & 3 & 4 & 1 \\ 2 & 5 & 3 & 5 & 6 \\ 1 & 0 & 8 & 9 & -6 \end{array}\right]$$

Reducing this matrix to reduced row echelon form we get

$$\left[\begin{array}{ccc|c|c} 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 \end{array}\right]$$

It follows from the last two columns that the solution of system (a) is $x_1 = 1, x_2 = 0, x_3 = 1$ and the solution of system (b) is $x_1 = 2, x_2 = 1, x_3 = -1$.

Theorem 1.18. Let A be a square matrix.

(a) If B is a square matrix satisfying BA = I, then B = A^{-1}.

(b) If B is a square matrix satisfying AB = I, then B = A^{-1}.

Proof. (a) Assume that $BA = I$. If we can show that A is invertible, the proof can be completed by multiplying $BA = I$ on both sides by A^{-1} to obtain

$BAA^{-1} = IA^{-1}$ or $BI = IA^{-1}$ or $B = A^{-1}$.

To show that A is invertible, it suffices to show that the system $Ax = 0$ has only the trivial solution. Let x_0 be any solution of this system. If we multiply both sides of $Ax_0 = 0$ on the left by B, we obtain

$BAx_0 = B0$ or $Ix_0 = 0$ or $x_0 = 0$.

Thus, the system of equations $Ax = 0$ has only the trivial solution.

Similarly (b).

Theorem 1.19. *Equivalent Statements*

If A is an $n \times n$ matrix, then the following statements are equivalent:

(a) A is invertible.

(b) $Ax = 0$ has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) $Ax = b$ is consistent for every $n \times 1$ matrix b.

(f) $Ax = b$ has exactly one solution for every $n \times 1$ matrix b

Proof.

Since we proved in Theorem 1.17 that (a), (b), (c), and (d) are equivalent, it will be sufficient to prove that

(a) $\Rightarrow$ (f) $\Rightarrow$ (e) $\Rightarrow$ (a).

(a) $\Rightarrow$ (f) Since $A(A^{-1}b) = b$, it follows that $x = A^{-1}b$ is a solution of $Ax = b$.

To show that this is the only solution, we will assume that x_0 is an arbitrary solution and then show that x_0 must be the solution $A^{-1}b$.

If x_0 is any solution of $Ax = b$, then $Ax_0 = b$. Multiplying both sides of this equation by A^{-1}, we obtain $x_0 = A^{-1}b$.

(f) $\Rightarrow$ (e) This is almost self-evident, for if $Ax = b$ has exactly one solution for every $n \times 1$ matrix b, then $Ax = b$ is consistent for every $n \times 1$ matrix b.

(e) $\Rightarrow$ (a) If the system $Ax = b$ is consistent for every $n \times 1$ matrix b, then, in particular, this is so for the systems

$$Ax = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad Ax = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \ldots \quad Ax = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Let $x_1, x_2, ..., x_n$ be solutions of the respective systems, and let us form an n $\times$ n matrix C having these solutions as columns. Thus C has the form

$$C = \begin{bmatrix} x_1 & | & x_2 & | & \ldots & | & x_n \end{bmatrix}.$$

The successive columns of the product AC will be $Ax_1, Ax_2, ..., Ax_n$.Thus,

$$AC = \begin{bmatrix} Ax_1 & | & Ax_2 & | & \ldots & | & Ax_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = I$$

By part (b) of the above theorem, it follows that C = A^{-1}. Thus, A is invertible.

Theorem 1.20. Let A and B be square matrices of the same size. If AB is invertible, then A and B must also be invertible.

Proof. We will show first that B is invertible by showing that the homogeneous system $Bx = 0$ has only the trivial solution. If we assume that x_0 is any solution of this system, then

$(AB)x_0 = A(Bx_0) = A0 = 0$

so $x_0 = 0$ by parts (a) and (b) of the above theorem applied to the invertible matrix AB. But the invertibility of B implies the invertibility of B^{-1}, which in turn implies that

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

is invertible since the left side is a product of invertible matrices. This completes the proof.

To answer the problem: Let A be a fixed m $\times$ n matrix. Find all m $\times$ 1 matrices b such that the system of equations $Ax = b$ is consistent, we have

if A is an invertible matrix, Theorem 1.16 completely solves this problem by asserting that for every m $\times$ 1 matrix b, the linear system $Ax = b$ has the unique solution $x = A^{-1}b$. If A is not square, or if A is square but not invertible, then Theorem 1.16 does not apply. In these cases b must usually satisfy certain conditions in order for $Ax = b$ to be consistent.

Example. What conditions must b_1, b_2, and b_3 satisfy in order for the system of equations

$x_1 + x_2 + 2x_3 = b_1$

$x_1 + x_3 = b_2$

$2x_1 + x_2 + 3x_3 = b_3$ to be consistent?

Solution. The augmented matrix is

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

which can be reduced to row echelon form as follows:

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

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

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

From the third row, it is clear that the system has a solution if and only if b_1, b_2, and b_3 satisfy the condition $b_3 - b_2 - b_1 = 0$ or $b_3 = b_1 + b_2$. To express this condition another way, $Ax = b$ is consistent if and only if b is a matrix of the form $b = \begin{bmatrix} b_1 \\ b_2 \\ b_1 + b_2 \end{bmatrix}$

where b_1, b_2 are arbitrary.

Example. What conditions must b_1, b_2, and b_3 satisfy in order for the system of equations

$x_1 + 2x_2 + 3x_3 = b_1$

$2x_1 + 5x_2 + 3x_3 = b_2$

$1x_1 + 8x_3 = b_3$ to be consistent?

Solution. The augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & b_1 \\ 2 & 5 & 3 & b_2 \\ 1 & 0 & 8 & b_3 \end{array}\right].$$

Reducing this to reduced row echelon we get

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -40b_1 & 16b_2 & 9b_3 \\ 0 & 1 & 0 & 13b_1 & -5b_2 & -3b_3 \\ 0 & 0 & 1 & 5b_1 & -2b_2 & -1b_3 \end{array}\right]$$

In this case there are no restrictions on b_1, b_2, and b_3, so the system has the unique solution $x_1 = -40b_1 + 16b_2 + 9b_3$, $x_2 = 13b_1 - 5b_2 - 3b_3$, $x_3 = 5b_1 - 2b_2 - b_3$ for all values of b_1, b_2, and b_3.

1.9 Matrix Transformations

An "ordered n-tuple" is a sequence of n real numbers, a solution of a linear system in n unknowns, say $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ can be expressed as the ordered n-tuple $(s_1, s_2, \ldots, s_n)$.

If n = 2, then the n-tuple is called an "ordered pair," and if n = 3, it is called an "ordered triple." Two ordered n-tuples are same,if they list the same numbers in the same order. Thus, for example, $(1, 2)$ and $(2, 1)$ are different ordered pairs.

The set of all ordered n-tuples of real numbers is denoted by the symbol R^n. The elements of R^n are called vectors and are denoted as **a, b, v, w,** and **x** or in hand written form as $\overrightarrow{a}$, $\overrightarrow{b}$, $\overrightarrow{v}$, $\overrightarrow{w}$ and $\overrightarrow{x}$. Ordered n-tuples can also be denoted in matrix notation as column vectors. For example, $\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$ can be used as an alternative to $(s_1, s_2, \ldots, s_n)$ which is called the **comma-delimited form** of a vector and the ordered n-tuple as a matrix is called **column-vector form**.

For each i = 1, 2,...,n, let e_i denote the vector in R^n with a 1 in the ith position and zeros elsewhere. In column form these vectors are

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad e_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

We call the vectors $e_1, e_2, \ldots, e_n$ the standard basis vectors for R^n. For example, the vectors

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

are the standard basis vectors for R_3.

The vectors $e_1, e_2, \ldots, e_n$ in R^n are termed "basis vectors" because all other vectors in R^n are expressible in exactly one way as a linear combination of them. For example, if $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ then we can express $\mathbf{x}$ as $\mathbf{x} = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$.

If f is a function with domain R^n and codomain R^m, then we say that f is a transformation from R^n to R^m or that f maps from R^n to R^m, which we denote by writing $f : \mathrm{R}^n \rightarrow \mathrm{R}^m$. In the special case where $m = n$, a transformation is sometimes called an operator on R^n.

Suppose that we have the system of linear equations

$$\begin{aligned}
w_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\
w_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\
\ldots &\qquad \ldots \qquad \ldots \qquad \ldots \\
w_m &= a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \qquad\qquad (1).
\end{aligned}$$

The above system of equations can be expressed as a single matrix equation $\mathbf{w}=A\mathbf{x}$ (2),

$$\text{where } A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, \mathbf{x}= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{w}= \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \qquad (3)$$

Although we could view (2) as a compact way of writing linear system (1), we will view it instead as a transformation that maps a vector $\mathbf{x}$ in R^n into the vector $\mathbf{w}$ in R^m by multiplying $\mathbf{x}$ on the left by A. We call this a matrix transformation (or matrix operator in the special case where $m = n$). We denote it by $T_A \colon \mathrm{R}^n \to \mathrm{R}^m$.

This notation is useful when it is important to make the domain and codomain clear. The subscript on T_A serves as a reminder that the transformation results from multiplying vectors in R^n by the matrix A. In situations where specifying the domain and codomain is not essential, we will express (3) as $\mathbf{w} = T_A(\mathbf{x})$. (4)

We call the transformation T_A **multiplication by** A. On occasion we will find it convenient to express (4) in the schematic form $\mathbf{x} \xrightarrow{T_A} \mathbf{w}$ (5)

which is read "T_A maps $\mathbf{x}$ into $\mathbf{w}$."

Example. The transformation from R^4 to R^3 defined by the equations

$w_1 = 2x_1 - 3x_2 + x_3 - 5x_4$

$w_2 = 4x_1 + x_2 - 2x_3 + x_4$

$w_3 = 5x_1 - x_2 + 4x_3$

can be expressed in matrix form as

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} 2 & -3 & 1 & -5 \\ 4 & 1 & -2 & 1 \\ 5 & -1 & 4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

from which we see that the transformation can be interpreted as multiplication by $A = \begin{bmatrix} 2 & -3 & 1 & -5 \\ 4 & 1 & -2 & 1 \\ 5 & -1 & 4 & 0 \end{bmatrix}$.

If $\mathbf{x} = \begin{bmatrix} 1 \\ -3 \\ 0 \\ 2 \end{bmatrix}$, then

$$T_A(\mathbf{x}) = \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = A\mathbf{x} = \begin{bmatrix} 2 & -3 & 1 & -5 \\ 4 & 1 & -2 & 1 \\ 5 & -1 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}.$$

Zero Transformations

If 0 is the m $\times$ n zero matrix, then for any $\mathbf{x} \in \mathrm{R}^n$,

$T_0(\mathbf{x}) = 0\mathbf{x} = 0 \in \mathrm{R}^m$. So multiplication by zero maps every vector in R^n into the zero vector in R^m. We call T_0 the **zero transformation** from R^n to R^m.

Identity Operators

If I is the n $\times$ n identity matrix, then for any $\mathbf{x} \in \mathrm{R}^n$, $T_I(\mathbf{x}) = \mathrm{I}\mathbf{x} = \mathbf{x} \in \mathrm{R}^n$. So multiplication by I maps every vector in R^n to itself. We call T_I the **identity operator** on R^n.

Basic properties of matrix transformations

Theorem 1.21. For every matrix A the matrix transformation T_A: $\mathrm{R}^n \to \mathrm{R}^m$ has the following properties for all vectors u and v and for every scalar k:

(a) $T_A(\mathbf{0}) = \mathbf{0}$

(b) $T_A(\mathrm{k}\mathbf{u}) = \mathrm{k}T_A(\mathbf{u})$ [Homogeneity property]

(c) $T_A(\mathbf{u} + \mathbf{v}) = T_A(\mathbf{u}) + T_A(\mathbf{v})$ [Additivity property]

(d) $T_A(\mathbf{u} - \mathbf{v}) = T_A(\mathbf{u}) - T_A(\mathbf{v})$

Proof. Let A be an m $\times$ n matrix. By applying the properties of matrix multiplication, we get the following results:

(a) For $\mathbf{0} \in \mathrm{R}^n$, $T_A(\mathbf{0}) = A\mathbf{0} = \mathbf{0} \in \mathrm{R}^m$.

(b) For $\mathbf{u} \in \mathrm{R}^n$ and $\mathrm{k} \in \mathrm{R}$, $T_A(\mathrm{k}\mathbf{u}) = A(\mathrm{k}\mathbf{u}) = \mathrm{k}(A\mathbf{u}) = \mathrm{k}T_A(\mathbf{u})$.

(c) For $\mathbf{u}$, $\mathbf{v} \in \mathrm{R}^n$,

$T_A(\mathbf{u} + \mathbf{v}) = A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T_A(\mathbf{u}) + T_A(\mathbf{v})$.

(d) For $\mathbf{u}$, $\mathbf{v} \in \mathrm{R}^n$,

$T_A(\mathbf{u} - \mathbf{v}) = A(\mathbf{u} - \mathbf{v}) = A\mathbf{u} - A\mathbf{v} = T_A(\mathbf{u}) - T_A(\mathbf{v})$.

Remark. A matrix transformation maps a linear combination of vectors in R^n into the corresponding linear combination of vectors in R^m in the sense that

$T_A(k_1 u_1 + k_2 u_2 + \cdots + k_r u_r) = k_1 T_A(u_1) + k_2 T_A(u_2) + \cdots + k_r T_A(u_r)$.

Theorem 1.22. T_A: $R^n \to R^m$ is a matrix transformation if and only if the following relationships hold for all vectors u and v in R^n and for every scalar k:

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ [Additivity property]

(ii) $T(k\mathbf{u}) = kT(\mathbf{u})$ [Homogeneity property]

Proof. By applying the properties of matrix multiplication, we get for $\mathbf{u}$, $\mathbf{v} \in R^n$ and $k \in R$

$T(\mathbf{u} + \mathbf{v}) = T_A(\mathbf{u} + \mathbf{v}) = A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = T_A(\mathbf{u}) + T_A(\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ and

$T(k\mathbf{u}) = T_A(k\mathbf{u}) = A(k\mathbf{u}) = k(A\mathbf{u}) = kT_A(\mathbf{u}) = kT(\mathbf{u})$.

Thus if T is a matrix transformation, it satisfies the properties (i) and (ii).

Conversely, assume that properties (i) and (ii) hold. We must show that there exists an m $\times$ n matrix A such that $T(\mathbf{x}) = A\mathbf{x}$ for every vector $\mathbf{x} \in R^n$. Then,

$T(\mathrm{k}_1\mathrm{u}_1 + \mathrm{k}_2\mathrm{u}_2 + \cdots + \mathrm{k}_r\mathrm{u}_r) = \mathrm{k}_1T(\mathrm{u}_1) + \mathrm{k}_2T(\mathrm{u}_2) + \cdots + \mathrm{k}_rT(\mathrm{u}_r)$

for all scalars k_1, k_2,...,k_r and all vectors u_1, u_2,..., u_r in R^n.

Let A be the matrix $A = \left[T(e_1) \;\; | \;\; T(e_2) \;\; |T(e_3) \;\; | \ldots \;\; | \;\; T(e_n)\right]$

where $e_1, e_2, \ldots, e_n$ are the standard basis vectors for R^n. Then it follows that $A\mathbf{x}$ is a linear combination of the columns of A in which the successive coefficients are the entries $x_1, x_2, \ldots, x_n$ of $\mathbf{x}$. That is, $A\mathbf{x} = x_1Te_1 + x_2Te_2 + \ldots + x_nTe_n$ Using the above remark, we can rewrite this as $A\mathbf{x} = T(x_1e_1 + x_2e_2 + \ldots + x_ne_n)$ $= T(\mathbf{x})$ which completes the proof.

Remark. The additivity and homogeneity properties in the above theorem are called **linearity conditions**, and a transformation that satisfies these conditions is called a **linear transformation**.

Remark. *Every linear transformation from $R^n \to R^m$ is a matrix transformation,and conversely, every matrix transformation from $R^n \to R^m$ is a linear transformation.*

Theorem 1.23. If T_A: $\mathrm{R}^n \to \mathrm{R}^m$ and T_B: $\mathrm{R}^n \to \mathrm{R}^m$ are matrix transformations, and if $T_A(\mathbf{x}) = T_B(\mathbf{x})$ for every vector $\mathbf{x}$ in R^n, then $A = B$.

Proof. Since $T_A(\mathbf{x}) = T_B(\mathbf{x})$ for every vector $\mathbf{x}$ in R^n, then $A\mathbf{x}$

$= B\mathbf{x}$ for every vector $\mathbf{x}$ in R^n.

Let $e_1, e_2, \ldots, e_n$ be the standard basis vectors for R^n. Then $Ae_j = Be_j$ for all j=1,2,...,n.

Since every entry of e_j is 0 except for the j th, which is 1, it follows that Ae_j is the j th column of A and Be_j is the j th column of B. Thus, $Ae_j = Be_j$ implies that corresponding columns of A and B are the same, and hence that $A = B$.

Remark. *Every $m \times n$ matrix A produces exactly one matrix transformation (multiplication by A) and every matrix transformation from R^n to R^m arises from exactly one $m \times n$ matrix; we call that matrix the **standard matrix** for the transformation.*

We showed that if $e_1, e_2, \ldots, e_n$ are the standard basis vectors for R^n (in column form), then the standard matrix for a linear transformation T: $\mathrm{R}^n \to \mathrm{R}^m$ is given by the formula

$$A = \left[T(e_1) \;\; | \;\; T(e_2) \;\; | T(e_3) \;\; | \ldots \;\; | \;\; T(e_n) \right]$$

Finding the Standard Matrix for a Matrix Transformation

Step 1. Find the images of the standard basis vectors $e_1, e_2, \ldots, e_n$ for R^n.

Step 2. Construct the matrix that has the images obtained in Step 1 as its successive columns. This matrix is the standard matrix for the transformation.

Example . Find the standard matrix A for the linear transformation T: $\mathrm{R}^2 \to \mathrm{R}^2$ defined by the formula

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 \\ x_1 - 3x_2 \\ -x_1 + x_2 \end{bmatrix}.$$

Also find $T(1,4)$ by matrix multiplication.

Let e_1, e_2 be the standard basis for R^2. Then,

$$T(e_1) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2.1 + 0 \\ 1 - 3.0 \\ -1 + 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

$$T(e_2) = T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2.0 + 1 \\ 0 - 3.1 \\ -0 + 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}$$

Hence the standard matrix A for the given linear transformation is

$$A = \begin{bmatrix} T(e_1) & | & T(e_2) \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -3 \\ -1 & 1 \end{bmatrix}$$

Now, $T(1,4) = T\left(\begin{bmatrix}1\\4\end{bmatrix}\right) = A\left(\begin{bmatrix}1\\4\end{bmatrix}\right) = \begin{bmatrix}2 & 1\\1 & -3\\-1 & 1\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}$

$= \begin{bmatrix}6\\-11\\3\end{bmatrix} = (6,-11,3).$

Example. Rewrite the transformation $T(x_1, x_2) = (3x_1 + x_2, 2x_1 - 4x_2)$ in column-vector form and find its standard matrix.

$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}3x_1 + x_2\\2x_1 - 4x_2\end{bmatrix} = \begin{bmatrix}3 & 1\\2 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}.$

Thus the standard matrix is $\begin{bmatrix}3 & 1\\2 & -1\end{bmatrix}$.

Question. Find the standard matrix for the operator T: $\mathrm{R}^3 \to \mathrm{R}^3$ defined by $w_1 = 3x_1 + 5x_2 - x_3$, $w_2 = 4x_1 - x_2 + x_3$, $w_3 = 3x_1 + 2x_2 - x_3$ and then compute $T(-1,2,4)$ by directly substituting in the equations and then by matrix multiplication.

Answer. $\begin{bmatrix}3\\-2\\-3\end{bmatrix}$.

MODULE

TWO

GENERAL VECTOR SPACES

2.1 Vector spaces

Let V be an arbitrary nonempty set of objects on which two operations are defined: addition, and multiplication by numbers called scalars.By **addition** we mean a rule for associating with each pair of objects $\mathbf{u}$ and $\mathbf{v}$ in V an object $\mathbf{u} + \mathbf{v}$, called the **sum** of $\mathbf{u}$ and $\mathbf{v}$; by **scalar multiplication** we mean a rule for associating with each scalar k and each object $\mathbf{u}$ in V an object ku, called the **scalar multiple** of $\mathbf{u}$ by k.

If the following axioms are satisfied by all objects **u, v, w** in V and all scalars k and m, then we call V a **vector space** and we call the objects in V **vectors**.

1. If **u** and **v** are objects in V, then **u + v** is in V.
2. **u + v** = **v + u**
3. **u + (v + w)** = **(u + v) + w**
4. There is an object **0** in V, called a **zero vector** for V, such that **0 + u** = **u + 0** = **u** for all **u** in V.
5. For each **u** in V, there is an object **-u** in V, called a **negative** of **u**, such that **u + (-u)** = **(-u) + u** = **0**.
6. If k is any scalar and **u** is any object in V, then k**u** is in V.
7. k(**u + v**) = k**u** + k**v**
8. (k + m)**u** = k**u** + m**u**
9. k(m**u**) = (km)(**u**)
10. 1**u** = **u**

*Vector spaces with real scalars will be called **real vector spaces** and those with complex scalars will be called **complex vector spaces**.For now, we will focus exclusively on real vector spaces, which we will refer to simply as "vector spaces."*

To Show That a Set with Two Operations Is a Vector Space

Step 1. Identify the set V of objects that will become vectors.

Step 2. Identify the addition and scalar multiplication operations on V.

Step 3. Verify Axioms 1 and 6; that is, adding two vectors in V produces a vector in V, and multiplying a vector in V by a scalar also produces a vector in V. Axiom 1 is called closure under addition, and Axiom 6 is called closure under scalar multiplication.

Step 4. Confirm that Axioms 2, 3, 4, 5, 7, 8, 9, and 10 hold.

Examples

1. $V = R$, the set of real numbers.

2. $V = R^n$, the set of n-tuples of real numbers.

3. Let V consist of a single object, which we denote by $\mathbf{0}$, and define $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and $\mathbf{k0} = \mathbf{0}$ for all scalars k. This vector space is called as the **zero vector space**.

4. $V = R^\infty$, the vector space of infinite sequences of real numbers.

Here V consist of objects of the form $u = (u_1, u_2, \ldots, u_n, \ldots)$ in which $u_1, u_2, \ldots, u_n, \ldots$ is an infinite sequence of real numbers. We define two infinite sequences to be equal if their corresponding components are equal, and we define addition and scalar multiplication componentwise by

$u+v = (u_1, u_2, \ldots, u_n, \ldots) + (v_1, v_2, \ldots, v_n, \ldots) = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n, \ldots)$ *and*

$ku = (ku_1, ku_2, \ldots, ku_n, \ldots)$.

5. $V = M_{22}$, the set of 2×2 matrices with real entries, and take the vector space operations on V to be the usual operations of matrix addition and scalar multiplication.

In general, $V = M_{mn}$, the set of m $\times$ n matrices with real entries, and take the vector space operations on V to be the usual operations of matrix addition and scalar multiplication is a vector space.

6. *The Vector Space of Real-Valued Functions* Let V be the set of real-valued functions that are defined at each **x** in the interval $(-\infty, \infty)$.

If $f = f(x)$ and $g = g(x)$ are two functions in V and if k is any scalar, then define the operations of addition and scalar multiplication by

$(f+g)(x) = f(x) + g(x) \quad \text{and} \quad (kf)(x) = kf(x)$

7. $V = C[a, b]$, the set of real-valued continuous functions that

are defined at each $\mathbf{x}$ on the interval $[a, b]$.

If $f = f(x)$ and $g = g(x)$ are two functions in V and if k is any scalar, then define the operations of addition and scalar multiplication by

$(f + g)(x) = f(x) + g(x)$ and $(kf)(x) = kf(x)$

8. $V = P_n$, the set of all polynomials of **degree less than or equal to** n and with real coefficients.

For any $p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$ and $q(x) = b_n x^n + b_{n-1} x^{n-1} + \ldots + b_1 x + b_0$ in P_n and $k \in R$, define $p(x) + q(x) = (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \ldots + (a_1 + b_1)x + (a_0 + b_0)$ and $kp(x) = ka_n x^n + ka_{n-1} x^{n-1} + \ldots + ka_1 x + ka_0$.

Then P_n is a vector space under these operations.

9. $V = R^+$, the set of all real positive numbers. Define the operations on R^+ by

$x + y = xy$ and $kx = x^k$ [Vector addition is numerical multiplication and scalar multiplication is numerical exponentiation.]

Then R^+ with the above operations is a vector space.

Examples of set which are not vector spaces.

10. Let $V = R^2$ and define addition and scalar multiplication operations as follows: If

$u = (u_1, u_2)$ and $v = (v_1, v_2)$, then define $u+v = (u_1+v_1, u_2+v_2)$ and if k is any real number, then define $ku = (ku_1, 0)$ For example, if $u = (2, 4), v = (-3, 5)$, and $k = 7$, then

$u + v = (2 + (-3), 4 + 5) = (-1, 9)$ and

$ku = 7u = (7 \times 2, 0) = (14, 0)$.

Now, Axiom 10 fails to hold for certain vectors. For example, if $u = (u_1, u_2)$ is such that $u_2 \neq 0$, then $1u = 1(u_1, u_2) = (1 \times u_1, 0) = (u_1, 0) \neq u$ Thus, V is not a vector space.

11. The set of all polynomials of **degree** n with usual rule of addition and scalar multiplication (as defined in example 8.) since zero polynomial is not present in the set. Also it is not closed with respect to addition.

12. $V = R^+$, the set of all real positive numbers with usual addition and scalar multiplication is not a vector space.

Properties

Theorem 2.1 Let V be a vector space, **u** a vector in V, and k a scalar; then:

(a) $0\mathbf{u} = \mathbf{0}$

(b) $k\mathbf{0} = \mathbf{0}$

(c) (-1)$\mathbf{u}$ = -$\mathbf{u}$

(d) If k$\mathbf{u}$ = $\mathbf{0}$, then k = 0 or $\mathbf{u} = \mathbf{0}$.

Proof. (a) We can write 0$\mathbf{u}$ + 0$\mathbf{u}$ = (0 + 0)$\mathbf{u}$ [Axiom 8]

= 0$\mathbf{u}$ [Property of the number 0]

By Axiom 5 the vector 0$\mathbf{u}$ has a negative, -0$\mathbf{u}$. Adding this negative to both sides above we get

(0$\mathbf{u}$ + 0$\mathbf{u}$) + (-0$\mathbf{u}$) = 0$\mathbf{u}$ + (-0$\mathbf{u}$)

0$\mathbf{u}$+ (0$\mathbf{u}$ + (-0$\mathbf{u}$)) = 0$\mathbf{u}$ + (-0$\mathbf{u}$) [Axiom 3]

0$\mathbf{u}$ + $\mathbf{0}$ =$\mathbf{0}$ [Axiom 5]

0$\mathbf{u} = \mathbf{0}$ [Axiom 4].

(b) By Axioms 7 and 4, k$\mathbf{0}$+k$\mathbf{0}$=k$\mathbf{0+0}$=k$\mathbf{0}$.

By Axiom 5 the vector k$\mathbf{0}$ has a negative, -k$\mathbf{0}$. Adding this negative to both sides above we get

(k$\mathbf{0}$+k$\mathbf{0}$)+(-k$\mathbf{0}$)=k$\mathbf{0}$+(-k$\mathbf{0}$)

k$\mathbf{0}$+(k$\mathbf{0}$+-k$\mathbf{0}$)=k$\mathbf{0}$+(-k$\mathbf{0}$) [Axiom 3]

k$\mathbf{0}$+$\mathbf{0}$ = $\mathbf{0}$ [Axiom 5]

k$\mathbf{0}$ = $\mathbf{0}$ [Axiom 4].

(c) To prove that (-1)$\mathbf{u}$ = -$\mathbf{u}$, we must show that

$\mathbf{u}$ + (-1)$\mathbf{u}$ = 0. The proof is as follows:

$\mathbf{u}$ + (-1)$\mathbf{u}$ = 1$\mathbf{u}$ + (-1)$\mathbf{u}$ [Axiom 10]

$= (1 + (-1))\mathbf{u}$ [Axiom 8]

$= 0\mathbf{u}$ [Property of numbers]

$= \mathbf{0}$ [Part (a) of this theorem]

(d) Suppose that $k\mathbf{u} = \mathbf{0}$ and $k \neq 0$. Then k has a multiplicative inverse 1/k in R. Hence we get

$\mathbf{u}=1\mathbf{u}$ [Axiom 10]

$=(1/k\ .\ k)\mathbf{u}=(1/k)(k\mathbf{u})$ [Axiom 9]

$=(1/k)\mathbf{0}=\mathbf{0}$ [by part (b)]

Also if k=0, then by part (a), even if $\mathbf{u} \neq \mathbf{0}$, $k\mathbf{u} = 0\mathbf{u}=\mathbf{0}$.

Hence if $k\mathbf{u} = \mathbf{0}$, then either $k = 0$ or $\mathbf{u} = \mathbf{0}$.

2.2 Subspaces

A subset W of a vector space V is called a subspace of V if W is itself a vector space under the addition and scalar multiplication defined on V.

Theorem 2.2 If W is a set of one or more vectors in a vector space V, then W is a subspace of V if and only if the following conditions are satisfied.

(a) If $\mathbf{u}$ and $\mathbf{v}$ are vectors in W, then $\mathbf{u} + \mathbf{v}$ is in W.

(b) If k is a scalar and $\mathbf{u}$ is a vector in W, then k$\mathbf{u}$ is in W.

Proof. If W is a subspace of V, then all the vector space axioms hold in W, including Axioms 1 and 6, which are precisely conditions (a) and (b).

Conversely, assume that conditions (a) and (b) hold. Since these are Axioms 1 and 6, and since Axioms 2,3,7,8,9, and 10 are inherited from V, we only need to show that Axioms 4 and 5 hold in W. For this purpose, let $\mathbf{u}$ be any vector in W. It follows from condition (b) that k$\mathbf{u}$ is a vector in W for every scalar k. In particular, $0\mathbf{u} = \mathbf{0}$ and $(-1)\mathbf{u} = -\mathbf{u}$ are in W, which shows that Axioms 4 and 5 hold in W.

Remark. The conditions (a) and (b) are equivalent to a single condition $k\mathbf{u}+m\mathbf{v} \in$ W, for all $\mathbf{u}$, $\mathbf{v} \in$ W and for all scalars k and m. Hence we can rewrite the above theorem as:

A non-empty subset W of a vector space V is a subspace of V if and only if $k\boldsymbol{u}+m\boldsymbol{v} \in W$*, for all* $\boldsymbol{u}$*,* $\boldsymbol{v} \in W$ *and for all scalars k and m.*

Examples

1. If V is any vector space, and if W = $\mathbf{0}$ is the subset of V

that consists of the zero vector only, then W is closed under addition and scalar multiplication since **0+0=0** and k**0=0** for any scalar k. We call W the **zero subspace** of V.

Note that every vector space has at least two subspaces, itself and its zero subspace.

2. Consider the vector space R^2, with component wise addition and scalar multiplication. Then both $W_1 = \{(x, 0) : x \in R\}$ and $W_2 = \{(0, y) : y \in R\}$ are subspaces of R^2. Geometrically, they are X and Y axes respectively. In general, if W is a line through the origin, then adding two vectors on a line or multiplying a vector on the line by a scalar produces another vector on the line and so W is closed under addition and scalar multiplication. Hence W is a subspace of R^2.

Let S be the set of all points (x,y) in R^2 for which x $\geq$ 0 and y $\geq$ 0. This set is **not a subspace** of R^2 because it is not closed under scalar multiplication. For example, **v** = (1, 1) is a vector in W, but (-1)**v** = (-1, -1) is not.

3. Let $V = M_{nn}$. Then the set of symmetric n $\times$ n matrices is closed under addition and scalar multiplication and hence is a subspace of M_{nn}. Similarly, the sets of upper triangular

matrices, lower triangular matrices, and diagonal matrices are subspaces of M_{nn}.

The set W of invertible n × n matrices is **not a subspace** of M_{nn}, as it is not closed under addition and not closed under scalar multiplication. As an example in M_{22} , consider the matrices
$U = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ and $V = \begin{bmatrix} -1 & 2 \\ -2 & 5 \end{bmatrix}$ The matrix $0U$ is the 2 × 2 zero matrix and hence is not invertible, and the matrix $U + V$ has a column of zeros so it also is not invertible.

Theorem 2.3 If $W_1, W_2, \ldots, W_r$ are subspaces of a vector space V, then the intersection of these subspaces is also a subspace of V.

Proof. Let W be the intersection of the subspaces $W_1, W_2, \ldots, W_r$. This set is not empty because each of these subspaces contains the zero vector of V, and hence so does their intersection. Thus, it remains to show that W is closed under addition and scalar multiplication.

To prove closure under addition, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in W. Since W is the intersection of $W_1, W_2, \ldots, W_r$, it follows that $\mathbf{u}$

and $\mathbf{v}$ also lie in each of these subspaces. Moreover, since these subspaces are closed under addition and scalar multiplication, they also all contain the vectors $\mathbf{u}+\mathbf{v}$ and k$\mathbf{u}$ for every scalar k, and hence so does their intersection W. This proves that W is closed under addition and scalar multiplication.

Remark. In general if U and W are subspaces of a vector space V, then $U \cup W$ need not be a subspace.

For example, $U = \{(x, 0, 0) : x \in R\}$ and $W = \{(0, y, 0) : y \in R\}$ are subspaces of R^3. But $(x, 0, 0) + (0, y, 0) = (x, y, 0) \notin U \cup W$. Hence $U \cup W$ is not a subspace.

Problems:

1. Define W as $W = \{(x_1, x_2, x_3) \in R^3 : a_1x_1 + a_2x_2 + a_3x_3 = 0\}$ for some fixed scalars a_1, a_2, a_3. Show that W is a subspace of R^3.

2. Let P_n be the space of all polynomials of degree less than or equal to n. Then show that

$S = \{p(x) \in P_n : p(1) = 0 \; and \; p(3) = 0\}$ is a subspace of P_n.

2.3 Linear combinations

If **w** is a vector in a vector space V, then **w** is said to be a **linear combination** of the vectors $v_1, v_2, \ldots, v_r$ in V if **w** can be expressed in the form $w = k_1v_1 + k_2v_2 + \cdots + k_rv_r$ where $k_1, k_2, \ldots, k_r$ are scalars. These scalars are called the **coefficients** of the linear combination.

Theorem 2.4 If $S = \{w_1, w_2, \ldots, w_r\}$ is a nonempty set of vectors in a vector space V, then:

(a) The set W of all possible linear combinations of the vectors in S is a subspace of V.

(b) The set W in part(a) is the "smallest" subspace of V that contains all of the vectors in S in the sense that any other subspace that contains those vectors contains W.

Proof. (a) Let W be the set of all possible linear combinations of the vectors in S. We must show that W is closed under addition and scalar multiplication. To prove closure under addition, let $u = c_1w_1 + c_2w_2 + \cdots + c_rw_r$ and $v = k_1w_1 + k_2w_2 + \cdots + k_rw_r$ be two vectors in W. It follows that their sum can be written as $u + v = (c_1 + k_1)w_1 + (c_2 + k_2)w_2 + \cdots + (c_r + k_r)w_r$ which is a linear combination of the vectors in S. Thus, W is closed under addition. Similarly, for any scalar k, $ku = k(c_1w_1 + c_2w_2 + \cdots + c_rw_r) = kc_1w_1 + kc_2w_2 + \cdots + kc_rw_r$ which is a linear combinations of the vectors in S. Thus W is closed under scalar multiplication. Hence is a subspace of V.

(b) Let W be any subspace of V that contains all of the vectors in S. Since W is closed under addition and scalar multiplication, it contains all linear combinations of the vectors in S and hence contains W.

If $S = \{w_1, w_2, \ldots, w_r\}$ is a nonempty set of vectors in a vector space V , then the subspace W of V that consists of all possible linear combinations of the vectors in S is called the **subspace of V generated by S**, and we say that the vectors $w_1, w_2, \ldots, w_r$ **span** W. We denote this subspace as $W = span\{w_1, w_2, \ldots, w_r\}$ or $W = span(S)$.

Examples

1. We know that the standard unit vectors in R^n are $e_1 = (1, 0, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, 0, \ldots, 1)$. These vectors span R^n since every vector $v = (v_1, v_2, \ldots, v_n)$ in R^n can be expressed as $v = v_1e_1 + v_2e_2 + \cdots + v_ne_n$

which is a linear combination of $e_1 = (1,0,0,\ldots,0), e_2 = (0,1,0,\ldots,0),\ldots,e_n = (0,0,0,\ldots,1)$. Thus, for example, the vectors $i = (1,0,0), j = (0,1,0), k = (0,0,1)$ span R^3 since every vector $v = (a,b,c)$ in this space can be expressed as $v = (a,b,c) = a(1,0,0) + b(0,1,0) + c(0,0,1) = ai + bj + ck$.

2. In R^3, span{(1,0,0),(0,0,1)}={x(1,0,0)+z(0,0,1):x,z∈R} $=$ {(x,0,z):x,z∈R}. Hence the subspace of R^3 spanned by the set {(1,0,0),(0,0,1)} is the xz-plane.

3. The polynomials $1, x, x^2, \ldots, x^n$ span the vector space of all polynomials of degree less than or equal to n, since each polynomial $p(x) \in P_n$ which is of the form $p(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ can be written as a linear combination of $1, x, x^2, \ldots, x^n$. Hence $P_n = span\{1, x, x^2, \ldots, x^n\}$.

Geometric view in R^2 and R^3.

If v is a nonzero vector in R^2 or R^3 that has its initial point at the origin, then spanv, which is the set of all scalar multiples of v, is the line through the origin determined by v.

If u and v are nonzero vectors in R^3 that have their initial points at the origin, then spanu,v, which consists of all linear combinations of u and v is the plane through the origin determined by these two vectors.

4. Consider the vectors u = (1, 2, -1) and v = (6, 4, 2) in R^3. Then w = (9, 2, 7) is a linear combination of u and v and that r = (4, -1, 8) is not a linear combination of u and v.

In order for w to be a linear combination of u and v, there must be scalars k_1 and k_2 such that w = k_1u + k_2v; that is, (9, 2, 7) = k_1(1, 2, -1) + k_2(6, 4, 2) = (k_1 + 6k_2, 2k_1 + 4k_2, -k_1 + 2k_2)

Equating corresponding components gives k_1 + 6k_2 = 9, 2k_1 + 4k_2 = 2, -k_1 + 2k_2 = 7. Solving this system using Gaussian elimination yields k_1 = -3, k_2 = 2, so w = -3u + 2v.

Similarly, for r to be a linear combination of u and v, there must be scalars k_1 and k_2 such that r = k_1u + k_2v; that is, (4, -1, 8) = k_1(1, 2, -1) + k_2(6, 4, 2) = (k_1 + 6k_2, 2k_1 + 4k_2, -k_1 + 2k_2) Equating corresponding components gives k_1 + 6k_2 = 4, 2k_1 + 4k_2 = -1, -k_1 + 2k_2 = 8. This system of equations is inconsistent, so no such scalars k_1 and k_2 exist. Consequently, r is not a linear combination of u and v.

5. Determine whether the vectors v_1 = (1, 1, 2), v_2 = (1, 0, 1), and v_3 = (2, 1, 3) span the vector space R^3.

Solution. We want to determine whether an arbitrary vector b $= (b_1, b_2, b_3)$ in R^3 can be expressed as a linear combination $b = k_1v_1 + k_2v_2 + k_3v_3$ of the vectors v_1, v_2, and v_3. Expressing this equation in terms of components gives us

$(b_1, b_2, b_3) = k_1(1, 1, 2) + k_2(1, 0, 1) + k_3(2, 1, 3)$

i.e; $(b_1, b_2, b_3) = (k_1 + k_2 + 2k_3, k_1 + k_3, 2k_1 + k_2 + 3k_3)$

i.e; $k_1 + k_2 + 2k_3 = b_1$, $k_1 + k_3 = b_2$, $2k_1 + k_2 + 3k_3 = b_3$

Thus, our problem reduces to ascertaining whether this system is consistent for all values of b_1, b_2, and b_3. We know that the system is consistent if and only if its coefficient matrix has a nonzero determinant. Now, here the coefficient matrix is,

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

But det(A) = 0, so v_1, v_2, and v_3 do not span R^3.

Problems:

1. Check whether (2,-5,3) in R^3 can be written as a linear combination of (1,-3,2), (2,-4,-1) and (1,-5,7).

2. Write (1,0) as a linear combination of (1,1) and (-1,2).

2.4 Solution Spaces of Homogeneous Systems

Theorem 2.5 The solution set of a homogeneous linear system $A\mathbf{x} = \mathbf{0}$ of m equations in n unknowns is a subspace of R^n.

Proof. Let W be the solution set of the system. The set W is not empty as it contains at least the trivial solution $\mathbf{x} = \mathbf{0}$.

To show that W is a subspace of R^n, we must show that it is closed under addition and scalar multiplication. To do this, let $\mathbf{x}_1$ and $\mathbf{x}_2$ be vectors in W. Since these vectors are solutions of $A\mathbf{x} = \mathbf{0}$, we have $\mathrm{A}\mathbf{x}_1 = 0$ and $\mathrm{A}\mathbf{x}_2 = 0$.

It follows from these equations and the distributive property of matrix multiplication that $\mathrm{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathrm{A}\mathbf{x}_1 + \mathrm{A}\mathbf{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0}$, so W is closed under addition. Similarly, if k is any scalar then $\mathrm{A}(\mathrm{k}\mathbf{x}_1) = \mathrm{kA}\mathbf{x}_1 = \mathrm{k}\mathbf{0} = \mathbf{0}$, so W is also closed under scalar multiplication.

Theorem 2.5 can be viewed as a statement about matrix transformations by letting T_A: $R^n \rightarrow R^m$ be multiplication by the coefficient matrix A. From this point of the solution space of $Ax = 0$ is the set of vectors in R^n that T_A maps into the *zero vector in R^m. This set is sometimes called the* ***kernel*** *of the transformation.*

Theorem 2.6 If A is an m $\times$ n matrix, then the kernel of the matrix transformation T_A: $\mathrm{R}^n \to \mathrm{R}^m$ is a subspace of R^n.

Remark. The solution set of every homogeneous system of m equations in n unknowns is a subspace of R^n, it is never true that the solution set of a nonhomogeneous system of m equations in n unknowns is a subspace of R^n. There are two possible cases: first, the system may not have any solutions at all, and second, if there are solutions, then the solution set will not be closed either under addition or under scalar multiplication.

Theorem 2.7 If $S = \{v_1, v_2, \ldots, v_r\}$ and $T = \{w_1, w_2, \ldots, w_k\}$ are nonempty sets of vectors in a vector space V, then $span\{v_1, v_2, \ldots, v_r\} = span\{w_1, w_2, \ldots, w_k\}$ if and only if each vector in S is a linear combination of those in T and each vector in T is a linear combination of those in S.

Problems

In each part, solve the system by any method and then give a geometric description of the solution set.

(a) $\begin{bmatrix} 1 & -2 & 3 \\ 2 & -4 & 6 \\ 3 & -6 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

(b) $\begin{bmatrix} 1 & -2 & 3 \\ 3 & 7 & -8 \\ -2 & 4 & -6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

(c) $\begin{bmatrix} 1 & -2 & 3 \\ 3 & 7 & -8 \\ 4 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

(d) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$

Solution. After solving we get

(a) The solutions are x = 2s - 3t, y = s, z = t from which it follows that x = 2y - 3z or x - 2y + 3z = 0. This is the equation of a plane through the origin that has n = (1, -2, 3) as a normal.

(b) The solutions are x = -5t, y = -t, z = t which are parametric equations for the line through the origin that is parallel to the vector v = (-5, -1, 1).

(c) The only solution is x = 0, y = 0, z = 0, so the solution space consists of the single point 0.

(d) This linear system is satisfied by all real values of x, y, and z, so the solution space is all of R^3.

2.5 Linear Independence

If $S = \{v_1, v_2, \ldots, v_r\}$ is a set of two or more vectors in a vector space V , then S is said to be a **linearly independent** set if no vector in S can be expressed as a linear combination of the others. A set that is not linearly independent is said to be **linearly dependent**.

Theorem 2.8 A nonempty set $S = \{v_1, v_2, \ldots, v_r\}$ in a vector space V is linearly independent if and only if the only coefficients satisfying the vector equation $k_1 v_1 + k_2 v_2 + \ldots + k_r v_r = 0$ are $k_1 = 0, k_2 = 0, \ldots, k_r = 0$.

Examples.

1. The most basic linearly independent set in R^n is the set of standard unit vectors $e_1 = (1, 0, 0, \ldots, 0), e_2 =$

$(0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, 0, \ldots, 1)$.

In R^3, consider the standard unit vectors $\mathbf{i} = (1, 0, 0), \mathbf{j} = (0, 1, 0), \mathbf{k} = (0, 0, 1)$. To prove linear independence we must show that $k_1\mathbf{i} + k_2\mathbf{j} + k_3\mathbf{k} = 0 \Rightarrow k_1 = 0, k_2 = 0, k_3 = 0$.

Now, $k_1\mathbf{i} + k_2\mathbf{j} + k_3\mathbf{k} = \mathbf{0} \Rightarrow k_1(1, 0, 0) + k_2(0, 1, 0) + k_3(0, 0, 1) = (0, 0, 0) \Rightarrow (k_1, k_2, k_3) = (0, 0, 0) \Rightarrow k_1 = 0, k_2 = 0, k_3 = 0$.

2. Determine whether the vectors $v_1 = (1, -2, 3)$, $v_2 = (5, 6, -1)$, $v_3 = (3, 2, 1)$ are linearly independent or linearly dependent in R^3.

Consider the vector equation $k_1 v_1 + k_2 v_2 + k_3 v_3 = \mathbf{0}$. That is,

$k_1(1, -2, 3) + k_2(5, 6, -1) + k_3(3, 2, 1) = \mathbf{(0, 0, 0)}$.

Equating corresponding components on the two sides yields the homogeneous linear system

$k_1 + 5k_2 + 3k_3 = 0$

$-2k_1 + 6k_2 + 2k_3 = 0$

$3k_1 - k_2 + k_3 = 0$

Thus, our problem reduces to determining whether this system has nontrivial solutions. One method is to solve the system, which yields $k_1 = -1/2\ t$, $k_2 = -1/2\ t$, $k_3 = t$

This shows that the system has nontrivial solutions and hence that the vectors are linearly dependent.

A second method for establishing the linear dependence is to consider the **coefficient matrix** and compute its determinant. If it's **determinant is equal to zero**, then the linear system will have nontrivial solutions and the vectors v_1, v_2, and v_3 will be **linearly dependent**.

Here, $A = \begin{bmatrix} 1 & 5 & 3 \\ -2 & 6 & 2 \\ 3 & -1 & 1 \end{bmatrix}$ and $\det(A) = 0$ and hence v_1, v_2, and v_3 are linearly dependent.

3. Determine whether the vectors $v_1 = (1, 2, 2, -1), v_2 = (4, 9, 9, -4), v_3 = (5, 8, 9, -5)$ in R^4 are linearly dependent or linearly independent.

The linear independence or linear dependence of these vectors is determined by whether there exist nontrivial solutions of the vector equation $k_1v_1 + k_2v_2 + k_3v_3 = 0$ or, equivalently, of $k_1(1, 2, 2, -1) + k_2(4, 9, 9, -4) + k_3(5, 8, 9, -5) = (0, 0, 0, 0)$.

Equating corresponding components on the two sides, we get the homogeneous linear system $k_1 + 4k_2 + 5k_3 = 0$

$2k_1 + 9k_2 + 8k_3 = 0$

$2k_1 + 9k_2 + 9k_3 = 0$

$-k_1 - 4k_2 - 5k_3 = 0$

Solving we get, this system has only the trivial solution $k_1 = 0, k_2 = 0, k_3 = 0$ from which you can conclude that v_1, v_2 and v_3 are linearly independent.

4. Show that the polynomials $1, x, x^2, \ldots, x^n$ form a linearly independent set in P_n.

Let us denote the polynomials as $p_0 = 1, p_1 = x, p_2 = x^2, \ldots, p_n = x^n$.

We must show that the only coefficients satisfying the vector equation $a_0p_0 + a_1p_1 + a_2p_2 + \ldots + a_np_n = 0$ are $a_0 = a_1 = a_2 = \ldots = a_n = 0$. But it is equivalent to the statement that $a_0 + a_1x + a_2x^2 + \ldots + a_nx^n = 0$ for all x in $(-\infty, \infty)$, so we must show that this is true if and only if each coefficient is zero.

We know that *a nonzero polynomial of degree n has at most n distinct roots.* That being the case, each coefficient must be zero, for otherwise the left side of the equation would be a nonzero polynomial with infinitely many roots. Thus, $a_0p_0 + a_1p_1 + a_2p_2 + \ldots + a_np_n = 0$ has only the trivial solution.

5. Determine whether the polynomials $p_1 = 1 - x, p_2 =$

$5+3x-2x^2, p_3 = 1+3x-x^2$ are linearly dependent or linearly independent in P_2.

The linear independence or dependence of these vectors is determined by whether the vector equation $k_1p_1 + k_2p_2 + k_3p_3 = 0$ can be satisfied with coefficients that are not all zero. To see this is so, let us write the equation in its polynomial form as $k_1(1-x)+k_2(5+3x-2x^2)+k_3(1+3x-x^2) = 0$ or, equivalently, as $(k_1 + 5k_2 + k_3) + (-k_1 + 3k_2 + 3k_3)x + (-2k_2 - k_3)x^2 = 0$ Since this equation must be satisfied by all x in $(-\infty, \infty)$, each coefficient must be zero (as explained in the previous example). Thus, the linear dependence or independence of the given polynomials determined by whether the following linear system has a nontrivial solution: $k_1 + 5k_2 + k_3 = 0$

$-k_1 + 3k_2 + 3k_3 = 0$

$-2k_2 - k_3 = 0.$

Since the coefficient matrix has determinant zero, this linear system has nontrivial solutions. Thus, the set p_1, p_2, p_3 is linearly dependent.

Theorem 2.9 (a) A finite set that contains **0** is linearly dependent.

(b) A set with exactly one vector is linearly independent if and only if that vector is not **0**.

(c) A set with exactly two vectors is linearly independent if and only if neither vector is a scalar multiple of the other.

Proof. (a) For any vectors $v_1, v_2, \ldots, v_r$, the set $S = \{v_1, v_2, \ldots, v_r, 0\}$ is linearly dependent since the equation $0v_1 + 0v_2 + \ldots + 0v_r + 1(0) = 0$ expresses **0** as a linear combination of the vectors in S with coefficients that are not all zero.

(b) Let V be a vector space and $v \in V$. Let $v \neq 0$. Then for any scalar α, $\alpha v = 0 \Rightarrow \alpha = 0$. Hence $\{v\}$ is linearly independent.

Conversely, let $\{v\}$ be linearly independent. Then from (a), we understand $v \neq 0$.

(c) Let V be a vector space and $S = \{v_1, v_2\}$ be a subset of V, with exactly two vectors. Suppose that one is a scalar multiple of the other. Then there exists a non zero scalar α such that $v_1 = \alpha v_2$. Then by axioms and properties of a vector space, $v_1 = \alpha v_2 \Rightarrow v_1 + (-\alpha v_2) = \alpha v_2 + (-\alpha v_2) \Rightarrow v_1 + (-\alpha)v_2 = 0$. Hence the set $S = \{v_1, v_2\}$ is linearly dependent.

Now let the set $S = \{v_1, v_2\}$ be linearly dependent. Then there exists scalars α_1, α_2 not both of them are zero, such that $\alpha_1 v_1 + \alpha_2 v_2 = 0$. If $\alpha_1 \neq 0$, then $v_1 = -(\alpha_2 \alpha_1/)v_2$ and if $\alpha_2 \neq 0$, then $v_2 = -(\alpha_1 \alpha_2/)v_1$. Thus if $S = \{v_1, v_2\}$ is linearly dependent, then one of them is a scalar multiple of the other.

Example 6. The functions $f_1 = x$ and $f_2 = sinx$ are linearly independent vectors in $F(-\infty, \infty)$ since neither function is a scalar multiple of the other. On the other hand, the two functions $g_1 = sin2x$ and $g_2 = sinxcosx$ are linearly dependent because the trigonometric identity $sin2x = 2sinxcosx$ reveals that g_1 and g_2 are scalar multiples of each other.

Exanmple 7. the functions $f_1 = sin^2x, f_2 = cos^2x$, and $f_3 = 5$ form a linearly dependent set in $F(-\infty, \infty)$, since the equation $5f_1 + 5f_2 - f_3 = 5sin^2x + 5cos^2x - 5 = 5(sin^2x + cos^2x) - 5 = 0$ expresses 0 as a linear combination of $f_1, f_2,$ and f_3 with coefficients that are not all zero.

Linear independence has the following useful geometric interpretations in R^2 and R^3:

- Two vectors in R^2 or R^3 are linearly independent if and only if they do not lie on the same line when they have their initial points at the origin. Otherwise one would be a scalar multiple of the other.

• Three vectors in R^3 are linearly independent if and only if they do not lie in the same plane when they have their initial points at the origin. Otherwise at least one would be a linear combination of the other two.

Theorem 2.10 Let $S = \{v_1, v_2, \ldots, v_r\}$ be a set of vectors in R^n. If $r > n$, then S is linearly dependent.

Proof. Suppose that

$v_1 = (v_{11}, v_{12}, \ldots, v_{1n})$

$v_2 = (v_{21}, v_{22}, \ldots, v_{2n})$

$\vdots$

$v_r = (v_{r1}, v_{r2}, \ldots, v_{rn})$

and consider the equation $k_1v_1 + k_2v_2 + \cdots + k_rv_r = 0$. If we express both sides of this equation in terms of components and then equate the corresponding components, we obtain the system

$v_{11}k_1 + v_{21}k_2 + \ldots + v_{r1}k_r = 0$

$v_{12}k_1 + v_{22}k_2 + \ldots + v_{r2}k_r = 0$

$\vdots$

$v_{1n}k_1 + v_{2n}k_2 + \ldots + v_{rn}k_r = 0$

This is a homogeneous system of n equations in the r unknowns $k_1, \ldots, k_r$. Since $r > n$, it follows that the system has nontrivial solutions. Therefore, $S = \{v_1, v_2, \ldots, v_r\}$ is a linearly dependent set.

Problems

1. Determine whether the set $\{(3, -2, 5), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}$ is linearly independent in R^3.

2. Determine whether the set $\{(6, 2, 3, 4), (0, 5, -3, 1), (0, 0, 7, -2)\}$ is linearly independent in R^4.

3. Determine whether the set $\{(2, 6, -4), (3, 9, -6)\}$ is linearly independent in R^3.

4. Determine whether the set $\{1 + x, x + x^2, x^2 + 1\}$, a subset of P_2 is linearly independent or not.

2.6 Wronskian

If $f_1 = f_1(x), f_2 = f_2(x), \ldots, f_n = f_n(x)$ are functions that are n - 1 times differentiable on the interval $(-\infty, \infty)$, then the determinant

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \ldots & f_n(x) \\ f_1'(x) & f_2'(x) & \ldots & f_n'(x) \\ \vdots & \vdots & & \vdots \\ f_1^{n-1}(x) & f_2^{n-1}(x) & \ldots & f_n^{n-1}(x) \end{vmatrix}$$

is called the **Wronskian** of $f_1, f_2, \ldots, f_n$.

Theorem 2.11 If the functions $f_1, f_2, \ldots, f_n$ have n-1 continuous derivatives on the interval $(-\infty, \infty)$, and if the Wronskian of these functions is not identically zero on $(-\infty, \infty)$, then these functions form a linearly independent set of vectors in $C^{(n-1)}(-\infty, \infty)$.

Proof. Suppose that $f_1 = f_1(x), f_2 = f_2(x), \ldots, f_n = f_n(x)$ are linearly dependent vectors in $C^{(n-1)}(-\infty, \infty)$ and also suppose that the Wronskian of these functions is not identically zero on $(-\infty, \infty)$. This implies that the vector equation

$k_1 f_1 + k_2 f_2 + \cdots + k_n f_n = 0$

is satisfied by values of the coefficients $k_1, \ldots, k_n$ that are not all zero, and for these coefficients the equation

$k_1 f_1(x) + k_2 f_2(x) + \cdots + k_n f_n(x) = 0$

is satisfied for all x in $(-\infty, \infty)$. Using this equation together with those that result by differentiating it n - 1 times we obtain the linear system

$$\begin{array}{lllll} k_1 f_1(x) & +k_2 f_2(x) & +\dots & +k_n f_n(x) & =0 \\ k_1 f_1'(x) & +k_2 f_2'(x) & +\dots & +k_n f_n'(x) & =0 \\ \vdots & \vdots & & \vdots & \vdots \\ k_1 f_1^{n-1}(x) & +k_2 f_2^{n-1}(x) & +\dots & +k_n f_n^{n-1}(x) & =0 \end{array}$$

This can be written as

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

Then the determinant of the coefficient matrix is the Wronskian, W, of $f_1, f_2, \dots, f_n$. Since W is not identically zero on $(-\infty, \infty)$, the coefficient matrix is invertible and the above linear system has only the trivial solution., that is $k_1 = 0, k_2 = 0, \dots, k_n = 0$. Thus the functions $f_1, f_2, \dots, f_n$ form a linearly independent set of vectors in $C^{(n-1)}(-\infty, \infty)$.

Examples.

1. The functions $f_1 = x$ and $f_2 = sinx$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

The Wronskian is $W(x) = \begin{vmatrix} x & sinx \\ 1 & cosx \end{vmatrix} = xcosx - sinx$ This function is not identically zero on the interval $(-\infty, \infty)$ since, for example, $W(\pi/2) = \pi/2\ cos(\pi/2) - sin(\pi/2) = \pi/2$. Thus, the functions are linearly independent.

2. The functions $f_1 = 1, f_2 = e^x$, and $f_3 = e^{2x}$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

The Wronskian is $W(x) = \begin{vmatrix} 1 & e^x & e^{2x} \\ 0 & e^x & 2e^{2x} \\ 0 & e^x & 4e^{2x} \end{vmatrix} = 2e^{3x}$ This function is obviously not identically zero on $(-\infty, \infty)$, so f_1, f_2, and f_3 form a linearly independent set.

2.7 Coordinates and Basis

A vector space V is said to be **finite-dimensional** if there is a finite set of vectors in V that spans V and is said to be **infinite-dimensional** if no such set exists.

Examples of finite dimensional vector spaces are R^n, P_n, M_{mn}.

Examples of infinite dimensional vector spaces are $R^\infty, P_\infty, F(-\infty, \infty), C($

If $S = \{v_1, v_2, \ldots, v_n\}$ is a set of vectors in a finite-dimensional vector space V , then S is called a **basis** for V if:

(a) S spans V.

(b) S is linearly independent.

Examples

1. Consider the standard unit vectors $S = e_1 = (1, 0, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, 0, \ldots, 1)$ of R^n. We have proved that $spanS = R^n$ and also S is linearly independent. Thus, S is a basis for R^n that we call **the standard basis for** R^n. In particular, $i = (1, 0, 0), j = (0, 1, 0), k = (0, 0, 1)$ is the standard basis for R^3.

2. $S = 1, x, x^2, \ldots, x^n$ is a basis for the vector space P_n of polynomials of degree n or less.

For that we must show that the polynomials in S are linearly independent and span P_n. Let us denote these polynomials by $p_0 = 1, p_1 = x, p_2 = x^2, \ldots, p_n = x^n$ We have showed that these vectors span P_n and that they are linearly independent. Thus, they form a basis for P_n that we call **the standard basis for**

P_n.

3. The vectors $v_1 = (1,2,1), v_2 = (2,9,0)$, and $v_3 = (3,3,4)$ form a basis for R^3.

To show this, we must show that these vectors are linearly independent and span R^3. To prove linear independence we must show that the vector equation

$c_1v_1 + c_2v_2 + c_3v_3 = 0$

has only the trivial solution and to prove that the vectors span R^3 we must show that every vector $b = (b_1, b_2, b_3)$ in R^3 can be expressed as

$c_1v_1 + c_2v_2 + c_3v_3 = b.$

By equating corresponding components on the two sides, these two equations can be expressed as the linear systems

$$\begin{array}{lcl} c_1 + 2c_2 + 3c_3 = 0 & & c_1 + 2c_2 + 3c_3 = b_1 \\ 2c_1 + 9c_2 + 3c_3 = 0 & \quad and \quad & 2c_1 + 9c_2 + 3c_3 = b_2 \\ c_1 \qquad\quad + 4c_3 = 0 & & c_1 \qquad\quad + 4c_3 = b_3 \end{array}$$

Thus we want to show that the homogeneous system has only the trivial solution and that the nonhomogeneous system is consistent for all values of b_1, b_2, and b_3. But the two systems have the same coefficient matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 3 \\ 1 & 0 & 4 \end{bmatrix}$$

and $det(A) = -1 \neq 0$. Hence the proof and thus the vectors v_1, v_2, and v_3 form a basis for R^3.

4. The matrices $M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, M_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, M_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, M_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ form a basis for the vector space M_{22} of 2×2 matrices.

For this we want to show that the matrices are linearly independent and span M_{22}. To prove linear independence we must show that the equation

$c_1M_1 + c_2M_2 + c_3M_3 + c_4M_4 = 0$

has only the trivial solution, where 0 is the 2×2 zero matrix; and to prove that the matrices span M_{22} we must show that every 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ can be expressed as

$c_1M_1 + c_2M_2 + c_3M_3 + c_4M_4 = B$

The matrix forms of the above two equations are

$$c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and $c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

which can be rewritten as $\begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and

$\begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Since the first equation has only the trivial solution $c_1 = c_2 = c_3 = c_4 = 0$ the matrices are linearly independent, and since the second equation has the solution $c_1 = a, c_2 = b, c_3 = c, c_4 = d$ the matrices span M_{22}. This proves that the matrices M_1, M_2, M_3, M_4 form a basis for M_{22}. More generally, the mn different matrices whose entries are zero except for a single entry of 1 form a basis for M_{mn} called the standard basis for M_{mn}.

The simplest of all vector spaces is the zero vector space $V = \{0\}$. *This space is finite-dimensional because it is spanned by the vector 0. However, it has no basis because* $\{0\}$ *is not a linearly independent set. However, we will define the empty set*

$\emptyset$ to be a basis for this vector space.

5. The vector space of P_∞ of all polynomials with real coefficients is infinite dimensional. (We will show this by P_∞ has no finite spanning set.)

If there were a finite spanning set, say $S = \{p_1, p_2, \ldots, p_r\}$, then the degrees of the polynomials in S would have a maximum value, say n; and this in turn would imply that any linear combination of the polynomials in S would have degree at most n. Thus, there would be no way to express the polynomial x^{n+1} as a linear combination of the polynomials in S, contradicting the fact that the vectors in S span P_∞.

Theorem 2.12 Uniqueness of Basis Representation

If $S = \{v_1, v_2, \ldots, v_n\}$ is a basis for a vector space V, then every vector v in V can be expressed in the form $v = c_1v_1 + c_2v_2 + \ldots + c_nv_n$ in exactly one way.

Proof. Since S spans V, every vector in V is expressible as a linear combination of the vectors in S. To see that there is only one way to express a vector as a linear combination of the vectors in S, suppose that some vector v can be written as $v = c_1v_1+c_2v_2+\ldots+c_nv_n$ and also as $v = k_1v_1+k_2v_2+\ldots+k_nv_n$.

Subtracting the second equation from the first gives $0 = (c_1 - k_1)v_1 + (c_2 - k_2)v_2 + \ldots + (c_n - k_n)v_n$.

Since the right side of this equation is a linear combination of vectors in S, the linear independence of S implies that $c_1 - k_1 = 0, c_2 - k_2 = 0, \ldots, c_n - k_n = 0$ that is, $c_1 = k_1, c_2 = k_2, \ldots, c_n = k_n$.

Thus, the two expressions for v are the same.

If $S = \{v_1, v_2, \ldots, v_n\}$ is a basis for a vector space V, and $v = c_1v_1 + c_2v_2 + \ldots + c_nv_n$ is the expression for a vector v in terms of the basis S, then the scalars $c_1, c_2, \ldots, c_n$ are called the **coordinates** of v relative to the basis S. The vector $(c_1, c_2, \ldots, c_n)$ in R^n constructed from these coordinates is called the **coordinate vector of v relative to S**; it is denoted by $(v)_S = (c_1, c_2, \ldots, c_n)$.

Remark. It is standard to regard two sets to be the same if they have the same members, even if those members are written in a different order. In particular, in a basis for a vector space V, which is a set of linearly independent vectors that span V, the order in which those vectors are listed does not generally matter. However, the order in which *they are listed is critical fo r coordinate vectors, since changing the order of the basis vectors changes the coordinate vectors. To deal with this complication, many authors define an* ***ordered basis*** *to be one in which the listing order of the basis vectors remains fixed. In all discussions involving coordinate vectors we will assume that the underlying basis is ordered, even though we may not say so explicitly. Observe that* $(v)_S$ *is a vector in* R^n*, so that once an ordered basis S is given for a vector space V, Theorem 2.12 establishes a one-to-one correspondence between vectors in V and vectors in* R^n

Examples

1. In the special case where $V = R^n$ and S is the standard basis, the coordinate vector $(v)_S$ and the vector v are the same; that is,$v = (v)_S$. For example, in R^3 the representation of a vector $v = (a, b, c)$ as a linear combination of the vectors in the standard basis $S = \{i, j, k\}$ is $v = ai + bj + ck$ so the coordinate vector relative to this basis is $(v)_S = (a, b, c)$, which is the same as the vector v.

2. Find the coordinate vector for the polynomial $p(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n$ relative to the standard basis for the vector space P_n.

Here $p(x)$ is expressed in the polynomial form as a linear combination of the standard basis vectors $S = 1, x, x^2, \ldots, x^n$. Thus, the coordinate vector for p relative to S is $(p)S = (c_0, c_1, c_2, \ldots, c_n)$.

3. Find the coordinate vector of $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ relative to the standard basis for M_{22}.

We have showed that the representation of a vector $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ as a linear combination of the standard basis vectors is $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ so the coordinate vector of B relative to S is $(B)_S = (a, b, c, d)$.

4. The vectors $v_1 = (1, 2, 1), v_2 = (2, 9, 0), v_3 = (3, 3, 4)$ form a basis for R_3. Find the coordinate vector of $v = (5, -1, 9)$ relative to the basis $S = \{v_1, v_2, v_3\}$. Also find the vector v in R_3 whose coordinate vector relative to S is $(v)_S = (-1, 3, 2)$.

To find $(v)_S$ we must first express v as a linear combination of the vectors in S; that is, we must find values of c_1, c_2, and c_3 such that $v = c_1v_1 + c_2v_2 + c_3v_3$ or, in terms of components, $(5, -1, 9) = c_1(1, 2, 1) + c_2(2, 9, 0) + c_3(3, 3, 4)$ Equating corresponding components gives

$c_1 + 2c_2 + 3c_3 = 5$

$2c_1 + 9c_2 + 3c_3 = -1$

$c_1 + 4c_3 = 9$

Solving this system we obtain $c_1 = 1, c_2 = -1, c_3 = 2$. Therefore, $(v)_S = (1, -1, 2)$.

Using the definition of $(v)_S$, we obtain $v = (-1)v_1 + 3v_2 + 2v_3 = (-1)(1, 2, 1) + 3(2, 9, 0) + 2(3, 3, 4) = (11, 31, 7)$.

5. Find the coordinate vector of $v = (5, -1, 9)$ relative to the basis $B = \{(1, 0, 1), (1, 1, 0), (0, 1, 1)\}$ of R_3. Also find the vector u in R_3 whose coordinate vector relative to B is $(u)_B = (-1, 3, 2)$.

Answer. $(v)_B = (2, 1, 3)$ and $u = (2, 5, 1)$.

2.8 Dimension

Theorem 2.13 Let V be an n-dimensional vector space, and let $\{v_1, v_2, \ldots, v_n\}$ be any basis.

(a) If a set in V has more than n vectors, then it is linearly dependent.

(b) If a set in V has fewer than n vectors, then it does not span V.

Proof. (a) Let $S' = \{w_1, w_2, \ldots, w_m\}$ be any set of m vectors in V, where m>n. We want to show that S' is linearly dependent. Since $S = \{v_1, v_2, \ldots, v_n\}$ is a basis, each w_i can be expressed as a linear combination of the vectors in S, say

$w_1 = a_{11}v_1 + a_{21}v_2 + \ldots + a_{n1}v_n$

$w_2 = a_{12}v_1 + a_{22}v_2 + \ldots + a_{n2}v_n$

$\vdots$

$w_m = a_{1m}v_1 + a_{2m}v_2 + \ldots + a_{nm}v_n \qquad (1)$

To show that S' is linearly dependent, we must find scalars $k_1, k_2, \ldots, k_m$, not all zero, such that

$k_1w_1 + k_2w_2 + \cdots + k_mw_m = 0 \qquad (2).$

Then the equations in (1) can be rewritten in the partitioned form

$$\left[\begin{array}{c|c|c|c} w_1 & w_2 & \ldots & w_m \end{array}\right] = \left[\begin{array}{c|c|c|c} v_1 & v_2 & \ldots & v_m \end{array}\right] \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{bmatrix} \qquad (3)$$

Since m>n, the linear system

$$\begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (4)$$

has more equations than unknowns and hence has a nontrivial solution $x_1 = k_1, x_2 = k_2, ..., x_m = k_m$. Creating a column vector from this solution and multiplying both sides of (3) on the right by this vector yields $\left[\begin{array}{c|c|c|c} w_1 & w_2 & \dots & w_m \end{array}\right] \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix}$

$$= \left[\begin{array}{c|c|c|c} v_1 & v_2 & \dots & v_m \end{array}\right] \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix}.$$

By (4), this simplifies to

$$\left[\begin{array}{c|c|c|c} w_1 & w_2 & \dots & w_m \end{array}\right] \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

which we can rewrite as $k_1 w_1 + k_2 w_2 + \cdots + k_m w_m = 0$. Since the scalar coefficients in this equation are not all zero, we have proved that $S' = w_1, w_2, ..., w_m$ is linearly independent.

(b) Similar to (a).

Theorem 2.14 All bases for a finite-dimensional vector space have the same number of vectors.

Proof. By Theorem 2.13, if $S = v_1, v_2, \ldots, v_n$ is an arbitrary basis for V, then the linear independence of S implies that any set in V with more than n vectors is linearly dependent and any set in V with fewer than n vectors does not span V. Thus, unless a set in V has exactly n vectors it cannot be a basis.

The **dimension** of a finite-dimensional vector space V is denoted by $dim(V)$ and is defined to be the number of vectors in a basis for V. In addition, the zero vector space is defined to have dimension zero.

Examples

1. Dimensions of some familiar vector spaces

$dim(R^n)$ = n [The standard basis has n vectors.]

$dim(P_n)$ = n + 1 [The standard basis has n + 1 vectors.]

$dim(M_{mn})$ = mn [The standard basis has mn vectors.]

2. If $S = \{v_1, v_2, \ldots, v_r\}$ then every vector in span(S)is expressible as a linear combination of the vectors in S. Thus, if the vectors in S are linearly independent, they automatically form a basis for span(S), from which we can conclude that $dim(span(S)) = dim[span\{v_1, v_2, \ldots, v_r\}] = r.$

In words, the dimension of the space spanned by a linearly independent set of vectors is equal to the number of vectors in that set.

3. Find a basis for and the dimension of the solution space of the homogeneous system $x_1 + 3x_2 - 2x_3 + 2x_5 = 0$

$2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 = 0$

$5x_3 + 10x_4 + 15x_6 = 0$

$2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 = 0$

After solving, the solution of this system will be $x_1 = -3r - 4s - 2t, x_2 = r, x_3 = -2s, x_4 = s, x_5 = t, x_6 = 0$ which can be written in vector form as $(x_1, x_2, x_3, x_4, x_5, x_6) = (-3r - 4s - 2t, r, -2s, s, t, 0)$ or, alternatively, as $(x_1, x_2, x_3, x_4, x_5, x_6) = r(-3, 1, 0, 0, 0, 0) + s(-4, 0, -2, 1, 0, 0) + t(-2, 0, 0, 0, 1, 0)$. This shows that the vectors $v_1 = (-3, 1, 0, 0, 0, 0), v_2 = (-4, 0, -2, 1, 0, 0), v_3 = (-2, 0, 0, 0, 1, 0)$ span the solution space. These vectors are linearly independent by showing that none of them is a linear combination of the other two. Thus, the solution space has dimension 3.

Theorem 2.15 Plus/Minus Theorem

Let S be a nonempty set of vectors in a vector space V.

(a) If S is a linearly independent set, and if v is a vector in V that is outside of span(S), then the set S ∪ {v} that results by inserting v into S is still linearly independent.

(b) If v is a vector in S that is expressible as a linear combination of other vectors in S, and if S - v denotes the set obtained by removing v from S, then S and S - v span the same space; that is, span(S) = span(S - v)

Proof. (a) Assume that $S = \{v_1, v_2, \ldots, v_r\}$ is a linearly independent set of vectors in V, and v is a vector in V that is outside of span(S). To show that $S = v_1, v_2, \ldots, v_r, v$ is a linearly independent set, we must show that the only scalars that satisfy $k_1 v_1 + k_2 v_2 + \ldots + k_r v_r + k_{r+1} v = 0$ (1)

are $k_1 = k_2 = \ldots = k_r = k_{r+1} = 0$. But it must be true that $k_{r+1} = 0$ for otherwise we could solve (1) for v as a linear combination of $v_1, v_2, \ldots, v_r$, contradicting the assump-

tion that v is outside of span(S). Thus, (1) simplifies to
$k_1 v_1 + k_2 v_2 + \ldots + k_r v_r = 0 \qquad (2)$
which, by the linear independence of $v_1, v_2, \ldots, v_r$, implies that
$k_1 = k_2 = \ldots = k_r = 0$
(b) Assume that $S = \{v_1, v_2, \ldots, v_r\}$ is a set of vectors in V, and suppose that v_r is a linear combination of $v_1, v_2, \ldots, v_{r-1}$, say $v_r = c_1 v_1 + c_2 v_2 + \ldots + c_{r-1} v_{r-1} \qquad (3)$.
We want to show that if v_r is removed from S, then the remaining set of vectors $v_1, v_2, \ldots, v_{r-1}$ still spans S; that is, we must show that every vector w in span(S) is expressible as a linear combination of $v_1, v_2, \ldots, v_{r-1}$. But if w is in span(S), then w is expressible in the form $w = k_1 v_1 + k_2 v_2 + \ldots + k_{r-1} v_{r-1} + k_r v_r$ or, on substituting (3),
$w = k_1 v_1 + k_2 v_2 + \ldots + k_{r-1} v_{r-1} + k_r (c_1 v_1 + c_2 v_2 + \ldots + c_{r-1} v_{r-1})$
which expresses w as a linear combination of $v_1, v_2, \ldots, v_{r-1}$.

Example. Show that $p_1 = 1 - x^2, p_2 = 2 - x^2$, and $p_3 = x^3$ are linearly independent vectors.

The set $S = p_1, p_2$ is linearly independent since neither vector in S is a scalar multiple of the other. Since the vector p_3 cannot be expressed as a linear combination of the vectors in

S, it can be adjoined to S to produce a linearly independent set $S \cup p_3 = p_1, p_2, p_3$.

Theorem 2.16 Let V be an n-dimensional vector space, and let S be a set in V with exactly n vectors. Then S is a basis for V if and only if S spans V or S is linearly independent.

Proof. Assume that S has exactly n vectors and spans V. To prove that S is a basis, we must show that S is a linearly independent set. But if this is not so, then some vector v in S is a linear combination of the remaining vectors. If we remove this vector from S, then it follows from Theorem 2.15(b) that the remaining set of n - 1 vectors still spans V. But this is impossible since Theorem 2.13(b) states that no set with fewer than n vectors can span an n-dimensional vector space. Thus S is linearly independent.

Assume that S has exactly n vectors and is a linearly independent set. To prove that S is a basis, we must show that S spans V. But if this is not so, then there is some vector v in V that is not in span(S). If we insert this vector into S, then it follows from Theorem 2.15(a) that this set of n + 1 vectors is still linearly independent. But this is impossible, since Theorem

2.13(a) states that no set with more than n vectors in an n-dimensional vector space can be linearly independent. Thus S spans V.

1. Explain why the vectors $v_1 = (-3, 7)$ and $v_2 = (5, 5)$ form a basis for R^2.

Since neither vector is a scalar multiple of the other, the two vectors form a linearly independent set in the two-dimensional space R^2, and hence they form a basis by Theorem 2.16.

2. Explain why the vectors $v_1 = (2, 0, -1), v_2 = (4, 0, 7)$, and $v_3 = (-1, 1, 4)$ form a basis for R^3.

The vectors v_1 and v_2 form a linearly independent set in the xz-plane. The vector v_3 is outside of the xz-plane, so the set v_1, v_2, v_3 is also linearly independent. Since R^3 is three-dimensional, Theorem 2.16 implies that v_1, v_2, v_3 is a basis for the vector space R^3.

Theorem 2.17 Let S be a finite set of vectors in a finite-dimensional vector space V.

(a) If S spans V but is not a basis for V, then S can be reduced to a basis for V by removing appropriate vectors from S.

(b) If S is a linearly independent set that is not already a basis for V, then S can be enlarged to a basis for V by inserting appropriate vectors into S.

Proof. (a) If S is a set of vectors that spans V but is not a basis for V, then S is a linearly dependent set. Thus some vector v in S is expressible as a linear combination of the other vectors in S. By the Plus/Minus Theorem 2.15(b), we can remove v from S, and the resulting set S will still span V. If S is linearly independent, then S is a basis for V, and we are done. If S is linearly dependent, then we can remove some appropriate vector from S to produce a set S that still spans V. We can continue removing vectors in this way until we finally arrive at a set of vectors in S that is linearly independent and spans V. This subset of S is a basis for V.

(b) Suppose that dim(V) = n. If S is a linearly independent set that is not already a basis for V, then S fails to span V, so there is some vector v in V that is not in span(S). By the Plus/Minus Theorem 2.15 (a), we can insert v into S, and the resulting set S will still be linearly independent. If S spans V, then S is a basis for V, and we are finished. If S does not span V, then we can insert an appropriate vector into S to produce a set S that is still linearly independent. We can continue inserting vectors in this way until we reach a set with n linearly independent vectors in V. This set will be a basis for V by Theorem 2.16.

Theorem 2.18 If W is a subspace of a finite-dimensional vector space V, then:

(a) W is finite-dimensional.

(b) $\dim(W) \leq \dim(V)$.

(c) $W = V$ if and only if $\dim(W) = \dim(V)$.

Proof. (a) Let V be a finite dimensional vector space and W be a subspace of it. Suppose dim(V)=n and S be a basis for V. Then the number of elements in S = n. Since W is a subspace of V, W has at most n vectors. Hence if B be a basis of W, B has at most n vectors. So W is finite-dimensional.

(b) Part (a) shows that W is finite-dimensional, so it has a basis $S = \{w_1, w_2, \ldots, w_m\}$. Either S is also a basis for V or it is not. If so, then dim(V) = m, which means that dim(V) = dim(W). If not, then because S is a linearly independent set it can be enlarged to a basis for V by part (b) of Theorem 2.17.

But this implies that dim(W) $<$ dim(V),so we have shown that dim(W) $\leq$ dim(V) in all cases.

(c) Assume that dim(W) = dim(V) and that S = $\{w_1, w_2, \ldots, w_m\}$ is a basis for W. If S is not also a basis for V, then being linearly independent S can be extended to a basis for V by part (b) of Theorem 2.17. But this would mean that dim(V) $>$ dim(W), which contradicts our hypothesis. Thus S must also be a basis for V, which means that W = V . The converse is obvious.

Problems

1. Find a basis for the solution space of the following homogeneous linear system and find the dimension of that space:

$x - 2y + z - w = 0$

$x + y - 2z + 3w = 0$

$4x + y - 5z + 8w = 0$

$5x - 7y + 2z - w = 0.$

The augmented matrix (same as the coefficient matrix) for the above system is

$$\begin{bmatrix} 1 & -2 & 1 & -1 \\ 1 & 1 & -2 & 3 \\ 4 & 1 & -5 & 8 \\ 5 & -7 & 2 & -1 \end{bmatrix}.$$

After reducing to row echelon form we get

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

The general solution is obtained as

$x = r - 5/3s, \quad y = r - 4/3s, \quad z = r, \quad w = s.$

This can be written in the vector form as

$(x, y, z, w) = (r - 5/3s,\ r - 4/3s,\ r,\ s)$

$= r(1, 1, 1, 0) + s(-5/3, -4/3, 0, 1)$

$= r(1, 1, 1, 0) - \frac{1}{3}s(5, 4, 0, -3).$

This shows that the vectors

$v_1 = (1, 1, 1, 0)$ and $v_2 = (5, 4, 0, -3)$ span the solution space.

also for any scalars k_1 and k_2

$k_1v_1 + k_2v_2 = 0 \Rightarrow k_1(1, 1, 1, 0) + k_2(5, 4, 0, -3) = (0, 0, 0, 0) \Rightarrow$

$k_1 + 5k_2 = 0, k_1 + 4k_2 = 0, k_1 = 0, -3k_2 = 0 \Rightarrow k_1 = k_2 = 0.$

Hence the vectors v_1 and v_2 are linearly independent and so

$\{v_1, v_2\}$ form the basis for the solution space. Thus the solution space is of dimension 2.

2. Find the basis of the subspace $W = \{(a, b, c, d) : c = a - b \quad d = a + b\}$ of R^4 and its dimension.

Any vector (a, b, c, d) in W can be written as $(a, b, c, d) = (a, b, a - b, a + b) = a(1, 0, 1, 1) + b(0, 1, -1, 1)$.

$W = span\{(1, 0, 1, 1), (0, 1, -1, 1)\}$.

For any scalars k_1 and k_2,

$k_1(1, 0, 1, 1) + k_2(0, 1, -1, 1) = (0, 0, 0, 0) \Rightarrow (k_1, k_2, k_1 - k_2, k_1 + k_2) = (0, 0, 0, 0) \Rightarrow k_1 = 0, k_2 = 0$.

Hence $\{(1, 0, 1, 1), (0, 1, -1, 1)\}$ is linearly independent and so is a basis for W. Since the basis has two elements, $dim(W) = 2$.

3. Let $S = \{2, x, x - x^2, x + x^2\}$ be a subset of P_2. Find the dimension of $span(S)$.

Any element $p(x)$ of $span(S)$ can be written as

$p(x) = a2 + bx + c(x - x^2) + d(x + x^2)$ for some $a, b, c, d \in R$.

Then $p(x) = a2 + bx + c(x - x^2) + d(x + x^2)$

$= (2a)1 + (b + c + d)x + (d - c)x^2$.

Hence $span(S) = \{1, x, x^2\}$.

Also for any scalars $l, m, n, \quad l1 + mx + nx^2 = 0 \Rightarrow l = m =$

$n = 0$. Hence $\{1, x, x^2\}$ is linearly independent.

Since $\{1, x, x^2\}$ is linearly independent and $\{1, x, x^2\} = span(S)$, $\{1, x, x^2\}$ is a basis of $span(S)$. So $dim(span(S)) = 3$.

MODULE

THREE

GENERAL VECTOR SPACES CONTINUED

3.1 Change of Basis

If $S = v_1, v_2, \ldots, v_n$ is a basis for a finite-dimensional vector space V, and if

$(v)_S = (c_1, c_2, \ldots, c_n)$

is the coordinate vector of v relative to S, then, the mapping

$v \to (v)_S$

creates a one-to-one correspondence between vectors in the general vector space V and vectors in the Euclidean vector space R^n. We call this map the coordinate map relative to S from V to R^n. We will express coordinate vectors in the matrix form

$[v]_S = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ where the square brackets emphasize the matrix notation.

The Change-of-Basis Problem *If v is a vector in a finite-dimensional vector space V, and if we change the basis for V from a basis B to a basis B', how are the coordinate vectors* $[v]_B$ *and* $[v]_{B'}$ *related?*

To solve this problem, it will be convenient to refer to B as the "old basis" and B' as the "new basis." Thus, our objective is to find a relationship between the old and new coordinates of a fixed vector v in V.

For convenience we solve this problem for two-dimensional spaces. The solution for n-dimensional spaces is similar. Let $B = \{u_1, u_2\}$ and B' $= \{v_1, v_2\}$ be the old and new bases, respectively. Suppose the coordinate vectors for the new basis vectors relative to the old basis are

$[v_1]_B = \begin{bmatrix} a \\ b \end{bmatrix}$ and $[v_2]_B = \begin{bmatrix} c \\ d \end{bmatrix}$

. That is $v_1 = au_1 + bu_2$ and $v_2 = cu_1 + du_2$.

Now let v be any vector in V, and let $[v]_{B'} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ be the new coordinate vector, so that $v = k_1 v_1 + k_2 v_2$. In order to find the old coordinates of v, we must express v in terms of the old basis B. Substituting, $v = k_1(au_1 + bu_2) + k_2(cu_1 + du_2)$ or $v = (k_1 a + k_2 c)u_1 + (k_1 b + k_2 d)u_2$ Thus, the old coordinate vector for v is $[v]_B = \begin{bmatrix} k_1 a + k_2 c \\ k_1 b + k_2 d \end{bmatrix}$ which can be written as

$$[v]_B = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} [v]_{B'}.$$

This equation states that the old coordinate vector $[v]_B$ results when we multiply the new coordinate vector $[v]_{B'}$ on the left by the matrix $P = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$.

Since the columns of this matrix are the coordinates of the new basis vectors relative to the old basis, we have the following solution of the change-of-basis problem.

Solution of the Change-of-Basis Problem *If we change the basis for a vector space V from an old basis* $B = u_1, u_2, \ldots, u_n$ *to a new basis* $B' = v_1, v_2, \ldots, v_n$*, then for each vector v in V, the old coordinate vector* $[v]_B$ *is related to the new coordinate vector* $[v]_{B'}$ *by the equation*

$$[v]_B = P[v]_{B'} \qquad (1)$$

where the columns of P are the coordinate vectors of the new basis vectors relative to the old basis; that is, the column vectors of P are $[v_1]_B, [v_2]_B, \ldots, [v_n]_B.$ (2)

The matrix P in Equation (1) is called the **transition matrix** from B' to B. For emphasis, we will often denote it by $P_{B' \to B}$ It follows from (2) that this matrix can be expressed in terms of its column vectors as

$$P_{B' \to B} = \left[\ [v_1]_B \ \middle|\ [v_2]_B \ \middle|\ \ldots \ \middle|\ [v_n]_B\ \right]. \qquad (3)$$

Similarly, the transition matrix from B to B' can be expressed in terms of its column vectors as

$$P_{B \to B'} = \left[\ [u_1]_{B'} \ \middle|\ [u_2]_{B'} \ \middle|\ \ldots \ \middle|\ [u_n]_{B'}\ \right]. \qquad (4)$$

Remark. There is a simple way to remember both of these formulas using the terms "old basis" and "new basis" : In Formula (3) the old basis is B' and the new basis is B, whereas in Formula (4) the old basis is B and the new basis is B'. Thus, both formulas can be restated as follows:

The columns of the transition matrix from an old basis to a new basis are the coordinate vectors of the old basis relative to the new basis.

Examples

1. Consider the bases $B = \{u_1, u_2\}$ and B' $= \{v_1, v_2\}$ for R^2, where $u_1 = (1,0), u_2 = (0,1), v_1 = (1,1), v_2 = (2,1)$.

(a) Find the transition matrix $P_{B' \to B}$ from B' to B.

(b) Find the transition matrix $P_{B \to B'}$ from B to B'.

(a) Here the old basis vectors are v_1 and v_2 and the new basis vectors are u_1 and u_2. We want to find the coordinate matrices of the old basis vectors v_1 and v_2 relative to the new basis vectors u_1 and u_2. To do this, observe that

$v_1 = u_1 + u_2$

$v_2 = 2u_1 + u_2$

from which it follows that

$[v_1]_B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $[v_2]_B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

and hence that $P_{B' \to B} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$.

(b) Here the old basis vectors are u_1 and u_2 and the new basis vectors are v_1 and v_2 we want to find the coordinate matrices of the old basis vectors v_1 and v_2 relative to the new basis vectors u_1 and u_2. To do this, observe that

$u_1 = -v_1 + v_2$

$u_2 = 2v_1 - v_2$

from which it follows that

$[u_1]_{B'} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $[u_2]_{B'} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

and hence that $P_{B \to B'} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$.

Suppose now that B and B' are bases for a finite-dimensional vector space V. Since multiplication by $P_{B' \to B}$ maps coordinate vectors relative to the basis B' into coordinate vectors relative to a basis B, and $P_{B \to B'}$ maps coordinate vectors relative to B into coordinate vectors relative to B', it *follows that for every vector v in V we have*

$[v]_B = P_{B' \to B}[v]_{B'}$ (5) and $[v]_{B'} = P_{B \to B'}[v]_B = P_{B'}$ (6)

2. Let B and B' be the bases in Example 1. Use an appropriate formula to find $[v]_B$ given that $[v]_{B'} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$.

To find $[v]_B$ we need to make the transition from B' to B. It follows from Formula (5) and part (a) of Example 1 that

$$[v]_B = P_{B' \to B}[v]_{B'} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -3 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 2 \end{bmatrix}.$$

If B and B' are bases for a finite-dimensional vector space V, then $P_{B' \to B}P_{B \to B'} = P_{B \to B}$ because multiplication by the product $P_{B' \to B}P_{B \to B'}$ first maps the B-coordinates of a vector into its B'-coordinates, and then maps those B'-coordinates back into the original B-coordinates. Since the net effect of the two operations is to leave each coordinate vector unchanged, we are led to conclude that $P_{B \to B}$ must be the identity matrix, that is, $P_{B' \to B}P_{B \to B'} = I.$ (7)

For example, for the transition matrices obtained in Example 1 we have $P_{B' \to B}P_{B \to B'} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

It follows from (7) that $P_{B' \to B}$ is invertible and that its inverse is $P_{B \to B'}$.

Theorem 3.1 If P is the transition matrix from a basis B' to a basis B for a finite dimensional vector space V, then P is invertible and P^{-1} is the transition matrix from B to B'.

An Efficient Method for Computing Transition Matrices for R^n

Procedure for Computing $P_{B \to B'}$ **(8)**

Step 1. Form the matrix $[B'|B]$.

Step 2. Use elementary row operations to reduce the matrix in Step 1 to reduced row echelon form.

Step 3. The resulting matrix will be $[I|P_{B \to B'}]$.

Step 4. Extract the matrix $P_{B \to B'}$ from the right side of the matrix in Step 3.

Examples

3. In example 1, consider the bases $B = \{u_1, u_2\}$ and B' $= \{v_1, v_2\}$ for R^2, where $u_1 = (1, 0), u_2 = (0, 1), v_1 = (1, 1), v_2 = (2, 1)$.

(a) Use formula (8) to find transition matrix from B' to B.

(b) Use formula (8) to find transition matrix from B to B'.

(a) Here B' is the old basis and B is the new basis, so $[new\ basis|old\ basis] = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{array}\right]$ Since the left side is already the identity matrix, no reduction is needed. We see that the transition matrix is

$P_{B'\to B} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$. which agrees with the result in Example 1.

(b) Here B is the old basis and B' is the new basis, so $[new\ basis|old\ basis] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right]$ By reducing this matrix, so the left side becomes the identity, we obtain $[I|transition\ from\ old\ to\ new] = \left[\begin{array}{cc|cc} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \end{array}\right]$ so the transition matrix is $P_{B\to B'} = \begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$.

which also agrees with the result in Example 1.

Theorem 3.2 Let $B' = \{u_1, u_2, \ldots, u_n\}$ be any basis for the vector space R^n and let $S = e_1, e_2, \ldots, e_n$ be the standard basis for R^n. If the vectors in these bases are written in column form, then $P_{B'\to S} = \left[\begin{array}{c|c|c|c} u_1 & u_2 & \ldots & u_n \end{array}\right]$

If $A = \left[\begin{array}{c|c|c|c} u_1 & u_2 & \dots & u_n \end{array}\right]$ is any invertible n $\times$ n matrix, then A can be viewed as the transition matrix from the basis $\{u_1, u_2, \dots, u_n\}$ for R^n to the standard basis for R^n. Thus, for example, the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}$$

which was shown to be invertible earlier, is the transition matrix from the basis $u_1 = (1, 2, 1), u_2 = (2, 5, 0), u_3 = (3, 3, 8)$ to the basis $e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1)$.

3.2 Row Space, Column Space, and Null Space

For an m$\times$n matrix

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

the vectors

$$r_1 = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{bmatrix}$$
$$r_2 = \begin{bmatrix} a_{21} & a_{22} & \dots & a_{2n} \end{bmatrix}$$
$$\vdots \qquad\qquad \vdots$$
$$r_m = \begin{bmatrix} a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

in R^n that are formed from the rows of A are called the **row vectors** of A, and the vectors

$$c_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad c_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \quad \dots, \quad c_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

in R^m formed from the columns of A are called the **column vectors** of A.

If A is an m×n matrix, then the subspace of R^n spanned by the row vectors of A is called the **row space** of A, and the subspace of R^m spanned by the column vectors of A is called the **column space** of A. The solution space of the homogeneous system of equations Ax = 0, which is a subspace of R^n, is called the **null space** of A.

We will sometimes denote the row space of A, the column space of A, and the null space of A by row(A), col(A), and null(A), respectively.

Theorem 3.3 A system of linear equations Ax = b is consistent if and only if b is in the column space of A.

Proof. Consider the linear system Ax=b where $A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$ and $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$.

If c_1, $c_2, \ldots$, c_n denote the column vectors of A, then the product Ax can be expressed as a linear combination of these vectors with coefficients from x;

that is, $Ax = x_1c_1 + x_2c_2 + \ldots + x_nc_n$. (1)

Thus, a linear system, Ax = b, of m equations in n unknowns can be written as $x_1c_1 + x_2c_2 + \ldots + x_nc_n = b$ (2)

from which we conclude that Ax = b is consistent if and only if b is expressible as a linear combination of the column vectors of A.

Example 4. Let Ax = b be the linear system

$$\begin{bmatrix} -1 & 3 & 2 \\ 1 & 2 & -3 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$$

Show that b is in the column space of A by expressing it as a linear combination of the column vectors of A.

Solving the system by Gaussian elimination yields $x_1 = 2$, $x_2 = -1$, $x_3 = 3$. It follows from this and Formula (2) that

$$2\begin{bmatrix}-1\\1\\2\end{bmatrix} - \begin{bmatrix}3\\2\\1\end{bmatrix} + 3\begin{bmatrix}2\\-3\\-2\end{bmatrix} = \begin{bmatrix}1\\-9\\-3\end{bmatrix}.$$

The general solution of a consistent linear system Ax = b can be obtained by adding any specific solution of the system to the general solution of the corresponding homogeneous system Ax = 0. Keeping in mind that the null space of A is the same as the solution space of Ax = 0, we can rephrase that theorem in the following vector form.

Theorem 3.4 If x_0 is any solution of a consistent linear system Ax = b, and if S = $v_1, v_2, \ldots, v_k$ is a basis for the null space of A, then every solution of Ax = b can be expressed in the form

$$x = x_0 + c_1 v_1 + c_2 v_2 + \cdots + c_k v_k. \qquad (3)$$

Conversely, for all choices of scalars $c_1, c_2, \ldots, c_k$, the vector x in this formula is a solution of Ax = b.

The vector x_0 in Formula (3) is called a **particular solution** of Ax = b, and the remaining part of the formula is called the **general solution** of Ax = 0.

The general solution of a consistent linear system can be expressed as the sum of a particular solution of that system and the general solution of the corresponding homogeneous system.

Example 5. For the matrix A and the column vector b, determine whether b is in the column space of A and if so, express b as a linear combination of the column vectors of A:

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix}; b = \begin{bmatrix} 6 \\ 2 \\ 1 \end{bmatrix}.$$

Consider the linear system Ax=b where

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 2 & 1 & -1 \end{bmatrix}; x = \begin{bmatrix} x \\ y \\ z \end{bmatrix}; b = \begin{bmatrix} 6 \\ 2 \\ 1 \end{bmatrix}.$$

The augmented matrix for the above system is $\begin{bmatrix} 1 & 1 & 1 & 6 \\ 1 & -1 & 1 & 2 \\ 2 & 1 & -1 & 1 \end{bmatrix}$.

After reducing to echelon form, we get $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$.

In matrix form, $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ which is equivalent to $x = 1, y = 2, z = 3$.

Hence the linear system Ax=b is consistent. So the vector b is in the column space of A and can be expressed as a linear combination of column vectors of A as

$$\begin{bmatrix} 6 \\ 2 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

Example 6. Suppose that $x_1 = 3, x_2 = 0, x_3 = -1, x_4 = 5$ is a solution of a nonhomogeneous linear system Ax=b and that the solution set of the homogeneous system $Ax = 0$ is given by $x_1 = 5r - 2s, \quad x_2 = s, \quad x_3 = s + t, \quad x_4 = t$.

(a) Find a column vector form of the general solution of $Ax = 0$.

(b) Find a column vector form of the general solution of Ax=b.

Solution. (a) Column vector form of the general solution of the homogeneous system $Ax = 0$ is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5r-2s \\ s \\ s+t \\ t \end{bmatrix} = r\begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

(b) Column vector form of the general solution of the nonhomogeneous system Ax=b is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -1 \\ 5 \end{bmatrix} + \begin{bmatrix} 5r-2s \\ s \\ s+t \\ t \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ -1 \\ 5 \end{bmatrix} + r\begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Theorem 3.5 Elementary row operations do not change the null space of a matrix.

Theorem 3.6 Elementary row operations do not change the row space of a matrix.

Theorems 3.5 and 3.6 might tempt you into incorrectly believing that elementary row operations do not change the column space of a matrix. To see why this is not true, compare the matrices

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 3 \\ 0 & 0 \end{bmatrix}.$$

The matrix B can be obtained from A by adding -2 times the first row to the second. However, this operation has changed the column space of A, since that column space consists of all scalar multiples of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ whereas the column space of B consists of all scalar multiples of $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and the two are different spaces.

Theorem 3.7 If a matrix R is in row echelon form, then the row vectors with the leading 1's (the nonzero row vectors) form a basis for the row space of R, and the column vectors with the leading 1's of the row vectors form a basis for the column space of R.

Theorem 3.8 If A and B are row equivalent matrices, then:

(a) A given set of column vectors of A is linearly independent if and only if the corresponding column vectors of B are linearly independent.

(b) A given set of column vectors of A forms a basis for the column space of A if and only if the corresponding column vectors of B form a basis for the column space of B.

Example 7. Find a bases for the null space, row space and column space of the matrix:

$$A = \begin{bmatrix} 1 & 4 & 5 & 6 & 9 \\ 3 & -2 & 1 & 4 & -1 \\ -1 & 0 & -1 & -2 & -1 \\ 2 & 3 & 5 & 7 & 8 \end{bmatrix}.$$

Solution. The null space of A is the solution space of the homogeneous linear system $Ax = 0$. First reduce the matrix to row reduced echelon form and which in turn yields

$$\begin{bmatrix} 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Since the elememtary row operations do not change the solution space of the system $Ax = 0$, it is equal to the solution space of the system of equations

$$\begin{bmatrix} 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

which is equivalent to the system of equations

$x_1 + x_3 + 2x_4 + x_5 = 0$, $x_2 + x_3 + x_4 + 2x_5 = 0$. Assinging

$x_3 = r,\ x_4 = s\ x_5 = t$ and solving we get the general solution as $x_1 = -r - 2s - t,\ x_2 = -r - s - 2t,\ x_3 = r,\ x_4 = s\ x_5 = t.$ Thus the column vector form of the general solution of $Ax = 0$ is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -r-2s-t \\ -r-s-2t \\ r \\ s \\ t \end{bmatrix} = r\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s\begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ -2 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Hence the null space of A is spanned by the vectors

$$v_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix},\ v_2 = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix},\ v_3 \begin{bmatrix} -1 \\ -2 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Also, for any scalars $k_1, k_2, k_3,\ k_1v_1 + k_2v_2 + k_3v_3 = 0 \Rightarrow$

$$k_1\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + k_2\begin{bmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + k_3\begin{bmatrix} -1 \\ -2 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow k_1 = 0, k_2 = 0, k_3 = 0.$$

Thus $span\{v_1, v_2, v_3\} = \text{null}(A)$ and v_1, v_2, v_3 are linearly independent. Hence $\{v_1, v_2, v_3\}$ is a basis for the null space of A.

Since the elememtary row operations do not change the row space of a matrix, row space of the given matrix A is same as the row space of the matrix below:

$\begin{bmatrix} 1 & 0 & 1 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ which we have obtained from A by elementary row operations. Hence the row space of A is spanned by the non zero vectors

$r_1 = \begin{bmatrix} 1 & 0 & 1 & 2 & 1 \end{bmatrix}$ and $r_2 = \begin{bmatrix} 0 & 1 & 1 & 1 & 2 \end{bmatrix}$

of the above matrix. Also, for any scalars k_1, k_2, $k_1 r_1 + k_2 r_2 = 0 \Rightarrow k_1 \begin{bmatrix} 1 & 0 & 1 & 2 & 1 \end{bmatrix} + k_2 \begin{bmatrix} 0 & 1 & 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$\Rightarrow k_1 = 0, k_2 = 0.$

Thus $span\{r_1, r_2\} = \text{row}(A)$ and r_1, r_2 are linearly independent.

Hence $\{r_1, r_2\}$ is a basis for the row space of A.

Now, to find column space, remember that A and R can have different column spaces. So we cannot directly find the column space of A from the column vectors of R. However, it follows from Theorem 3.8(b) that if we can find a set of column vectors of R that forms a basis for the column space of R, then the corresponding column vectors of A will form a basis for the column space of A. Since the first and second columns of R contain the leading 1's of the row vectors, these vectors

$$c_1' = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } c_2' = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

foem a basis for the column space of R. Thus, the corresponding column vectors of A, that is first and second columns of A, which are

$$c_1 = \begin{bmatrix} 1 \\ 3 \\ -1 \\ 2 \end{bmatrix} \text{ and } c_2 = \begin{bmatrix} 4 \\ -2 \\ 0 \\ 3 \end{bmatrix}$$

form a basis for the column space of A.

In Example 7, we found a basis for the row space of a matrix by reducing that matrix to row echelon form. However, the basis vectors produced by that method were not all row vectors of the original matrix. To find a basis for the row space of a matrix A that consists entirely of row vectors of A, we proceed as follows:
First take the transpose of A, A^T, there by converting the row space of A into column space of A^T. appluing the method described in the Example 7.. find the basis of the basis of the column space of A^T, consisting entirely of column vectors of A^T.Transposing these vectors back to row vectors, we get a basis for the row space of A that consists entirely of row vectors from A.

Example 8. Find a basis for the row space of

$$A = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 6 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{bmatrix}$$

consisting entirely of row vectors from A.

Transposing A, we get

$$A^T = \begin{bmatrix} 1 & 2 & 0 & 2 \\ -2 & -5 & 5 & 6 \\ 0 & -3 & 15 & 18 \\ 0 & -2 & 10 & 8 \\ 3 & 6 & 0 & 6 \end{bmatrix}$$

and when reducing this matrix to row echelon form we obtain

$$\begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & 1 & -5 & -10 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The first, second, and fourth columns contain the leading 1's, so the corresponding column vectors in A^T form a basis for the column space of A^T ; these are

$$c_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 0 \\ 3 \end{bmatrix}, c_2 = \begin{bmatrix} 2 \\ -5 \\ -3 \\ -2 \\ 6 \end{bmatrix}, c_4 = \begin{bmatrix} 2 \\ 6 \\ 18 \\ 8 \\ 6 \end{bmatrix}.$$

Transposing again and adjusting the notation appropriately yields the basis vectors

$r_1 = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \end{bmatrix}$, $r_2 = \begin{bmatrix} 2 & -5 & -3 & -2 & 6 \end{bmatrix}$, $r_4 = \begin{bmatrix} 2 & 6 & 18 & 8 & 6 \end{bmatrix}$.
for the row space of A.

3.2.1 Basis for the Space Spanned by a Set of Vectors

Consider the following general problem in R^n.
Given a set of vectors $S = \{v_1, v_2, \ldots, v_k\}$ in R^n, find a subset of these vectors that forms a basis for $span(S)$, and express each vector that is not in that basis as a linear combination of the basis vectors.
The following is a summary of the steps that we follow to solve the problem posed above.
Step 1. Form the matrix A whose columns are the vectors in the set $S = \{v_1, v_2, \ldots, v_k\}$.
Step 2. Reduce the matrix A to reduced row echelon form R.
Step 3. Denote the column vectors of R by $w_1, w_2, \ldots, w_k$.
Step 4. Identify the columns of R that contain the leading

1's. The corresponding column vectors of A form a basis for $span(S)$.

This completes the first part of the problem.

Step 5. Obtain a set of dependency equations for the column vectors $w_1, w_2, \ldots, w_k$ of R by successively expressing each w_i that does not contain a leading 1 of R as a linear combination of predecessors that do.

Step 6. In each dependency equation obtained in Step 5, replace the vector w_i by the vector v_i for $i = 1, 2, \ldots, k$.

This completes the second part of the problem.

Example 9.(a) Find a subset of the vectors $v_1 = (1, -2, 0, 3), v_2 = (2, -5, -3, 6), v_3 = (0, 1, 3, 0), v_4 = (2, -1, 4, -7), v_5 = (5, -8, 1, 2)$ that forms a basis for the subspace of R^4 spanned by these vectors.

(b) Express each vector not in the basis as a linear combination of the basis vectors.

Solution (a) We begin by constructing a matrix that has v_1, v_2, v_3, v_4, v_5 as its column vectors:

$$\begin{bmatrix} 1 & 2 & 0 & 2 & 5 \\ -2 & -5 & 1 & -1 & -8 \\ 0 & -3 & 3 & 4 & 1 \\ 3 & 6 & 0 & -7 & 2 \end{bmatrix}. \qquad (1)$$

The first part of our problem can be solved by finding a basis for the column space of this matrix. Reducing the matrix to reduced row echelon form and denoting the column vectors of the resulting matrix by w_1, w_2, w_3, w_4, and w_5 yields

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (2)$$

The leading 1's occur in columns 1, 2, and 4, so by Theorem 3.7, $\{w_1, w_2, w_4\}$ is a basis for the column space of (2), and consequently, $\{v_1, v_2, v_4\}$ is a basis for the column space of (1).

(b) We will start by expressing w_3 and w_5 as linear combinations of the basis vectors w_1, w_2, w_4. The simplest way of doing this is to express w_3 and w_5 in terms of basis vectors with smaller subscripts. Accordingly, we will express w_3 as a linear combination of w_1 and w_2, and we will express w_5 as a linear combination of w_1, w_2, and w_4. By inspection of (2), these linear combinations are

$w_3 = 2w_1 - w_2$

$w_5 = w_1 + w_2 + w_4$.

We call these the **dependency equations**. The corresponding relationships in (1) are

$v_3 = 2v_1 - v_2$

$v_5 = v_1 + v_2 + v_4$.

3.3 Rank, Nullity, and the Fundamental Matrix Spaces

Theorem 3.9 The row space and the column space of a matrix A have the same dimension.

Proof. It follows from Theorems 3.6 and 3.8 (b) that elementary row operations do not change the dimension of the row space or of the column space of a matrix. Thus, if R is any row echelon form of A, it must be true that

dim(row space of A) = dim(row space of R)

dim(column space of A) = dim(column space of R) so it suffices to show that the row and column spaces of R have the same dimension. But the dimension of the row space of R is the number of nonzero rows, and by Theorem 3.7 the dimension of the column space of R is the number of leading 1's. Since these two numbers are the same, the row and column space have the same dimension.

The rank of A can be interpreted as the number of leading 1's in any row echelon form of A.

The common dimension of the row space and column space of a matrix A is called the **rank** of A and is denoted by **rank(A)**; the dimension of the null space of A is called the **nullity** of A and is denoted by **nullity(A)**.

Example 10. Find the rank and nullity of the matrix

$$A = \begin{bmatrix} -1 & 2 & 0 & 4 & 5 & -3 \\ 3 & -7 & 2 & 0 & 1 & 4 \\ 2 & -5 & 2 & 4 & 6 & 1 \\ 4 & -9 & 2 & -4 & -4 & 7 \end{bmatrix}$$

Solution. The reduced row echelon form of A is

$$\begin{bmatrix} 1 & 0 & -4 & -28 & -37 & 13 \\ 0 & 1 & -2 & -12 & -16 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (1)$$

Since this matrix has two leading 1's, its row and column spaces are two dimensional and rank(A) = 2.

To find the nullity of A, we must find the dimension of the solution space of the linear system Ax = 0. This system can be solved by reducing its augmented matrix to reduced row echelon form. The resulting matrix will be identical to (1), except that it will have an additional last column of zeros, and hence the corresponding system of equations will be

$x_1 - 4x_3 - 28x_4 - 37x_5 + 13x_6 = 0$

$x_2 - 2x_3 - 12x_4 - 16x_5 + 5x_6 = 0$

Solving these equations for the leading variables yields

$x_1 = 4x_3 + 28x_4 + 37x_5 - 13x_6$

$x_2 = 2x_3 + 12x_4 + 16x_5 - 5x_6 \qquad (2)$

from which we obtain the general solution

$x_1 = 4r + 28s + 37t - 13u$

$x_2 = 2r + 12s + 16t - 5u$

$x_3 = r$

$x_4 = s$

$x_5 = t$

$x_6 = u$

or in column vector form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = r \begin{bmatrix} 4 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 28 \\ 12 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 37 \\ 16 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + u \begin{bmatrix} -13 \\ -5 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad (3)$$

Because the four vectors on the right side of (3) form a basis for the solution space, nullity(A) $= 4$.

Example 11. What is the maximum possible rank of an m $\times$ n matrix A that is not square?

Solution. Since the row vectors of A lie in R^n and the column vectors in R^m, the row space of A is at most n-dimensional and the column space is at most m-dimensional. Since the rank of A is the common dimension of its row and column space, it follows that the rank is at most the smaller of m and n. We denote this by writing rank(A) $\leq$ min(m,n) in which min(m,n) is the minimum of m and n.

Theorem 3.11 - Dimension Theorem for Matrices

If A is a matrix with n columns, then rank(A) + nullity(A) = n.

Proof. Since A has n columns, the homogeneous linear system $Ax = 0$ has n unknowns (variables). These fall into two distinct categories: the leading variables and the free variables. Thus, number of leading variables + number of free variables = n. But the number of leading variables is the same as the number of leading 1's in any row echelon form of A, which is the same as the dimension of the row space of A, which is the same as the rank of A. Also, the number of free variables in the general solution of $Ax = 0$ is the same as the number of parameters in that solution, which is the same as the dimension of the solution space of $Ax = 0$, which is the same as the nullity of A. Hence rank(A) + nullity(A) = n.

In Example 10., the matrix A has 6 columns. Also we have showed that rank(A) = 2 and nullity(A) = 4. Thus rank(A) + nullity(A) = 6.

Theorem 3.12 If A is an $m \times n$ matrix, then

(a) rank(A) = the number of leading variables in the general solution of Ax=0.

(b) nullity(A) = the number of parameters in the general solution of Ax=0.

Example 12. (a) Find the number of parameters in the general solution of Ax=0 if A is a 5×7 matrix of rank 3.

(b) Find the rank of a 5×7 matrix A for which Ax=0 has a two-dimensional solution space.

Solution. (a) From Dimension Theorem for Matrices,

nullity(A) = n - rank(A) = 7 - 3 = 4.

Thus, there are four parameters.

(b) The matrix A has nullity 2, so

rank(A) = n - nullity(A) = 7 - 2 = 5.

Theorem 3.13 If Ax=b is a consistent linear system of m equations in n unknowns, and if A has rank r, then the general solution of the system contains n - r parameters.

Proof. Suppose Ax=b is a consistent linear system of m equations in n unknowns. Then its coefficient matrix A is an $m \times n$ matrix. We know if Ax=b is a consistent linear system, then its general solution can be expressed as the sum of a particular solution of that system and the general solution of the corresponding homogeneous system (Theorem 3.4). Since the particular solution contains no parameters, the number of parameters in the general solution of the system Ax=b is same as the number of parameters in the general solution of the system Ax=0. But the number of parameters in the general solution of Ax=0 is equal to the nullity(A) (Theorem 3.12). Hence if A has rank r, then by Dimension Theorem for Matrices, we get (the number of parameters in the general solution of Ax=b)+r=nullity(A)+rank(A)=number of columns of A=n. Hence the number of parameters in the general solution of system of equations Ax=b is n-r.

There are six important vector spaces associated with a matrix A and its transpose A^T :

row space of A	row space of A^T
column space of A	column space of A^T
null space of A	null space of A^T

However, transposing a matrix converts row vectors into column vectors and conversely, so except for a difference in notation, the row space of A^T is the same as the column space of A, and the column space of A^T is the same as the row space of A. Thus, of the six spaces listed above, only the following four are distinct:

row space of A	column space of A
null space of A	null space of A^T

These are called the **fundamental spaces** of a matrix A.

If A is an $m \times n$ matrix, then the row space and null space of A are subspaces of R^n, and the column space of A and the null space of A^T are subspaces of R^m.

Theorem 3.14 If A is any matrix, then $\text{rank}(A) = \text{rank}(A^T)$.

Proof. Since transposing a matrix converts row vectors into column vectors and conversely, except for a difference in notation, the row space of A^T is the same as the column space of A, and the column space of A^T is the same as the row space of A. Thus, we have

$\text{rank}(A) = \dim(\text{row space of } A) = \dim(\text{column space of } A^T) = \text{rank}(A^T)$.

The above result has some important implications. For example, if A is an m $\times$ n matrix, then applying Dimension Theorem for Matrices to the matrix A^T and using the fact that this matrix has m columns (A^T is of order n $\times$ m) yields

rank(A^T) + nullity(A^T) = m

which, by Theorem 3.14, can be rewritten as

rank(A) + nullity(A^T) = m

The above equation makes it possible to express the dimensions of all four fundamental spaces in terms of the size and rank of A. Specifically, if rank(A) = r, then

dim[row(A)] = r   dim[col(A)] = r

dim[null(A)] = n - r   dim[null(A^T)] = m - r.

3.3.1 A Geometric Link Between the Fundamental Spaces

If $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ are vectors in R^n, then the **dot product** (also called the **Euclidean inner product**) of u and v is denoted by **u . v** and is defined by

$u.v = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$

Two nonzero vectors u and v in R^n are said to be **orthogonal (or perpendicular)** if $u.v = 0$. Also the zero vector in R^n is orthogonal to every vector in R^n.

Recall that a linear equation in the variables $x_1, x_2, \ldots, x_n$ has the form

$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b(a_1, a_2, \ldots, a_n \textit{ not all zero})$

and that the corresponding homogeneous equation is

$a1x_1 + a2x_2 + \ldots + anx_n = 0(a_1, a_2, \ldots, a_n \textit{ not all zero})$

These equations can be rewritten in vector form by letting

$a = (a_1, a_2, \ldots, a_n)$ and $x = (x_1, x_2, \ldots, x_n)$

and can be written as $a.x = b$ and $a.x = 0$

This equation reveals that each solution vector x of a homogeneous equation is orthogonal to the coefficient vector a. To take this geometric observation a step further, consider the homogeneous system $a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0$

$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0$

$\vdots$

$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0$

If we denote the successive row vectors of the coefficient matrix by $r_1, r_2, \ldots, r_m$, then we can rewrite this system in dot product form as

$r_1.x = 0$

$r_2.x = 0$

$\vdots$

$r_m . x = 0$

from which we see that every solution vector x is orthogonal to every row vector of the coefficient matrix. In summary, we have the following result.

If A is an m × n matrix, then the solution set of the homogeneous linear system Ax = 0 consists of all vectors in R^n that are orthogonal to every row vector of A.

From the above statement it follows that if A is an m × n matrix, then the null space of A consists of those vectors that are orthogonal to each of the row vectors of A. To develop that idea in more detail, we make the following definition.

If W is a subspace of R^n, then the set of all vectors in R^n that are orthogonal to every vector in W is called the **orthogonal complement** of W and is denoted by the symbol $W^{\perp}$.

Theorem 3.15 If W is a subspace of R^n, then:

(a) $W^{\perp}$ is a subspace of R^n.

(b) The only vector common to W and $W^{\perp}$ is 0.

(c) The orthogonal complement of $W^{\perp}$ is W.

Theorem 3.16 If A is an m × n matrix, then:

(a) The null space of A and the row space of A are orthogonal complements in R^n.

(b) The null space of A^T and the column space of A are orthogonal complements in R^m.

Theorem 3.17 - Equivalent statements

If A is an n $\times$ n matrix, then the following statements are equivalent.

(a) A is invertible.

(b) Ax = 0 has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) Ax = b is consistent for every n $\times$ 1 matrix b.

(f) Ax = b has exactly one solution for every n $\times$ 1 matrix b.

(g) det(A) $\neq$ 0.

(h) The column vectors of A are linearly independent.

(i) The row vectors of A are linearly independent.

(j) The column vectors of A span R^n.

(k) The row vectors of A span R^n.

(l) The column vectors of A form a basis for R^n.

(m) The row vectors of A form a basis for R^n.

(n) A has rank n.

(o) A has nullity 0.

(p) The orthogonal complement of the null space of A is R^n.

(q) The orthogonal complement of the row space of A is 0.

The linear systems that have more constraints than unknowns, called **overdetermined systems**, or with fewer constraints than unknowns, called **underdetermined systems**.

Theorem 3.18 Let A be an m $\times$ n matrix.

(a) (Overdetermined Case). If m $>$ n, then the linear system Ax $=$ b is inconsistent for at least one vector b in R^n.

(b) (Underdetermined Case). If m $<$ n, then for each vector b in R^m the linear system Ax $=$ b is either inconsistent or has infinitely many solutions.

Proof. (a) Assume that m>n, in which case the column vectors of A cannot span R^m (fewer vectors than the dimension of R^m). Thus, there is at least one vector b in R^m that is not in the column space of A, and for any such b the system Ax $=$ b is inconsistent by Theorem 3.3.

(b) Assume that m<n. For each vector b in R^n there are two

possibilities: either the system Ax = b is consistent or it is inconsistent. If it is inconsistent, then the proof is complete. If it is consistent, then Theorem 3.13 implies that the general solution has n - r parameters, where r = rank(A). Since A is an $m \times n$ matrix, r=rank(A) $\leq$ min(m,n) and so
n - r $\geq$ n - m $>$ 0. This means that the general solution has at least one parameter and hence there are infinitely many solutions.

Example 12.

(a) What can you say about the solutions of an overdetermined system Ax = b of 7 equations in 5 unknowns in which A has rank r = 4?

(b) What can you say about the solutions of an underdetermined system Ax = b of 5 equations in 7 unknowns in which A has rank r = 4?

Solution. (a) The system is consistent for some vector b in R^7, and for any such b the number of parameters in the general solution is n - r = 5 - 4 = 1.

(b) The system may be consistent or inconsistent, but if it is consistent for the vector b in R^5, then the general solution has n - r = 7 - 4 = 3 parameters.

Example 13. What conditions must be satisfied by b_1, b_2, b_3, b_4, b_5 for the following linear system to be consistent?

$x_1 - 2x_2 = b_1$

$x_1 - x_2 = b_2$

$x_1 + x_2 = b_3$

$x_1 + 2x_2 = b_4$

$x_1 + 3x_2 = b_5$

Solution. In matrix notation, the given system is Ax=b, where

$$A = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix}.$$

Since the number of equations is greater than the number of unknowns, the given system is over determined. Hence the given system cannot be consistent for all values of b. The augmented matrix of the given system is $\begin{bmatrix} 1 & -2 & b_1 \\ 1 & -1 & b_2 \\ 1 & 1 & b_3 \\ 1 & 2 & b_4 \\ 1 & 3 & b_5 \end{bmatrix}$.

Its row equivalent form is $\begin{bmatrix} 1 & 0 & 2b_2 - b_1 \\ 0 & 1 & b_2 - b_1 \\ 0 & 0 & b_3 - 3b_2 + 2b_1 \\ 0 & 0 & b_4 - 4b_2 + 3b_1 \\ 0 & 0 & b_5 - 5b_2 + 4b_1 \end{bmatrix}$.

Thus, the system is consistent if and only if b_1, b_2, b_3, b_4, b_5 satisfy the conditions

$2b_1 - 3b_2 + b_3 = 0$

$3b_1 - 4b_2 + b_4 = 0$

$4b_1 - 5b_2 + b_5 = 0$

This is a homogeneous system of linear equations in 5 unknowns. Its coefficient matrix is

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

which is equivalent to

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

Thus, $$\begin{bmatrix} 1 & 0 & 0 & -5 & 4 \\ 0 & 1 & 0 & -4 & 3 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

which is equivalent to the system of equations

$$b_1 = 5r - 4s, \quad b_2 = 4r - 3s, \quad b_3 = 2r - s, \quad b_4 = r, \quad b_5 = s$$

where r and s are arbitrary.

3.4 Basic Matrix Transformations in R^2 and R^3

There are many ways to transform the vector spaces R^2 and R^3, some of the most important of which can be accomplished by matrix transformations. For example, rotations about the origin, reflections about lines and planes through the origin, and projections onto lines and planes through the origin can all be accomplished using a linear operator T_A in which A is an appropriate 2×2 or 3×3 matrix.

Some of the most basic matrix operators on R^2 and R^3 are those that map each point into its symmetric image about a fixed line or a fixed plane that contains the origin; these are called **reflection operators**. Table 1 shows the standard matrices for the reflections about the coordinate axes in R^2, and Table 2 shows the standard matrices for the reflections about the coordinate planes in R^3.

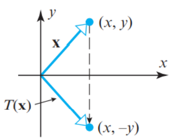
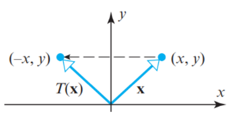
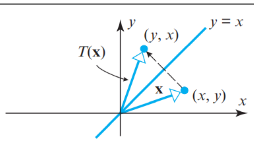
Operator	Illustration	Images of $\mathbf{e}_1$ and $\mathbf{e}_2$	Standard Matrix
Reflection about the x-axis $T(x, y) = (x, -y)$		$T(\mathbf{e}_1) = T(1, 0) = (1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, -1)$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Reflection about the y-axis $T(x, y) = (-x, y)$		$T(\mathbf{e}_1) = T(1, 0) = (-1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, 1)$	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
Reflection about the line $y = x$ $T(x, y) = (y, x)$		$T(\mathbf{e}_1) = T(1, 0) = (0, 1)$ $T(\mathbf{e}_2) = T(0, 1) = (1, 0)$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Figure 3.1: Table 1

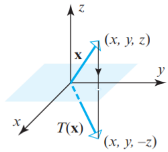
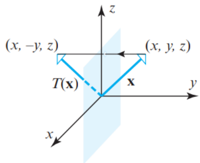
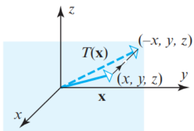
Operator	Illustration	Images of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$	Standard Matrix
Reflection about the xy-plane $T(x, y, z) = (x, y, -z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, -1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$
Reflection about the xz-plane $T(x, y, z) = (x, -y, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, -1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Reflection about the yz-plane $T(x, y, z) = (-x, y, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (-1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Figure 3.2: Table 2

Matrix operators on R^2 and R^3 that map each point into its orthogonal projection onto a fixed line or plane through the origin are called **projection operators** (or more precisely, **orthogonal projection operators**). Table 3 shows the standard matrices for the orthogonal projections onto the coordinate axes in R^2, and Table 4 shows the standard matrices for the orthogonal projections onto the coordinate planes in R^3.

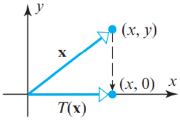
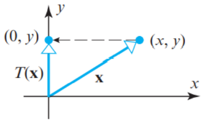
Operator	Illustration	Images of e_1 and e_2	Standard Matrix
Orthogonal projection onto the x-axis $T(x, y) = (x, 0)$		$T(\mathbf{e}_1) = T(1, 0) = (1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, 0)$	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Orthogonal projection onto the y-axis $T(x, y) = (0, y)$		$T(\mathbf{e}_1) = T(1, 0) = (0, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, 1)$	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

Figure 3.3: Table 3

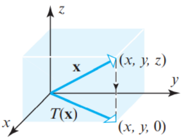
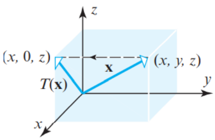
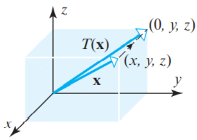
Operator	Illustration	Images of e_1, e_2, e_3	Standard Matrix
Orthogonal projection onto the xy-plane $T(x, y, z) = (x, y, 0)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 0)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Orthogonal projection onto the xz-plane $T(x, y, z) = (x, 0, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 0, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Orthogonal projection onto the yz-plane $T(x, y, z) = (0, y, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (0, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Figure 3.4: Table 4

Matrix operators on R^2 and R^3 that move points along arcs of circles centered at the origin are called **rotation operators**. Let us consider how to find the standard matrix for the rotation operator $T : R^2 \to R^2$ that moves points counterclockwise about the origin through a positive angle θ. As illustrated in Figure 3.5, the images of the standard basis vectors are

$T(e_1) = T(1, 0) = (cos\theta, sin\theta)$ and

$T(e_2) = T(0, 1) = (-sin\theta, cos\theta)$

so it follows that the standard matrix for T is

$$A = [T(e_1)|T(e_2)] = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

In keeping with common usage we will denote this operator by R_θ and call

$$\begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \qquad (1)$$

the **rotation matrix** for R^2. If $x = (x, y)$ is a vector in R^2, and if $w = (w_1, w_2)$ is its image under the rotation, then the relationship $w = R_\theta x$ can be written in component form as

$$w_1 = xcos\theta - ysin\theta \qquad w_2 = xsin\theta + ycos\theta \qquad (2)$$

These are called the **rotation equations** for R^2. These ideas are summarized in Table 5.

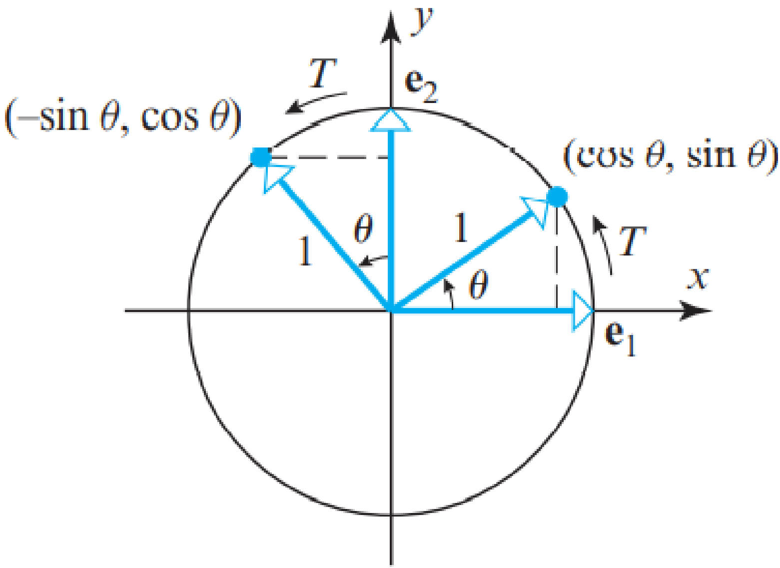


Figure 3.5

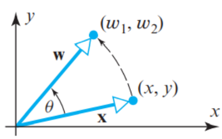
Operator	Illustration	Rotation Equations	Standard Matrix
Counterclockwise rotation about the origin through an angle θ		$w_1 = x\cos\theta - y\sin\theta$ $w_2 = x\sin\theta + y\cos\theta$	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Figure 3.6: Table 5

Example 14. Find the image of $x = (1, 1)$ under a rotation of $\frac{\pi}{6}$ radians ($= 30$ °) about the origin.

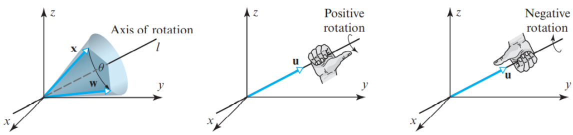
Solution. It follows from (1) with $\theta = \frac{\pi}{6}$ that

$$R_{\frac{\pi}{6}} x = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}-1}{2} \\ \frac{1+\sqrt{3}}{2} \end{bmatrix} \approx \begin{bmatrix} 0.37 \\ 1.37 \end{bmatrix}$$

or in comma-delimited notation, $R_{\frac{\pi}{6}}(1, 1) \approx (0.37, 1.37)$.

A rotation of vectors in R^3 is commonly described in relation to a line through the origin called the **axis of rotation** and a unit vector $\mathbf{u}$ along that line (Figure 3.7a). The unit vector and what is called the **right-hand rule** can be used to establish a sign for the angle of rotation by cupping the fingers of your right hand so they curl in the direction of rotation and observing the direction of your thumb. If your thumb points in the direction of $\mathbf{u}$, then the angle of rotation is regarded to be **positive** relative to $\mathbf{u}$, and if it points in the direction opposite to $\mathbf{u}$, then it is regarded to be **negative** relative to $\mathbf{u}$ (Figure 3.7b).

For rotations about the coordinate axes in R^3, we will take the unit vectors to be **i, j, k**, in which case an angle of rotation will be positive if it is counterclockwise looking toward the origin along the positive coordinate axis and will be negative if it is clockwise. Table 6 shows the standard matrices for the **rotation operators** on R^3 that rotate each vector about one of the coordinate axes through an angle θ.



(*a*) Angle of rotation (*b*) Right-hand rule

Figure 3.7

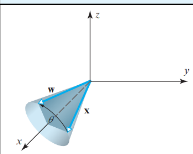
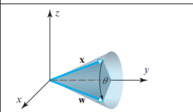
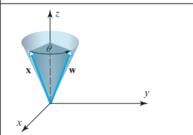
Operator	Illustration	Rotation Equations	Standard Matrix
Counterclockwise rotation about the positive x-axis through an angle θ		$w_1 = x$ $w_2 = y\cos\theta - z\sin\theta$ $w_3 = y\sin\theta + z\cos\theta$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$
Counterclockwise rotation about the positive y-axis through an angle θ		$w_1 = x\cos\theta + z\sin\theta$ $w_2 = y$ $w_3 = -x\sin\theta + z\cos\theta$	$\begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$
Counterclockwise rotation about the positive z-axis through an angle θ		$w_1 = x\cos\theta - y\sin\theta$ $w_2 = x\sin\theta + y\cos\theta$ $w_3 = z$	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Figure 3.8: Table 6

If k is a nonnegative scalar, then the operator $T(x) = kx$ on R^2 or R^3 has the effect of increasing or decreasing the length of each vector by a factor of k. If $0 \leq k < 1$ the operator is called a **contraction** with factor k, and if $k > 1$ it is called a **dilation** with factor k (Figure 3.9). Tables 7 and 8 illustrate these operators. If $k = 1$, then T is the **identity operator**.

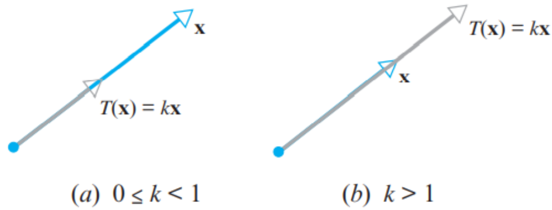


Figure 3.9

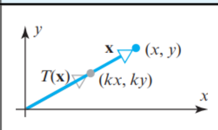
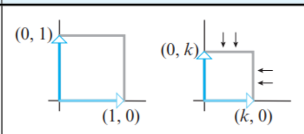
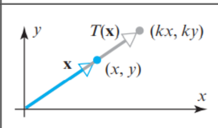
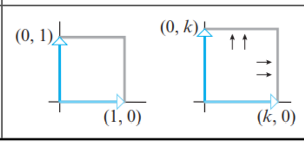
Operator	Illustration $T(x, y) = (kx, ky)$	Effect on the Unit Square	Standard Matrix
Contraction with factor k in R^2 $(0 \leq k < 1)$			$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$
Dilation with factor k in R^2 $(k > 1)$			

Figure 3.10: Table 7

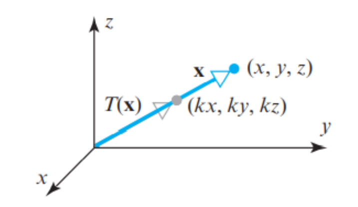
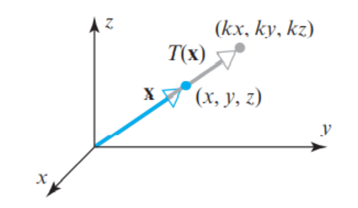
Operator	Illustration $T(x, y, z) = (kx, ky, kz)$	Standard Matrix
Contraction with factor k in R^3 $(0 \leq k < 1)$		$\begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k \end{bmatrix}$
Dilation with factor k in R^3 $(k > 1)$		

Figure 3.11: Table 8

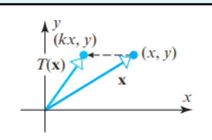
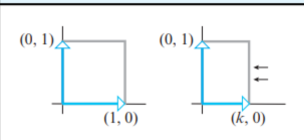
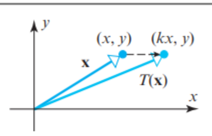
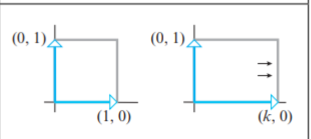
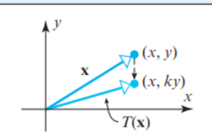
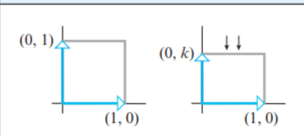
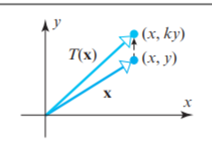
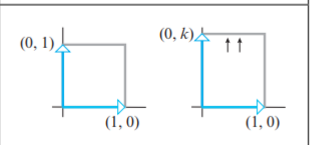
Operator	Illustration $T(x, y) = (kx, y)$	Effect on the Unit Square	Standard Matrix
Compression in the x-direction with factor k in R^2 $(0 \leq k < 1)$			$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Expansion in the x-direction with factor k in R^2 $(k > 1)$			
Operator	**Illustration $T(x, y) = (x, ky)$**	**Effect on the Unit Square**	**Standard Matrix**
Compression in the y-direction with factor k in R^2 $(0 \leq k < 1)$			$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$
Expansion in the y-direction with factor k in R^2 $(k > 1)$			

Figure 3.12: Table 9

In a dilation or contraction of R^2 or R^3, all coordinates are multiplied by a nonnegative factor k. If only one coordinate is multiplied by k, then, depending on the value of k, the resulting operator is called a **compression or expansion** with factor k in the direction of a coordinate axis. This is illustrated above in Table 9 for R^2. The extension to R^3 is left as an exercise.

A matrix operator of the form $T(x, y) = (x + ky, y)$ translates a point (x, y) in the xy-plane parallel to the x-axis by an amount ky that is proportional to the y-coordinate of the point. This operator leaves the points on the x-axis fixed (since y = 0), but as we progress away from the x-axis, the translation distance increases. We call this operator the **shear in the x-direction by a factor k**. Similarly, a matrix operator of the form $T(x, y) = (x, y + kx)$ is called the **shear in the y-direction by a factor k**. Table 10, which illustrates the basic information about shears in R^2, shows that a shear is in the positive direction if $k > 0$ and the negative direction if $k < 0$.

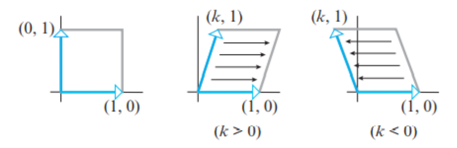
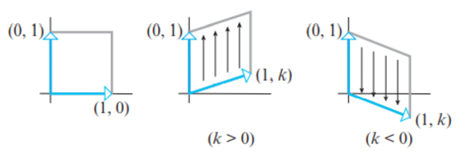
Operator	Effect on the Unit Square	Standard Matrix
Shear in the x-direction by a factor k in R^2 $T(x, y) = (x + ky, y)$		$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Shear in the y-direction by a factor k in R^2 $T(x, y) = (x, y + kx)$		$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

Figure 3.13: Table 10

Example 15. In each part, describe the matrix operator whose standard matrix is shown, and show its effect on the unit square.

(a) $A_1 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ (b) $A_2 = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$

(c) $A_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ (d) $A_4 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$

Solution. By comparing the forms of these matrices to those in Tables 7, 9, and 10, we see that the matrix A_1 corresponds to a shear in the x-direction by a factor 2, the matrix A_2 corresponds to a shear in the x-direction by a factor -2, the matrix A_3 corresponds to a dilation with factor 2, and the matrix A_4 corresponds to an expansion in the x direction with factor 2.

The effects of these operators on the unit square are shown in Figure 3.14.

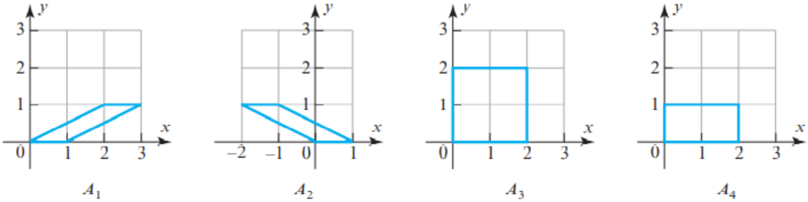


Figure 3.14

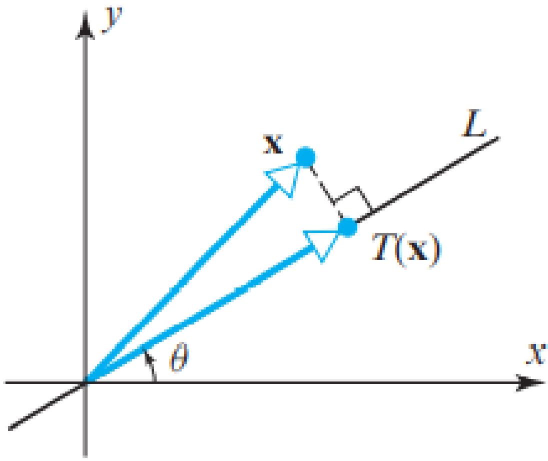


Figure 3.15

In Table 3 we listed the standard matrices for the orthogonal projections onto the coordinate axes in R^2. These are special cases of the more general matrix operator $T_A : R^2 \rightarrow R^2$ that maps each point into its **orthogonal projection onto a line L through the origin** that makes an angle θ with the positive x-axis (Figure 3.15).

Let a be the unit vector along the line L. Then since the line makes an angle θ with the positive x-axis, $a = (cos\theta, sin\theta)$.

Then $\quad \|a\| = \sqrt{cos^2\theta + sin^2\theta} = 1,$

$e_1.a = (1, 0).(cos\theta, sin\theta) = cos\theta,$

$e_2.a = (0, 1).(cos\theta, sin\theta) = sin\theta.$

Hence $T_A(e_1) =$ projection of e_1 along L $= \dfrac{e_1.a}{(\|a\|)^2}a$

$= cos\theta(cos\theta, sin\theta) = (cos^2\theta, sin\theta cos\theta)$

and $T_A(e_2) =$ projection of e_2 along L $= \dfrac{e_2.a}{(\|a\|)^2}a$

$= sin\theta(cos\theta, sin\theta) = (sin\theta cos\theta, sin^2\theta).$

Hence the standard matrix A for the transformation is

$$A = [T_A(e_1)|T_A(e_2)] = \begin{bmatrix} cos^2\theta & sin\theta cos\theta \\ sin\theta cos\theta & sin^2\theta \end{bmatrix} = \begin{bmatrix} cos^2\theta & \frac{1}{2}sin2\theta \\ \frac{1}{2}sin2\theta & sin^2\theta \end{bmatrix}.$$

In keeping with common usage, we will denote this operator by

$$P_\theta = \begin{bmatrix} cos^2\theta & sin\theta cos\theta \\ sin\theta cos\theta & sin^2\theta \end{bmatrix} = \begin{bmatrix} cos^2\theta & \frac{1}{2}sin2\theta \\ \frac{1}{2}sin2\theta & sin^2\theta \end{bmatrix}. \qquad (3)$$

Example 16. Use Formula (3) to find the orthogonal projection of the vector $x = (1, 5)$ onto the line through the origin that makes an angle of $\frac{\pi}{6}$ radians ($= 30\ °$) with the positive x-axis.

Solution. Since $sin(\frac{\pi}{6}) = 1/2$ and $cos(\frac{\pi}{6}) = \sqrt{3}/2$, it follows from (3) that the standard matrix for this projection is

$$P_{\frac{\pi}{6}} = \begin{bmatrix} cos^2\frac{\pi}{6} & sin\frac{\pi}{6}cos\frac{\pi}{6} \\ sin\frac{\pi}{6}cos\frac{\pi}{6} & sin^2\frac{\pi}{6} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix}.$$

Thus, $P_{\frac{\pi}{6}}x = \begin{bmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{3+5\sqrt{3}}{4} \\ \frac{\sqrt{3}+5}{4} \end{bmatrix} = \begin{bmatrix} 2.91 \\ 1.68 \end{bmatrix}.$

In Table 1 we listed the reflections about the coordinate axes in R^2. These are special cases of the more general operator $H_\theta : R^2 \to R^2$ that maps each point into its reflection about a line L through the origin that makes an angle θ with the positive x-axis (Figure 3.16). We could find the standard matrix for H_θ by finding the images of the standard basis vectors, but instead we will take advantage of our work on orthogonal projections by using Formula (3) for P_θ to find a formula for H_θ.

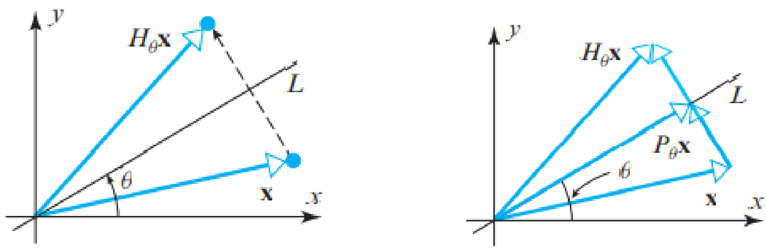
You should be able to see from Figure 3.17 that for every vector x in R^n

$P_\theta x - x = \dfrac{1}{2}(H_\theta x - x)$ or equivalently $H_\theta x = (2P_\theta - I)x.$

Thus, it follows that $H_\theta = 2P_\theta - I$ (4)

and hence from (3) that

$$H_\theta = \begin{bmatrix} cos2\theta & sin2\theta \\ sin2\theta & -cos2\theta \end{bmatrix}.$$



(a) (b)

Figure 3.16

Example 17. Find the reflection of the vector $x = (1, 5)$ about the line through the origin that makes an angle of $\frac{\pi}{6}$ radians ($= 30\ ^\circ$) with the x-axis.

Solution. $H_{\frac{\pi}{6}} = \begin{bmatrix} cos2\frac{\pi}{6} & sin2\frac{\pi}{6} \\ sin2\frac{\pi}{6} & -cos2\frac{\pi}{6} \end{bmatrix} = \begin{bmatrix} cos\frac{\pi}{3} & sin\frac{\pi}{3} \\ sin\frac{\pi}{3} & -cos\frac{\pi}{3} \end{bmatrix}$

$$= \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix}.$$

Hence $H_{\frac{\pi}{6}} x = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{(1+5\sqrt{3})}{2} \\ \frac{(\sqrt{3}-5)}{2} \end{bmatrix} \approx \begin{bmatrix} 4.83 \\ -1.63 \end{bmatrix}.$

3.5 Properties of Matrix Transformations

Suppose that T_A is a matrix transformation from R^n to R^k and T_B is a matrix transformation from R^k to R^m. If x is a vector in R^n, then T_A maps this vector into a vector $T_A(x)$ in R^k, and T_B, in turn, maps that vector into the vector $T_B(T_A(x))$ in R^m. This process creates a transformation from R^n to R^m that we call the **composition of** T_B **with** T_A and denote by the symbol $T_B \circ T_A$ which is read "T_B circle T_A."

As illustrated in Figure 3.18, the transformation T_A in the formula is performed first; that is,

$(T_B \circ T_A)(x) = T_B(T_A(x)) \qquad (1)$

This composition is itself a matrix transformation since

$(T_B \circ T_A)(x) = T_B(T_A(x)) = B(T_A(x)) = B(Ax) = (BA)x$ which shows that it is multiplication by BA. This is expressed by the formula $T_B \circ T_A = T_{BA} \qquad (2)$

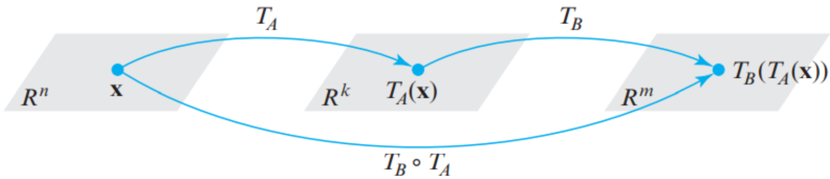


Figure 3.17

Compositions can be defined for any finite succession of matrix transformations whose domains and ranges have the appropriate dimensions. For example, to extend Formula (2) to three factors, consider the matrix transformations

$T_A : R^n \to R^k,\ T_B : R^k \to R^l,\ T_C : R^l \to R^m$

We define the composition $(T_C \circ T_B \circ T_A) : R^n \to Rm$ by $(T_C \circ T_B \circ T_A)(x) = T_C(T_B(T_A(x)))$.

As above, it can be shown that this is a matrix transformation whose standard matrix is CBA and that

$T_C \circ T_B \circ T_A = T_{CBA}$ (3).

Sometimes we will want to refer to the standard matrix for a matrix transformation $T : R^n \to R^m$ without giving a name to the matrix itself. In such cases we will denote the standard matrix for T by the symbol $[T]$.

Thus, the equation $T(x) = [T]x$ states that $T(x)$ is the product of the standard matrix $[T]$ and the column vector x. For example, if $T_1 : R^n \to R^k$ and if $T_2 : R^k \to R^m$, then Formula (2) can be restated as $[T_2 \circ T_1] = [T_2][T_1]$. (4)

Similarly, Formula (3) can be restated as $[T_3 \circ T_2 \circ T_1] = [T_3][T_2][T_1]$. (5)

Example 18. An example for a transformation where composition is not commutative.

Let $T_1 : R^2 \to R^2$ be the reflection about the line y = x, and let $T_2 : R^2 \to R^2$ be the orthogonal projection onto the y-axis. Figure 3.18 illustrates graphically that $T_1 \circ T_2$ and $T_2 \circ T_1$ have different effects on a vector x. This same conclusion can be reached by showing that the standard matrices for T_1 and T_2 do not commute:

$$[T_1 \circ T_2] = [T_1]\,[T_2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$[T_2 \circ T_1] = [T_2]\,[T_1] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

and so $[T_1 \circ T_2] \neq [T_2 \circ T_1]$.

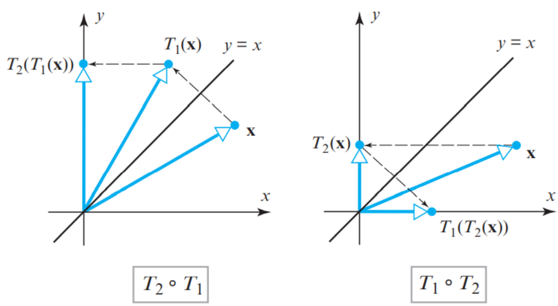


Figure 3.18

Example 19. An example for a transformation where composition is commutative.

Let $T_1 : R^2 \to R^2$ and $T_2 : R^2 \to R^2$ be the matrix operators that rotate vectors about the origin through the angles θ_1 and θ_2, respectively. Thus the operation $(T_2 \circ T_1)(x) = T_2(T_1(x))$ first rotates x through the angle θ_1, then rotates $T_1(x)$ through the angle θ_2. It follows that the net effect of $T_2 \circ T_1$ is to rotate each vector in R^2 through the angle $\theta_1 + \theta_2$ (Figure 3.19).

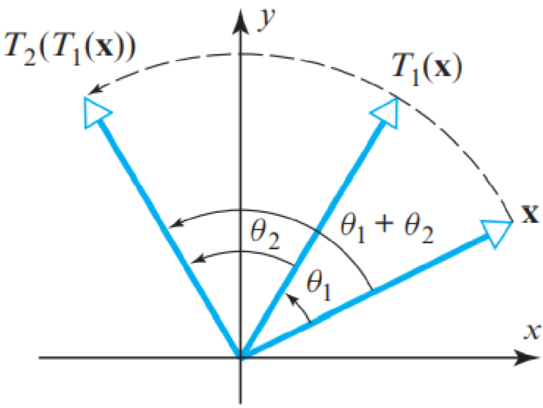


Figure 3.19

The standard matrices for these matrix operators, which are

$$[T_1] = \begin{bmatrix} cos\theta_1 & -sin\theta_1 \\ sin\theta_1 & cos\theta_1 \end{bmatrix}, \ [T_2] = \begin{bmatrix} cos\theta_2 & -sin\theta_2 \\ sin\theta_2 & cos\theta_2 \end{bmatrix},$$

$$[T_2 \circ T_1] = \begin{bmatrix} cos(\theta_1 + \theta_2) & -sin(\theta_1 + \theta_2) \\ sin(\theta_1 + \theta_2) & cos(\theta_1 + \theta_2) \end{bmatrix}$$

should satisfy (4). With the help of some basic trigonometric identities, we can confirm that this is so as follows:

$$[T_2]\,[T_1] = \begin{bmatrix} cos\theta_2 & -sin\theta_2 \\ sin\theta_2 & cos\theta_2 \end{bmatrix} \begin{bmatrix} cos\theta_1 & -sin\theta_1 \\ sin\theta_1 & cos\theta_1 \end{bmatrix}$$

$$= \begin{bmatrix} cos\theta_2 cos\theta_1 - sin\theta_2 sin\theta_1 & -(cos\theta_2 sin\theta_1 + sin\theta_2 cos\theta_1) \\ sin\theta_2 cos\theta_1 + cos\theta_2 sin\theta_1 & -sin\theta_2 sin\theta_1 + cos\theta_2 cos\theta_1 \end{bmatrix}$$

$$= \begin{bmatrix} cos(\theta_1 + \theta_2) & -sin(\theta_1 + \theta_2) \\ sin(\theta_1 + \theta_2) & cos(\theta_1 + \theta_2) \end{bmatrix}$$

$$= [T_2 \circ T_1].$$

One-to-One Matrix Transformations

A matrix transformation $T_A : R^n \rightarrow R^m$ is said to be **one-to-one** if T_A maps distinct vectors (points) in R^n into distinct vectors (points) in R^m.

(See Figure 3.20) This idea can be expressed in various ways. For example, you should be able to see that the following are just restatements of the above definition.

1. T_A is one-to-one if for each vector b in the range of A there

is exactly one vector x in R^n such that $T_A x = b$.

2. T_A is one-to-one if the equality $T_A(u) = T_A(v)$ implies that $u = v$.

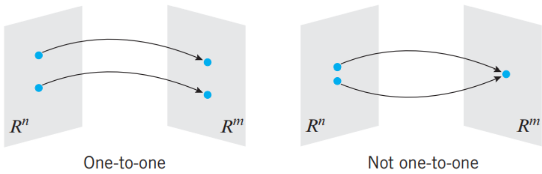


Figure 3.20

If $T_A : R^n \to R^m$ is a matrix transformation, then the set of all vectors in R^n that T_A maps into 0 is called the **kernel** of T_A and is denoted by $ker(T_A)$. The set of all vectors in R^m that are images under this transformation of at least one vector in R^n is called the **range** of T_A and is denoted by $R(T_A)$.

In brief:

$ker(T_A) = null \quad space \quad of \quad A$ (6)

$R(T_A) = column \quad space \quad of \quad A$ (7)

For a linear operator $T_A : R^n \to R^n$, the following theorem establishes fundamental relationships between the invertibility of A and properties of T_A.

Theorem 3.19 If A is an n $\times$ n matrix and $T_A : R^n \to R^n$ is the corresponding matrix operator, then the following statements are equivalent.

(a) A is invertible.

(b) The kernel of T_A is $\{0\}$.

(c) The range of T_A is R^n.

(d) T_A is one-to-one.

Proof. We can prove this theorem by establishing the chain of implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (a).

(a) $\Rightarrow$ (b)

Assume that A is invertible. We know

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

Since A is invertible, the system $Ax = 0$ has only the trivial solution and so the kernel of T_A is $\{0\}$.

(b) $\Rightarrow$ (c)

Assume that the kernel of T_A is $\{0\}$. It follows from Formula (6) that the null space of A is $\{0\}$ and hence that A has nullity 0. This in turn implies that the rank of A is n and hence that the column space of A is all of R^n. Formula (7) now implies that the range of T_A is R^n.

(c) $\Rightarrow$ (d)

Assume that the range of T_A is R^n. It follows from Formula (6) that the column space of A is R^n and hence is of dimension n. Thus rank of A is n and so by dimenson theorem for matrices, we get that A has nullity 0. Hence null space of A is $\{0\}$.

For any two vectors u and v in R^n, we have

$$\begin{aligned} T_A(u) = T_A(v) &\Rightarrow Au = Av \Rightarrow A(u-v) = Au - Av = 0 \\ &\Rightarrow u - v \in nullspace \;\; of \;\; A \Rightarrow u - v = 0 \\ &\Rightarrow u = v. \end{aligned}$$

Hence T_A is one-to-one.

(d) $\Rightarrow$ (a)

Assume T_A is one-to-one. Then $T_A(0) = A0 = 0$.

Since T_A is one-to-one, 0 is the only vector in R^n such that $T_A(0) = A0 = 0$. Hence the system $Ax = 0$ has only trivial solution and so A is invertible.

Remark. If A is an m $\times$ n matrix, here are three ways of viewing the same subspace of R^n:

- Matrix view: the null space of A
- System view: the solution space of Ax $=$ 0
- Transformation view: the kernel of T_A

and here are three ways of viewing the same subspace of R^m:

- Matrix view: the column space of A
- System view: all b in R^m for which Ax = b is consistent
- Transformation view: the range of T_A

Example 20. The rotation operator on R^2 is one-to-one.

Let T denote operator R^2 that rotates vectors counterclockwise about the origin through an angle θ. Then the standard matrix for T is

$$[T] = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}.$$

Then $det\,[T] = cos^2\theta + sin^2\theta = 1 \neq 0$.

Since $det\,[T] \neq 0$, T is one-to-one.

nverse of a One-to-One Matrix Operator If $T_A : R^n \to R^n$ is a one-to-one matrix operator, then it follows from Theorem 3.19 that A is invertible. The matrix operator $T_{A^{-1}} : R^n \to R^n$ that corresponds to A^{-1} is called the **inverse operator** or (more simply) the **inverse of** T_A. This terminology is appropriate because T_A and $T_{A^{-1}}$ cancel the effect of each other in the sense that if x is any vector in R^n, then

$T_A(T_{A^{-1}}(x)) = AA^{-1}x = Ix = x$

$T_{A^{-1}}(TA(x)) = A^{-1}Ax = Ix = x$

or, equivalently,

$T_A \circ T_{A^{-1}} = T_{AA^{-1}} = T_I$

$T_{A^{-1}} \circ T_A = T_{A^{-1}A} = T_I$

From a more geometric viewpoint, if w is the image of x under T_A, then $T_{A^{-1}}$ maps w backinto x, since $T_{A^{-1}}(w) = T_{A^{-1}}(T_A(x)) = x$.

This is illustrated in Figure 3.21 for R^2. **Example 21.**

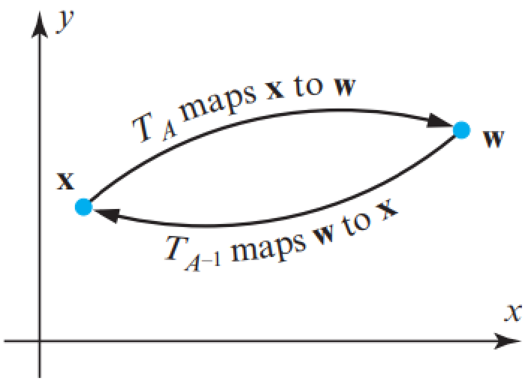


Figure 3.21

Standard Matrix for T^{-1}

Let $T : R^2 \to R^2$ be the operator that rotates each vector in R^2 through the angle θ. The standard matrix for T is

$$[T] = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}.$$

Then $det\,[T] = cos^2\theta + sin^2\theta = 1 \neq 0$.

Since $det\,[T] \neq 0$, T is invertible.

By interchanging the diagonal elements and changing the signs of the off-diagonal elements, we get the adjoint of $[T]$ as

$$adj\,[T] = \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix}.$$

$$\therefore [T^{-1}] = [T]^{-1} = \frac{1}{det\,[T]} \times adj\,[T]$$

$$= 1 \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} = \begin{bmatrix} cos(-\theta) & -sin(-\theta) \\ sin(-\theta) & cos(-\theta) \end{bmatrix},$$

which is the standard matrix for the rotation through the angle $-\theta$.

Example 22. Show that the operator $T : R^2 \to R^2$ defined by the equations $w_1 = 2x_1 + x_2, \;\; w_2 = 3x_1 + 4x_2$ is one-to-one, and find $T^{-1}(w_1, w_2)$.

Solution. The matrix form of these equations is

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

so the standard matrix for T is

$$[T] = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}.$$

This matrix is invertible (so T is one-to-one) and the standard matrix for T^{-1} is

$$[T^{-1}] = [T]^{-1} = \begin{bmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix}.$$

Thus $[T^{-1}] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \frac{4}{5}w_1 - \frac{1}{5}w_2 \\ -\frac{3}{5}w_1 + \frac{2}{5}w_2 \end{bmatrix}$

from which we conclude that

$T^{-1}(w_1, w_2) = (\frac{4}{5}w_1 - \frac{1}{5}w_2, -\frac{3}{5}w_1 + \frac{2}{5}w_2).$

Theorem 3.21 - Equivalent statements

If A is an n $\times$ n matrix, then the following statements are equivalent.

(a) A is invertible.

(b) Ax $=$ 0 has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) Ax $=$ b is consistent for every n $\times$ 1 matrix b.

(f) Ax $=$ b has exactly one solution for every n $\times$ 1 matrix b.

(g) det(A) $\neq$ 0.

(h) The column vectors of A are linearly independent.

(i) The row vectors of A are linearly independent.

(j) The column vectors of A span R^n.

(k) The row vectors of A span R^n.

(l) The column vectors of A form a basis for R^n.

(m) The row vectors of A form a basis for R^n.

(n) A has rank n.

(o) A has nullity 0.

(p) The orthogonal complement of the null space of A is R^n.

(q) The orthogonal complement of the row space of A is 0.

(r) The kernel of T_A is $\{0\}$.

(s) The range of T_A is R^n.

(t) T_A is one-to-one.

Problems

1. Let $T_1 : R^2 \to R^2$ and $T_2 : R^2 \to R^2$ be $T_1(x, y) = (-x, y)$ and $T_2(x, y) = (x, -y)$ respectively. Show that these transformations commute.

2. Let $T_1(x, y) = (x + y, x - y)$ and $T_2(x, y) = (3x, 2x + 4y)$.

(a) Find the standard matrices for T_1 and T_2.

(b) Find the standard matrices for $T_1 \circ T_2$ and $T_2 \circ T_1$.

(c) Use the matrices in part (b) to find the formulas for $T_1(T_2(x, y))$ and $T_2(T_1(x, y))$.

MODULE

FOUR

EIGENVALUES, EIGENVECTORS, INNER PRODUCT SPACES, DIAGONALIZATION

4.1 Geometry of Matrix Operators on R^2

The effect of a matrix operator on R2 can often be deduced by studying how it transforms the points that form the unit square. The following theorem, which we state without proof, shows that if the operator is invertible, then it maps each line segment in the unit square into the line segment connecting the images of its endpoints. In particular, the edges of the unit square get mapped into edges of the image (see Figure 4.1 in which the edges of a unit square and the corresponding edges of its image have been numbered).

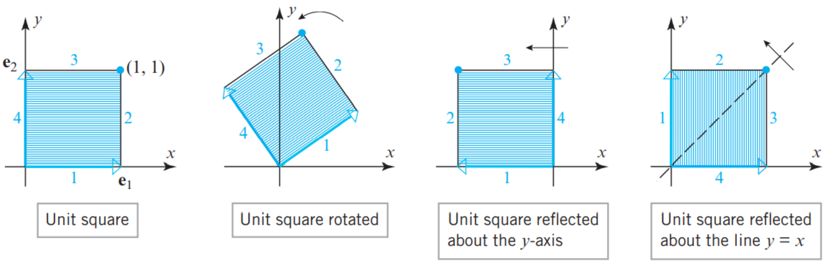


Figure 4.1

Theorem 4.1 If $T : R^2 \to R^2$ is multiplication by an invertible matrix, then:

(a) The image of a straight line is a straight line.

(b) The image of a line through the origin is a line through the origin.

(c) The images of parallel lines are parallel lines.

(d) The image of the line segment joining points P and Q is the line segment joining the images of P and Q.

(e) The images of three points lie on a line if and only if the points themselves lie on a line.

Example 1. According to Theorem 4.1, the invertible matrix A $= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$ maps the line $y = 2x + 1$ into another line. Find its equation.

Solution. Let (x, y) be a point on the line $y = 2x + 1$, and let $(x^{'}, y^{'})$ be its image under multiplication by A. Then

$$\begin{bmatrix} x^{'} \\ y^{'} \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \text{ and}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x^{'} \\ y^{'} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} x^{'} \\ y^{'} \end{bmatrix}$$

so $x = x^{'} - y^{'}$

$y = -2x^{'} + 3y^{'}$.

Substituting these expressions in $y = 2x + 1$ yields

$-2x^{'} + 3y^{'} = -2(x^{'} - y^{'}) + 1$

or, equivalently, $y^{'} = \frac{4}{5}x^{'} + \frac{1}{5}$.

Example 2. Sketch the image of the unit square under multiplication by the invertible matrix A $= \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$. Label the vertices of the image with their coordinates, and number the edges of the unit square and their corresponding images.

Since $\begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$,

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

the image of the unit square is a parallelogram with vertices (0, 0), (0, 2), (1, 1), and (1, 3).

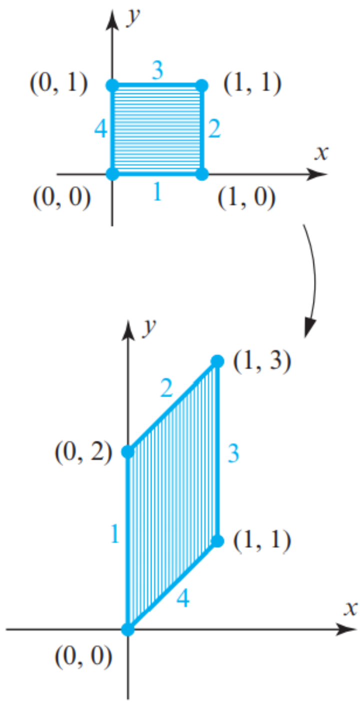


Figure 4.2

Example 3. Transformation of the Unit Square

(a) Find the standard matrix for the operator on R^2 that first shears by a factor of 2 in the x-direction and then reflects the result about the line $y = x$. Sketch the image of the unit square under this operator. (b) Find the standard matrix for the operator on R^2 that first reflects about $y = x$ and then shears by a factor of 2 in the x-direction. Sketch the image of the unit square under this operator. (c) Confirm that the shear and the reflection in parts (a) and (b) do not commute.

Solution. (a) The standard matrix for the shear is $A_1 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and for the reflection is $A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Thus, the standard matrix for the shear followed by the reflection is

$$A_2 A_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}.$$

(b) The standard matrix for the reflection followed by the shear is

$$A_1 A_2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}.$$

(c) The computations in solutions (a) and (b) show that $A_1 A_2 \neq A_2 A_1$, so the standard matrices, and hence the operators, do not commute. The same conclusion follows from Figures 4.3 and 4.4 since the two operators produce different images of the unit square.

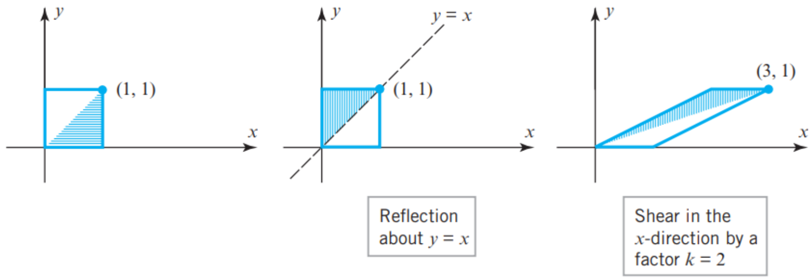


Figure 4.3

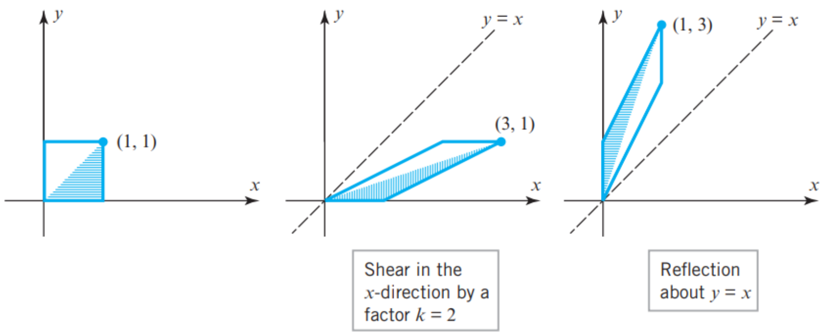
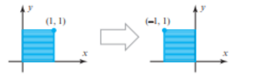
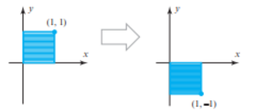
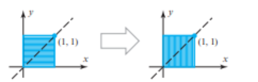
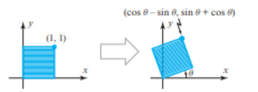
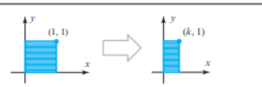
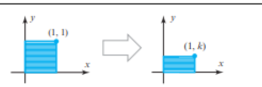
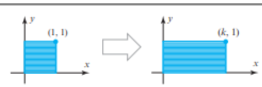


Figure 4.4

In Example 3 we illustrated the effect on the unit square in R^2 of a composition of shears and reflections. Our next objective is to show how to decompose any 2×2 invertible matrix into a product of matrices in Table 1, thereby allowing us to analyze the geometric effect of a matrix operator in R^2 as a composition of simpler matrix operators.

Operator	Standard Matrix	Effect on the Unit Square
Reflection about the y-axis	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	
Reflection about the x-axis	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	
Reflection about the line $y = x$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	
Rotation about the origin through a positive angle θ	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$	
Compression in the x-direction with factor k $(0 < k < 1)$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$	
Compression in the y-direction with factor k $(0 < k < 1)$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$	
Expansion in the x-direction with factor k $(k > 1)$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$	

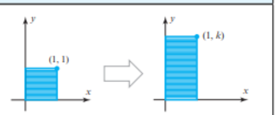
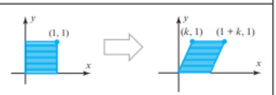
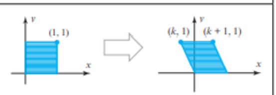
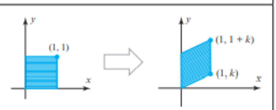
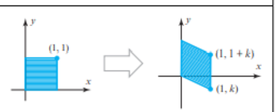
Operator	Standard Matrix	Effect on the Unit Square
Expansion in the y-direction with factor k $(k > 1)$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$	
Shear in the positive x-direction by a factor k $(k > 0)$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	
Shear in the negative x-direction by a factor k $(k < 0)$	$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$	
Shear in the positive y-direction by a factor k $(k > 0)$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$	
Shear in the negative y-direction by a factor k $(k < 0)$	$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$	

Figure 4.5: Table 1

Theorem 4.2 If E is an elementary matrtix, then $T_E : R^2 \to R^2$ is one of the following:

(a) A shear along a coordinate axis.

(b) A reflection about $y = x$.

(c) A compression along a coordinate axis.

(d) An expansion along a coordinate axis.

(e) A reflection about a coordinate axis.

(f) A compression or expansion along a coordinate axis followed by a reflection about a coordinate axis.

Proof. Because a 2×2 elementary matrix results from performing a single elementary row operation on the 2×2 identity matrix, such a matrix must have one of the following forms: $\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}, \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

The first two matrices represent shears along coordinate axes, and the third represents a reflection about $y = x$. If $k > 0$, the last two matrices represent compressions or expansions along coordinate axes, depending on whether $0 \le k < 1$ or $k > 1$. If $k < 0$, and if we express k in the form $k = -k_1$, where $k_1 > 0$, then the last two matrices can be written as

$$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -k_1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 & 0 \\ 0 & 1 \end{bmatrix} \qquad (1)$$

$$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -k_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & k_1 \end{bmatrix} \qquad (2)$$

Since $k_1 > 0$, the product in (1) represents a compression or expansion along the x-axis followed by a reflection about the y-axis, and (2) represents a compression or expansion along the y-axis followed by a reflection about the x-axis. In the case where $k = -1$, transformations (1) and (2) are simply reflections about the y-axis and x-axis, respectively.

We know from that an invertible matrix can be expressed as a product of elementary matrices, so above theorem implies the following result.

Theorem 4.3 If $T_A : R^2 \to R^2$ is multiplication by an invertible matrix A, then the geometric effect of T_A is the same as an appropriate succession of shears, compressions, expansions, and reflections.

Example 4. In Example 2 we illustrated the effect on the unit square of multiplication by A $= \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$ (see Figure 4.2).

Express this matrix as a product of elementary matrices, and then describe the effect of multiplication by A in terms of shears, compressions, expansions, and reflections.

Solution. The matrix A can be reduced to the identity matrix as follows:

$$\begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

These three successive row operations can be performed by multiplying A on the left successively by

$$E_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, E_2 = \begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}, E_3 = \begin{bmatrix} 1 & -1/2 \\ 0 & 1 \end{bmatrix}$$

Inverting these matrices we have

$$\text{A} = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/2 \\ 0 & 1 \end{bmatrix}$$

Reading from right to left we can now see that the geometric effect of multiplying by A is equivalent to successively

1. shearing by a factor of 1

2 in the x-direction

2. expanding by a factor of 2 in the x-direction

3. reflecting about the line $y = x$.

This is shown in Figure 4.6, which agrees with that in Ex. 2.

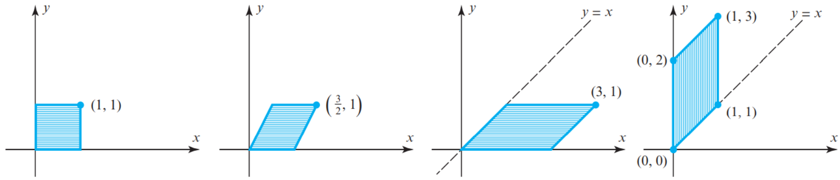


Figure 4.6

Example 5. Discuss the geometric effect on the unit square of multiplication by a diagonal matrix A $= \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$ in which the entries k_1 and k_2 are positive real numbers ($\neq 1$).

Solution. The matrix A is invertible and can be expressed as

$$A = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & k_2 \end{bmatrix} \begin{bmatrix} k_1 & 0 \\ 0 & 1 \end{bmatrix}$$

which show that multiplication by A causes a compression or expansion of the unit square by a factor of k_1 in the x-direction followed by an expansion or compression of the unit square by a factor of k_2 in the y-direction.

Example 6. As illustrated in Figure 4.7, multiplication by the matrix A $= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ has the geometric effect of reflecting the unit square about the origin. Note, however, that the matrix equation

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

together with Table 1 shows that the same result can be obtained by first reflecting the unit square about the x-axis and then reflecting that result about the y-axis. See Figure 4.7.

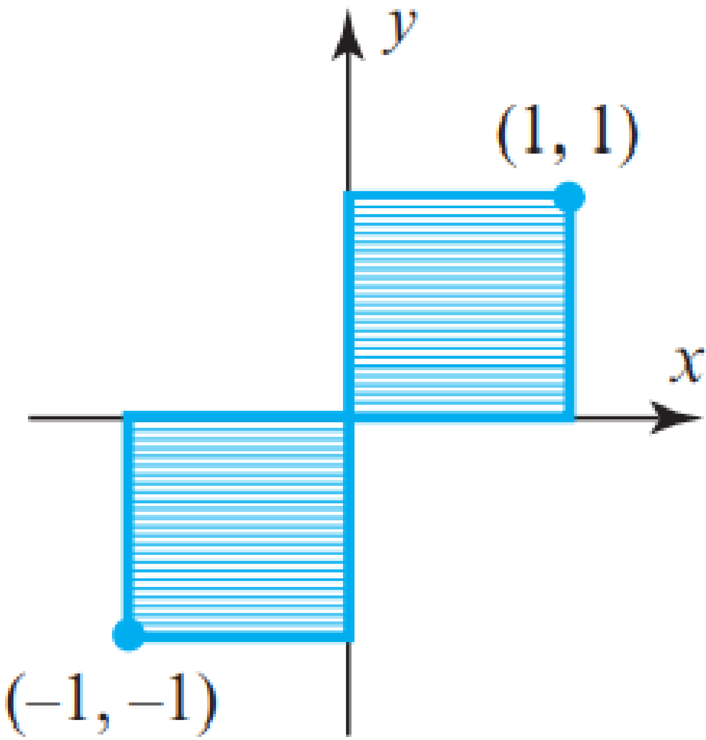


Figure 4.7

Example 7. Reflection About the Line $y = -x$

Verify that multiplication by the matrix A$=\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ reflects the unit square about the line $y = -x$ (Figure 4.8).

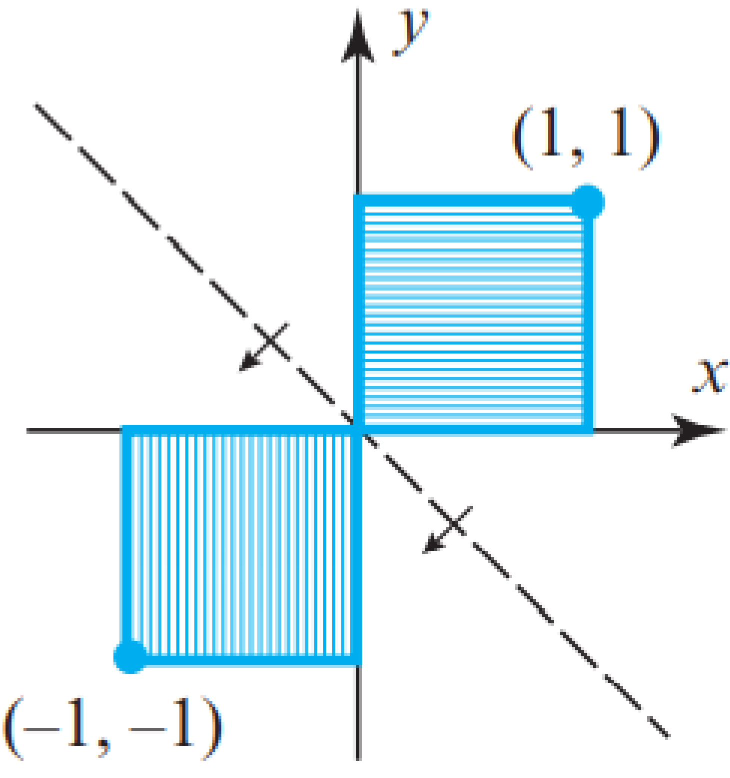


Figure 4.8

4.2 Eigenvalues and Eigenvectors

If A is an n $\times$ n matrix, then a nonzero vector x in R^n is called an **eigenvector** of A (or of the matrix operator T_A) if Ax is a scalar multiple of x; that is, Ax $= \lambda$x for some scalar λ. The scalar λ is called an **eigenvalue** of A (or of T_A), and x is said to be an **eigenvector corresponding** to λ.

Example 8. The vector x $= \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector of

A $= \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$ corresponding to the eigenvalue $\lambda = 3$, since

$$Ax = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3x.$$

Computing Eigenvalues and Eigenvectors

If A is an n $\times$ n matrix, then λ is an eigenvalue of A if and only if it satisfies the equation det(λI - A) = 0. (1)

This is called the **characteristic equation** of A.

When the determinant det(λI - A) in (1) is expanded, the characteristic equation of A takes the form

$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$ (2)

where the left side of this equation is a polynomial of degree n in which the coefficient of λ^n is 1. The polynomial

$p(\lambda) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n$ (3)

is called the **characteristic polynomial** of A.

Since a polynomial of degree n has at most n distinct roots, it follows from (2) that the characteristic equation of an n $\times$ n matrix A has at most n distinct solutions and consequently the matrix has at most n distinct eigenvalues.

Solving roots of the equation (2) are called the eigen values of A.

Steps to find eigen values

1. Consider the matrix λI - A and det(λI - A).

(characteristic matrix and characteristic polynomial respectively)

2. Equate det(λI - A)=0. (characteristic equation)

3. Find the roots of characteristic equation, which are the eigen values.

Example 9. Find the eigenvalues of $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$.

The characteristic polynomial of A is det(λI - A)

$$= \det \begin{bmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -4 & 17 & \lambda - 8 \end{bmatrix} = \lambda^3 - 8\lambda^2 + 17\lambda - 4.$$

The eigenvalues of A must therefore satisfy the cubic equation

$\lambda^3 - 8\lambda^2 + 17\lambda - 4 = 0 \Rightarrow (\lambda - 4)(\lambda^2 - 4\lambda + 1) = 0 \Rightarrow$

$\lambda = 4, \quad \lambda = 2 + \sqrt{3}, \quad \lambda = 2 - \sqrt{3}.$

Example 10. Find the eigenvalues of the upper triangular matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}.$$

The characteristic equation, det(λI - A)=0

$$\Rightarrow \det \begin{bmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & -a_{14} \\ 0 & \lambda - a_{22} & -a_{23} & -a_{24} \\ 0 & 0 & \lambda - a_{33} & -a_{34} \\ 0 & 0 & 0 & \lambda - a_{44} \end{bmatrix} = 0$$

$\Rightarrow (\lambda - a_{11})(\lambda - a_{22})(\lambda - a_{33})(\lambda - a_{44}) = 0$

$\Rightarrow$ the eigen values are $\lambda = a_{11}, \lambda = a_{22}, \lambda = a_{33}, \lambda = a_{44}$

which are precisely the diagonal entries of A.

Example 11. Find the eigenvalues of the lower triangular matrix

$$A=\begin{bmatrix} 1/2 & 0 & 0 \\ -1 & 2/3 & 0 \\ 5 & -8 & -1/4 \end{bmatrix}.$$

Proceeding as in the above example, we get eigen values are precisely the diagonal entries of A.

Theorem 4.4. If A is an n $\times$ n triangular matrix (upper triangular, lower triangular, or diagonal), then the eigenvalues of A are the entries on the main diagonal of A.

Proof. Let $A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}.$

The characteristic equation, det(λI - A)=0

$$\Rightarrow \det \begin{bmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ 0 & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda - a_{nn} \end{bmatrix} = 0$$

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

$\Rightarrow$ the eigen values are $\lambda = a_{11}, \lambda = a_{22}, \dots, \lambda = a_{nn}$

which are precisely the diagonal entries of A.

Theorem 4.5. If A is an n $\times$ n matrix, the following statements are equivalent.

(a) λ is an eigenvalue of A.

(b) λ is a solution of the characteristic equation det(λI - A) = 0.

(c) The system of equations (λI - A)x = 0 has nontrivial solutions.

(d) There is a nonzero vector x such that Ax = λx.

Finding Eigenvectors and Bases for Eigenspaces

Now that we know how to find the eigenvalues of a matrix, we will consider the problem of finding the corresponding eigenvectors. By definition, the eigenvectors of A corresponding to an eigenvalue λ are the nonzero vectors that satisfy (λI - A)x = 0 Thus, we can find the eigenvectors of A corresponding to λ by finding the nonzero vectors in the solution space of this linear system. This solution space, which is called the eigenspace of A corresponding to λ, can also be viewed as:

1. the null space of the matrix λI - A
2. the kernel of the matrix operator $T_{\lambda I-A} : R^n \to R^n$
3. the set of vectors for which Ax = λx.

Notice that x = 0 is in every eigenspace but is not an eigen vector. This is the only vector that distinct eigenspaces have in common.

Example 12. Find bases for the eigenspaces of the matrix A= $\begin{bmatrix} -1 & 3 \\ 2 & 0 \end{bmatrix}$.

Solution. The characteristic equation of A is

$$\det \begin{bmatrix} \lambda+1 & -3 \\ -2 & \lambda \end{bmatrix} = 0$$

$\Rightarrow \lambda(\lambda+1) - 6 = (\lambda-2)(\lambda+3) = 0 \Rightarrow \lambda = 2, \lambda = -3.$

So the eigenvalues of A are λ=2 and λ= -3. Thus, there are two eigenspaces of A, one for each eigenvalue.

By definition, x = $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is an eigenvector of A corresponding to an eigenvalue λ if and only if (λI - A)x = 0,

that is, $\begin{bmatrix} \lambda+1 & -3 \\ -2 & \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

In the case where $\lambda = 2$ this equation becomes

$$\begin{bmatrix} 3 & -3 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

whose general solution is $x_1 = t, x_2 = t$ (since solving we get $x_1 = x_2$).

Since this can be written in matrix form as

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} t \\ t \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

it follows that $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a basis for the eigenspace corresponding to $\lambda = 2$.

Similarly $\begin{bmatrix} -3/2 \\ 1 \end{bmatrix}$ is a basis for the eigenspace corresponding to $\lambda =$ -3.

Example 13. Find bases for the eigenspaces of A= $\begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$.

Solution. The characteristic equation of A is $\lambda^3 - 5\lambda^2 + 8\lambda - 4 = 0$, or $(\lambda - 1)(\lambda - 2)^2 = 0$. Thus, the distinct eigenvalues of A are $\lambda = 1$ and $\lambda = 2$, so there are two eigenspaces of A.

By definition, $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is an eigenvector of A corresponding to λ if and only if x is a nontrivial solution of (λI - A)x = 0, or in matrix form,

$$\begin{bmatrix} \lambda & 0 & 2 \\ -1 & \lambda - 2 & -1 \\ -1 & 0 & \lambda - 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \qquad (*)$$

In the case where $\lambda = 2$, (*) becomes

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

Solving this system using Gaussian elimination yields

$x_1 = -s, x_2 = t, x_3 = s.$

Thus, the eigenvectors of A corresponding to $\lambda = 2$ are the nonzero vectors of the form

$$x = \begin{bmatrix} -s \\ t \\ s \end{bmatrix} = \begin{bmatrix} -s \\ 0 \\ s \end{bmatrix} + \begin{bmatrix} 0 \\ t \\ 0 \end{bmatrix} = s\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + t\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Since $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ are linearly independent, these vectors form a basis for the eigenspace corresponding to $\lambda = 2$.

If $\lambda = 1$, then (*) becomes

$\begin{bmatrix} 1 & 0 & 2 \\ -1 & -1 & -1 \\ -1 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. Solving this system yields

$x_1 = -2s, x_2 = s, x_3 = s.$

Thus, the eigenvectors corresponding to $\lambda = 1$ are the nonzero vectors of the form $x = \begin{bmatrix} -2s \\ s \\ s \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$. Thus $\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ is a basis for the eigenspace corresponding to $\lambda = 1$.

4.2.1 Eigenvalues and Invertibility

Theorem 4.6. A square matrix A is invertible if and only if $\lambda = 0$ is not an eigen value of A.

Proof. Assume that A is an n $\times$ n matrix. Observe first that $\lambda = 0$ is a solution of the characteristic equation $\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$ if and only if the constant term c_n is zero. Thus, it suffices to prove that A is invertible if and only if $c_n \neq 0$. But det(λI - A) $= \lambda^n + c_1\lambda^{n-1} + \cdots + c_n$.

On setting $\lambda = 0$, det(-A) = c_n or (-1)^n det(A) = c_n.

It follows from the last equation that det(A) = 0 if and only if $c_n = 0$, and this in turn implies that A is invertible if and only if $c_n \neq 0$.

Example 14. The matrix A=$\begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$ in the above example is invertible since it has eigenvalues $\lambda = 1$ and $\lambda = 2$, neither of which is zero. One can also check that det(A) $\neq 0$.

Theorem 4.7. - Equivalent Statements

If A is an n $\times$ n matrix, then the following statements are equivalent.

(a) A is invertible.

(b) Ax $=$ 0 has only the trivial solution.

(c) The reduced row echelon form of A is I_n.

(d) A is expressible as a product of elementary matrices.

(e) Ax $=$ b is consistent for every n $\times$ 1 matrix b.

(f) Ax $=$ b has exactly one solution for every n $\times$ 1 matrix b.

(g) det(A) $\neq 0$.

(h) The column vectors of A are linearly independent.

(i) The row vectors of A are linearly independent.

(j) The column vectors of A span R^n.

(k) The row vectors of A span R^n.

(l) The column vectors of A form a basis for R^n.

(m) The row vectors of A form a basis for R^n.

(n) A has rank n.

(o) A has nullity 0.

(p) The orthogonal complement of the null space of A is R^n.

(q) The orthogonal complement of the row space of A is 0.

(r) The kernel of T_A is 0.

(s) The range of T_A is R^n.

(t) T_A is one-to-one.

(u) $\lambda = 0$ is not an eigenvalue of A.

4.2.2 Eigenvalues of General LinearTransformations

If T : V $\rightarrow$ V is a linear operator on a vector space V, then a nonzero vector $\mathbf{x}$ in V is called an **eigenvector** of T if T($\mathbf{x}$) is a scalar multiple of $\mathbf{x}$; that is, T($\mathbf{x}$) = $\lambda\mathbf{x}$ for some scalar λ. The scalar λ is called an **eigenvalue** of T, and $\mathbf{x}$ is said to be an **eigenvector corresponding** to λ.

Problem.

Find the eigenvalues and the bases for the eigenspace corresponding to the eigenvalues of the follwing matrices:

1. $\begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}$. 2. $\begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}$. 3. $\begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}$.

4.3 Diagonalization

Products of the form $P^{-1}AP$ in which A and P are n $\times$ n matrices and P is invertible will be our main topic of study in this section. There are various ways to think about such products, one of which is to view them as transformations

$A \to P^{-1}AP$ in which the matrix A is mapped into the matrix $P^{-1}AP$. These are called **similarity transformations**. Such transformations are important because they preserve many properties of the matrix A.

For example, if we let $B = P^{-1}AP$, then A and B have the same determinant since

$\det(B) = \det(P^{-1}AP) = \det(P^{-1}) \det(A) \det(P)$

$= 1/\det(P)\ \det(A)\ \det(P) = \det(A)$.

In general, any property that is preserved by a similarity transformation is called a **similarity invariant** and is said to be **invariant under similarity**. Table below lists the most important similarity invariants.

Property	**Description**
Determinant	A and $P^{-1}AP$ have the same determinant.
Invertibility	A is invertible if and only if $P^{-1}AP$ is invertible.
Rank	A and $P^{-1}AP$ have the same rank.
Nullity	A and $P^{-1}AP$ have the same nullity.
Trace	A and $P^{-1}AP$ have the same trace.
Characteristic polynomial	A and $P^{-1}AP$ have the same characteristic polynomial.
Eigenvalues	A and $P^{-1}AP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of A (and hence of $P^{-1}AP$) then the eigenspace of A corresponding to λ and the eigenspace of $P^{-1}AP$ corresponding to λ have the same dimension.

If A and B are square matrices, then we say that B is **similar** to A if there is an invertible matrix P such that $B = P^{-1}AP$.

Note that if B is similar to A, then it is also true that A is similar to B since we can express A as $A = Q^{-1}BQ$ by taking $Q = P^{-1}$. This being the case, we will usually say that A and B are ***similar matrices*** *if either is similar to the other.*

A square matrix A is said to be **diagonalizable** if it is similar to some diagonal matrix; that is, if there exists an invertible matrix P such that $\mathrm{P}^{-1}\mathrm{AP}$ is diagonal. In this case the matrix P is said to **diagonalize** A.

Theorem 4.8. If A is an n $\times$ n matrix, the following statements are equivalent.

(a) A is diagonalizable.

(b) A has n linearly independent eigenvectors.

Proof. (a) $\Rightarrow$ (b)

Since A is assumed to be diagonalizable, it follows that there exist an invertible matrix P and a diagonal matrix D such that $\mathrm{P}^{-1}\mathrm{AP} = \mathrm{D}$ or, equivalently, $\mathrm{AP} = \mathrm{PD}$ (1).

If we denote the column vectors of P by $p_1, p_2, \ldots, p_n$, and if we assume that the diagonal entries of D are $\lambda_1, \lambda_2, \ldots, \lambda_n$. We know that to multiply a matrix P on the right by a diagonal matrix D, multiply successive columns of P by the successive diagonal entries of D. So the left side of (1) can be expressed as

$$AP = A\begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix} = \begin{bmatrix} Ap_1 & Ap_2 & \dots & Ap_n \end{bmatrix}$$

and, the right side of (1) can be expressed as

$$PD = \begin{bmatrix} \lambda_1 p_1 & \lambda_2 p_2 & \dots & \lambda_n p_n \end{bmatrix}.$$

Thus, it follows from (1) that

$$Ap_1 = \lambda_1 p_1, \quad Ap_2 = \lambda_2 p_2, \quad \dots, \quad Ap_n = \lambda_n p_n \qquad (2).$$

Since P is invertible, we know from Theorem 4.7. above that its column vectors $p_1, p_2, \dots, p_n$ are linearly independent (and hence nonzero). Thus, it follows from (2) that these n column vectors are eigenvectors of A.

(b) $\Rightarrow$ (a)

Assume that A has n linearly independent eigenvectors, $p_1, p_2, \dots, p_n$, and that $\lambda_1, \lambda_2, \dots, \lambda_n$ are the corresponding eigenvalues. If we let $P = \begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix}$ and if we let D be the diagonal matrix that has $\lambda_1, \lambda_2, \dots, \lambda_n$ as its successive diagonal entries, then

$AP = A\begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix} = \begin{bmatrix} Ap_1 & Ap_2 & \dots & Ap_n \end{bmatrix} = \begin{bmatrix} \lambda_1 p_1 & \lambda_2 p_2 & \dots & \lambda_n p_n \end{bmatrix} = PD$. Since the column vectors of P are linearly independent, it follows from Theorem 4.7. above that P is invertible, so that this last equation can be rewritten as $P^{-1}AP = D$, which shows that A is diagonalizable.

Theorem 4.9. (a) If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of a matrix A, and if $v_1, v_2, \ldots, v_k$ are corresponding eigenvectors, then $\{v_1, v_2, \ldots, v_k\}$ is a linearly independent set.

(b) An n $\times$ n matrix with n distinct eigenvalues is diagonalizable.

Proof. (a) Let $v_1, v_2, \ldots, v_k$ be eigenvectors of A corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. We will assume that $v_1, v_2, \ldots, v_k$ are linearly dependent and obtain a contradiction. We can then conclude that $v_1, v_2, \ldots, v_k$ are linearly independent.

Since an eigenvector is nonzero by definition, $\{v_1\}$ is linearly independent. Let r be the largest integer such that $\{v_1, v_2, \ldots, v_r\}$ is linearly independent. Since we are assuming that $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent, r satisfies $1 \leq r < k$. Moreover, by the definition of r, $\{v_1, v_2, \ldots, v_{r+1}\}$ is linearly dependent. Thus, there are scalars $c_1, c_2, \ldots, c_{r+1}$, not all zero, such that

$$c_1 v_1 + c_2 v_2 + \cdots + c_{r+1} v_{r+1} = 0 \qquad (1)$$

Multiplying both sides of (1) by A and using the fact that

$Av_1 = \lambda_1 v_1, Av_2 = \lambda_2 v_2, \ldots, Av_{r+1} = \lambda_{r+1} v_{r+1}$

we obtain $c_1\lambda_1 v_1 + c_2\lambda_2 v_2 + \cdots + c_{r+1}\lambda_{r+1}v_{r+1} = 0$ (2)

If we now multiply both sides of (1) by λ_{r+1} and subtract the resulting equation from (2) we obtain

$c_1(\lambda_1 - \lambda_{r+1})v_1 + c_2(\lambda_2 - \lambda_{r+1})v_2 + \cdots + cr(\lambda_r - \lambda_{r+1})vr = 0.$

Since $\{v_1, v_2, \ldots, v_r\}$ is a linearly independent set, this equation implies that

$c_1(\lambda_1 - \lambda_{r+1}) = c_2(\lambda_2 - \lambda_{r+1}) = \cdots = cr(\lambda_r - \lambda_{r+1}) = 0$ and since $\lambda_1, \lambda_2, \ldots, \lambda_{r+1}$ are assumed to be distinct, it follows that

$c_1 = c_2 = \cdots = c_r = 0.$ (3)

Substituting these values in (1) yields $c_{r+1}v_{r+1} = 0.$

Since the eigenvector v_{r+1} is nonzero, it follows that $c_{r+1} = 0.$ (4)

But equations (3) and (4) contradict the fact that $c_1, c_2, \ldots, c_{r+1}$ are not all zero so the proof is complete.

(b) Let A be an n $\times$ n matrix with n distinct eigenvalues. By (a) part of this theorem, the n eigenvectors corresponding to these eigenvalues are linearly independent. Since A has n linearly independent eigenvectors, by Theorem 8. A is diagonalizable.

The Converse of Theorem 4.9(b) Is False

Consider the matrices $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Both of these matrices have only one distinct eigenvalue, namely $\lambda = 1$, and hence only one eigenspace. One can solve the characteristic equations

$(\lambda I - I)\mathbf{x} = 0$ and $(\lambda I - J)\mathbf{x} = 0$ with $\lambda = 1$

and show that for I the eigenspace is three-dimensional (all of R^3) and for J it is one-dimensional, consisting of all scalar multiples of $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

This shows that the converse of Theorem 4.9(b) is false, since we have produced two 3×3 matrices with fewer than three distinct eigenvalues, one of which is diagonalizable and the other of which is not.

A Procedure for Diagonalizing an $n \times n$ Matrix

Step 1. Determine first whether the matrix is actually diagonalizable by searching for n linearly independent eigenvectors.

One way to do this is to find a basis for each eigenspace and count the total number of vectors obtained. If there is a total of n vectors, then the matrix is diagonalizable, and if the total is less than n, then it is not.

Step 2. If you ascertained that the matrix is diagonalizable, then form the matrix $P = \begin{bmatrix} p_1 & p_2 & \dots & p_n \end{bmatrix}$ whose column vectors are the n basis vectors you obtained in Step 1.

Step 3. $P^{-1}AP$ will be a diagonal matrix whose successive diagonal entries are the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ that correspond to the successive columns of P.

Example 15. Find a matrix P that diagonalizes A= $\begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$.

We have already found in Example 13. that the characteristic equation of A to be $(\lambda - 1)(\lambda - 2)^2 = 0$ and we found the following bases for the eigenspaces:

$$\lambda = 2 : p_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ and } p_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

$$\lambda = 1 : p_3 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}.$$

There are three basis vectors in total, so the matrix

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

One can verify that $P^{-1}AP = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

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

Example 16. Show that the following matrix is not diagonalizable: $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ -3 & 5 & 2 \end{bmatrix}.$

Solution. The characteristic equation of A is

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

$\Rightarrow (\lambda-1)(\lambda-2)^2 = 0 \Rightarrow \lambda = 1, \quad \lambda = 2$. Thus the distinct eigenvalues of A are $\lambda = 1$ and $\lambda = 2$. We can find that the bases for the eigenspaces are

$$\lambda = 1 : p_1 = \begin{bmatrix} 1/8 \\ -1/8 \\ 1 \end{bmatrix} \qquad \lambda = 2 : p_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Since A is a 3×3 matrix and there are only two basis vectors in total, A is not diagonalizable.

Example 17. We saw in Example 9. that A= $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$ has three eigen values $\lambda = 4, \quad \lambda = 2+\sqrt{3}, \quad \lambda = 2-\sqrt{3}$.

So A is diagonalizable and

$\mathrm{P}^{-1}\mathrm{AP} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2+\sqrt{3} & 0 \\ 0 & 0 & 2-\sqrt{3} \end{bmatrix}$ for some invertible matrix P^{-1}.

Example 18. From Theorem 4.4., the eigenvalues of a triangular matrix are the entries on its main diagonal. Thus, a triangular matrix with distinct entries on the main diagonal is diagonalizable. For example,

$$A = \begin{bmatrix} -1 & 2 & 4 & 0 \\ 0 & 3 & 1 & 7 \\ 0 & 0 & 5 & 8 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

is a diagonalizable matrix with eigenvalues

$\lambda_1 = -1, \lambda_2 = 3, \lambda_3 = 5, \lambda_4 = -2.$

Theorem 4.10. If k is a positive integer, λ is an eigenvalue of a matrix A, and $\mathbf{x}$ is a corresponding eigenvector, then λ^k is an eigenvalue of A^k and $\mathbf{x}$ is a corresponding eigenvector.

Example 19. The eigenvalues and corresponding eigenvectors of the matrix $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ -3 & 5 & 2 \end{bmatrix}$ are found in Example 16.

The eigenvalues of A are $\lambda = 1$ and $\lambda = 2$, so the eigenvalues of A^7 are $\lambda = 1^7 = 1$ and $\lambda = 2^7$ =128.

The eigenvectors p_1 and p_2 obtained in Example 16. corresponding to the eigenvalues $\lambda = 1$ and $\lambda = 2$ of A are also the eigenvectors corresponding to the eigenvalues $\lambda = 1$ and $\lambda = 128$ of A^7.

Computing Powers of a Matrix

The problem of computing powers of a matrix is greatly simplified when the matrix is diagonalizable. Suppose that A is a diagonalizable n × n matrix, that P diagonalizes A, and

that $P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = D$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigen values of A.

Squaring both sides of this equation yields

$$(P^{-1}AP)^2 = \begin{bmatrix} \lambda_1^2 & 0 & \dots & 0 \\ 0 & \lambda_2^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n^2 \end{bmatrix} = D^2.$$

We can rewrite the left side of this equation as

$(P^{-1}AP)^2 = P^{-1}AP\,P^{-1}AP = P^{-1}AIAP = P^{-1}A^2P.$

from which we obtain the relationship $P^{-1}A^2P = D^2$. More generally, if k is a positive integer, then a similar computation

will show that $P^{-1}A^kP = D^k = \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix}$,

which we can rewrite as

$$A^k = PD^kP^{-1} = P \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix} P^{-1}.$$

Example 20. Find A^{13}, where A= $\begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$.

As done in Example 13. we can show that A is diagonalizable and

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

where P= $\begin{bmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ and P^{-1}= $\begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix}$.

Hence $A^{13} = PD^{13}P^{-1} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2^{13} & 0 & 0 \\ 0 & 2^{13} & 0 \\ 0 & 0 & 1^{13} \end{bmatrix} \begin{bmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$

$$= \begin{bmatrix} -8190 & 0 & -16382 \\ 8191 & 8192 & 8191 \\ 8191 & 0 & 16383 \end{bmatrix}.$$

4.3.1 Geometric and Algebraic Multiplicity

It can be proved that if λ_0 is an eigenvalue of A, then the dimension of the eigenspace corresponding to λ_0 cannot exceed the number of times that λ - λ_0 appears as a factor of the characteristic polynomial of A. For example, in Examples 15 and 16 the characteristic polynomial is $(\lambda - 1)(\lambda - 2)^2$. Thus, the eigenspace corresponding to $\lambda = 1$ is at most (hence exactly) one-dimensional, and the eigenspace corresponding to $\lambda = 2$ is at most two-dimensional. In Example 15 the eigenspace corresponding to $\lambda = 2$ actually had dimension 2, resulting in diagonalizability, but in Example 16 the eigenspace corresponding to $\lambda = 2$ had only dimension 1, resulting in non diagonalizability.

There is some terminology that is related to these ideas. If λ_0

is an eigenvalue of an n $\times$ n matrix A, then the dimension of the eigenspace corresponding to λ_0 is called the **geometric multiplicity** of λ_0, and the number of times that λ - λ_0 appears as a factor in the characteristic polynomial of A is called the **algebraic multiplicity** of λ_0.

Theorem 4.11 - Geometric and Algebraic Multiplicity

If A is a square matrix, then:

(a) For every eigenvalue of A, the geometric multiplicity is less than or equal to the algebraic multiplicity.

(b) A is diagonalizable if and only if the geometric multiplicity of every eigenvalue is equal to the algebraic multiplicity.

4.4 Inner Product Spaces

An inner product is a generalization of the dot product.

An **inner product** on a real vector space V is a function that associates a real number $\langle u, v\rangle$ with each pair of vectors in V in such a way that the following axioms are satisfied for all vectors u, v, and w in V and all scalars k.

1. $\langle u, v\rangle = \langle v, u\rangle$ [Symmetry axiom]
2. $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ [Additivity axiom]
3. $\langle ku, v\rangle = k\langle u, v\rangle$ [Homogeneity axiom]
4. $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ if and only if v = 0 [Positivity axiom]

A real vector space with an inner product is called a **real inner product space**.

Because the axioms for a real inner product space are based on properties of the dot product, these inner product space axioms will be satisfied automatically if we define the inner product of two vectors u and v in R^n to be

$\langle u, v\rangle = u.v = u_1v_1 + u_2v_2 + \cdots + u_nv_n$.

This inner product is commonly called the **Euclidean inner product (or the standard inner product)** on R^n.

If V is a real inner product space, then the **norm (or length)** of a vector v in V is denoted by $\|v\|$ and is defined by

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

and the **distance** between two vectors is denoted by **d(u, v)** and is defined by

$d(u, v) = \|u - v\| = \sqrt{\langle u - v, u - v\rangle}$.

A vector of norm 1 is called a **unit vector**.

Theorem 4.12 If u and v are vectors in a real inner product space V, and if k is a scalar, then:

(a) $\|v\| \geq 0$with equality if and only if v = 0.

(b) $\|kv\| = |k|\|v\|$.

(c) $d(u, v) = d(v, u)$.

(d) $d(u, v) \geq 0$ with equality if and only if u = v.

Proof. (a) Let v be any vector in a real inner product space V. Then by the definition of norm of a vector and positivity axiom of inner product, we get

$\|v\| = \sqrt{\langle v, v\rangle} \geq 0$ and

$\|v\| = \sqrt{\langle v, v\rangle} \geq 0 \Leftrightarrow \langle v, v\rangle = 0 \Leftrightarrow v = 0.$

(b) Let v be any vector in a real inner product space V and k be any scalar. Then by the definition of norm of a vector and homogeneity axiom of inner product, we get

$\|kv\| = \sqrt{\langle kv, kv\rangle} = \sqrt{k\langle v, kv\rangle} = \sqrt{k\langle kv, v\rangle} \quad (by\ symmetry\ axiom)$

$= \sqrt{k^2\langle v, v\rangle} = \sqrt{k^2}\sqrt{\langle v, v\rangle} = |k|\|v\|.$

(c) Let u and v any two vectors in a real inner product space V. Then from the definition of distance between two vectors, we obtain

$d(u, v) = \|u - v\| = \|(-1)(v - u)\|$

$= |(-1)|\|(v - u)\| \qquad$ by part (b)

$= \|(v - u)\| = d(v, u).$

(d) Let u and v any two vectors in a real inner product space V. Then from the definition of distance between two vectors and part (a) of this theorem, we get

$d(u, v) = \|u - v\| \geq 0$ and

$d(u, v) = \|u - v\| = 0 \Leftrightarrow u - v = 0 \Leftrightarrow u = v.$

Example 21. The Standard Inner Product on P_n

If $p = a_0 + a_1 x + \cdots + a_n x^n$ and $q = b_0 + b_1 x + \cdots + b_n x^n$ are polynomials in P_n, then the following formula defines an inner product on P_n that we will call the standard inner product on this space:

$\langle p, q \rangle = a_0 b_0 + a_1 b_1 + \cdots + a_n b_n.$

The norm of a polynomial p relative to this inner product is

$\|p\| = \sqrt{\langle p, p \rangle} = \sqrt{a_0^2 + a_1^2 + \cdots + a_n^2}$

Verification:

For any $p = a_0 + a_1 x + \cdots + a_n x^n$, $q = b_0 + b_1 x + \cdots + b_n x^n$ and $r = c_0 + c_1 x + \cdots + c_n x^n$ and for any real number k, we get

1. $\langle p, q \rangle = a_0 b_0 + a_1 b_1 + \cdots + a_n b_n$

$$= b_0 a_0 + b_1 a_1 + \cdots + b_n a_n = \langle q, p \rangle$$

2. $\langle p+q, r \rangle = (a_0 + b_0)c_0 + (a_1 + b_1)c_1 + \cdots + (a_n + b_n)c_n$

$$= (a_0 c_0 + a_1 c_1 + \cdots + a_n c_n) + (b_0 c_0 + b_1 c_1 + \cdots + b_n c_n)$$

$$= \langle p, r \rangle + \langle q, r \rangle.$$

3. $\langle kp, q \rangle = (ka_0)b_0 + (ka_1)b_1 + \cdots + (ka_n)b_n$

$$= k(a_0 b_0 + a_1 b_1 + \cdots + a_n b_n) = k\langle p, q \rangle.$$

4. $\langle p, p \rangle = a_0^2 + a_1^2 + \cdots + a_n^2 \geq 0,$

$\langle p, p \rangle = 0 \Leftrightarrow a_0^2 + a_1^2 + \cdots + a_n^2 = 0 \Leftrightarrow a_i = 0 \forall i \Leftrightarrow p = 0.$

Example 22. An Integral Inner Product on $C[a, b]$

Let $f = f(x)$ and $g = g(x)$ be two functions in $C[a, b]$ and define $\langle f, g \rangle = \int_a^b f(x)g(x)dx$.

We will show that this formula defines an inner product on C[a, b] by verifying the four inner product axioms for functions $f = f(x)$, $g = g(x), h = h(x)$ in $C[a, b]$:

Axiom 1: $\langle f, g \rangle = \int_a^b f(x)g(x)dx = \int_a^b g(x)f(x)dx = \langle g, f \rangle$.

Axiom 2: $\langle f+g, h \rangle = \int_a^b [f(x)+g(x)]h(x)dx = \int_a^b f(x)h(x)dx + \int_a^b g(x)h(x)dx = \langle f, h \rangle + \langle g, h \rangle$.

Axiom 3: $\langle kf, g \rangle = \int_a^b kf(x)g(x)dx = k\int_a^b f(x)g(x)dx = k\langle f, g \rangle$.

Axiom 4: $\langle f, f \rangle = \int_a^b f(x)f(x)dx = \int_a^b f^2(x)dx \geq 0,$

since $f^2(x) \geq 0$ for all x in the interval $[a, b]$. Moreover, because f is continuous on $[a, b]$, the equality in the above equation holds if and only if the function f is identically zero on $[a, b]$, that is, if and only if $f = 0$.

Problem

Show that the vector space P_2 of all polynomials with degree less than or equal to two, is an innerproduct spcae with inner product defined by $\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)dx \forall p = p(x), q = q(x) \in P_2$.

4.5 Angle and Orthogonality in Inner Product Spaces

We know that the angle θ between two vectors u and v in R^n is given by the formula

$$cos\theta = \frac{u.v}{\|u\|\|v\|} = \frac{\langle u, v \rangle}{\|u\|\|v\|},$$

where $\langle u, v \rangle$ is the standard inner product on R^n.

In sequel with the above formula, the angle θ between two vectors u and v in an innerproduct space is given by the formula

$$cos\theta = \frac{\langle u, v\rangle}{\|u\|\|v\|}.$$

Two vectors u and v in an inner product space V is called **orthogonal** if $\langle u, v\rangle = 0$.

Example 23. The vectors $u = (1, 1)$ and $v = (1, -1)$ are orthogonal with respect to the Euclidean inner product on R^2 since $u.v = (1)(1) + (1)(-1) = 0$.

However, they are not orthogonal with respect to the weighted Euclidean inner product $\langle u, v\rangle = 3u_1v_1 + 2u_2v_2$ since $\langle u, v\rangle = 3(1)(1) + 2(1)(-1) = 1 \neq 0$.

Example 24. Let P_2 have the inner product

$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx$ and let $p = x$ and $q = x^2$.

Now, $\|p\| = \langle p, p\rangle^{1/2} = [\int_{-1}^{1} xxdx]^{1/2} = [\int_{-1}^{1} x^2dx]^{1/2} = \sqrt{\frac{2}{3}}$

and $\|q\| = \langle q, q\rangle^{1/2} = [\int_{-1}^{1} x^2x^2dx]^{1/2} = [\int_{-1}^{1} x^4dx]^{1/2} = \sqrt{\frac{2}{5}}$.

Also, $\langle p, q\rangle = \int_{-1}^{1} xx^2dx = \int_{-1}^{1} x^3dx = 0$.

Because $\langle p, q\rangle = 0$, the vectors $p = x$ and $q = x^2$ are orthogonal relative to the given inner product.

Problem

Show that the vectors $u = (-4, 6, 1, 1)$, $(2, 1, -7, 9)$ are orthogonal with respect to the Euclidian inner product.

4.6 Gram–Schmidt Process

A set of two or more vectors in a real inner product space is said to be **orthogonal** if all pairs of distinct vectors in the set are orthogonal. An orthogonal set in which each vector has norm 1 is said to be **orthonormal**.

A simple way to convert an orthogonal set of nonzero vectors into an orthonormal set is to multiply each vector v in the orthogonal set by the reciprocal of its length to create a vector of norm 1 (called a unit vector). Suppose that v is a nonzero vector in an inner product space, and let

$u = \frac{1}{\|v\|}v.$

Then by the properties of norm of a vector, we obtain

$\|u\| = \|\frac{1}{\|v\|}v\| = |\frac{1}{\|v\|}| \; \|v\| = 1.$

This process of multiplying a vector v by the reciprocal of its norm(length) is called **normalizing v**.

Example 25. Let $v_1 = (0, 1, 0), v_2 = (1, 0, 1), v_3 = (1, 0, -1)$

and assume that R^3 has the Euclidean inner product. It follows that the set of vectors $S = \{v_1, v_2, v_3\}$ is orthogonal since $\langle v_1, v_2\rangle = \langle v_1, v_3\rangle = \langle v_2, v_3\rangle = 0$.

The Euclidean norms of the above vectors are

$\|v_1\| = 1,\ \|v_2\| = \sqrt{2},\ \|v_3\| = \sqrt{2}$.

Consequently, normalizing v_1, v_2, v_3 gives,

$u_1 = \dfrac{v_1}{\|v_1\|} = (0, 1, 0),\ u_2 = \dfrac{v_2}{\|v_2\|} = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}),$

$u_3 = \dfrac{v_3}{\|v_3\|} = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$.

Also, $\langle v_1, v_2\rangle = \langle v_1, v_3\rangle = \langle v_2, v_3\rangle = 0$ and

$\|u_1\|,\ \|u_2\|,\ \|u_3\| = 1$.

Hence the set $S = \{u_1, u_2, u_3\}$ is orthonormal in R^3.

Theorem 4.13 If $S = \{v_1, v_2, \ldots, v_n\}$ is an orthogonal set of nonzero vectors in an inner product space, then S is linearly independent.

Proof. Assume that $k_1 v_1 + k_2 v_2 + \cdots + k_n v_n = 0$ (1)

To demonstrate that $S = \{v_1, v_2, \ldots, v_n\}$ is linearly independent, we must prove that $k_1 = k_2 = \cdots = k_n = 0$. For each v_i in S, it follows from (1) that

$\langle k_1 v_1 + k_2 v_2 + \cdots + k_n v_n, v_i\rangle = \langle 0, v_i\rangle = 0$

or equivalently,

$k_1\langle v_1, v_i\rangle + k_2\langle v_2, v_i\rangle + \cdots + k_n\langle v_n, v_i\rangle = 0. \qquad (2)$

From the orthogonality of S it follows that $\langle v_j, v_i\rangle = 0$ when $j \neq i$, so this equation reduces to $k_i\langle v_i, v_i\rangle = 0$.

Since the vectors in S are assumed to be nonzero, it follows from the positivity axiom for inner products that $\langle v_i, v_i\rangle \neq 0$. Thus, the preceding equation implies that each k_i in equation (1) is zero. Thus S is linearly independent.

In an inner product space, a basis consisting of orthonormal vectors is called an **orthonormal basis**, and a basis consisting of orthogonal vectors is called an **orthogonal basis**.

A familiar example of an orthonormal basis is the standard basis for R^n with the Euclidean inner product:

$e_1 = (1, 0, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, 0, \ldots, 1).$

Example 26. Recall the standard inner product of the polynomials. If $p = a_0 + a_1x + \cdots + a_nx^n$ and $q = b_0 + b_1x + \cdots + b_nx^n$ are polynomials in P_n, then their inner product on P_n is:

$\langle p, q\rangle = a_0b_0 + a_1b_1 + \cdots + a_nb_n$ and

the norm of a polynomial p relative to this inner product is

$\|p\| = \sqrt{\langle p, p\rangle} = \sqrt{a_0^2 + a_1^2 + \cdots + a_n^2}$

Clearly the standard basis $S = \{1, x, x^2, \ldots, x^n\}$ is orthonormal with respect to this inner product.

Example 27. In Example 25. we see that $u_1 = (0,1,0)$, $u_2 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$, $u_3 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})$ form an orthonormal set with respect to the Euclidean inner product on R^3. By Theorem 4.13, these vectors form a linearly independent set, and since R^3 is three-dimensional, it follows that $S = \{u_1, u_2, u_3\}$ is an orthonormal basis for R^3.

Let V be an inner product space and $\{v_1, v_2, \ldots v_n\}$ be a basis for V. Let **u** be a vector in V and $\mathbf{u} = c_1v_1 + c_2v_2 + \cdots + c_nv_n$ where $c_1, c_2, \ldots, c_n$ are scalars. Then the scalars $c_1, c_2, \ldots, c_n$ are called the **coordinates** of **u** relative to S. The vector $(c_1, c_2, \ldots, c_n)$ in R^n constructed from these coordinates is called the **coordinate vector of u relative to S** and is denoted by $(\mathbf{u})_S = (c_1, c_2, \ldots, c_n)$.

One way to express a vector **u** as a linear combination of basis vectors $S = \{v_1, v_2, \ldots v_n\}$ is to convert the vector equation $u = c_1v_1 + c_2v_2 + \cdots + c_nv_n$ to a linear system and solve for the coefficients $c_1, c_2, \ldots, c_n$. However, if the basis happens to be orthogonal or orthonormal, then the following theorem shows that the coefficients can be obtained more simply by computing appropriate inner products.

Theorem 4.14 (a) If $S = \{v_1, v_2, \dots v_n\}$ is an orthogonal basis for an inner product space V, and if u is any vector in V, then
$u = \dfrac{\langle u, v_1\rangle}{\|v_1\|^2}v_1 + \dfrac{\langle u, v_2\rangle}{\|v_2\|^2}v_2 + \cdots + \dfrac{\langle u, v_n\rangle}{\|v_n\|^2}v_n.$
(b) If $S = \{v_1, v_2, \dots v_n\}$ is an orthonormal basis for an inner product space V, and if u is any vector in V, then

$u = \langle u, v_1\rangle v_1 + \langle u, v_2\rangle v_2 + \cdots + \langle u, v_n\rangle v_n.$

Proof. (a) Since $S = \{v_1, v_2, \dots v_n\}$ is a basis for V, every vector u in V can be expressed in the form

$u = c_1v_1 + c_2v_2 + \cdots + c_nv_n \qquad (1).$

We will complete the proof by showing that
$c_i = \dfrac{\langle u, v_i\rangle}{\|v_i\|^2}$
for $i = 1, 2, \dots, n$. To do this, observe first that

$$\begin{aligned}\langle u, v_i\rangle &= \langle c_1v_1 + c_2v_2 + \cdots + c_nv_n, v_i\rangle \\ &= c_1\langle v_1, v_i\rangle + c_2\langle v_2, v_i\rangle + \cdots + c_n\langle v_n, v_i\rangle.\end{aligned}$$

Since S is an orthogonal set, all of the inner products in the last equality are zero except the ith, so we have

$\langle u, v_i\rangle = c_i\langle v_i, v_i\rangle = c_i\|v_i\|^2.$
So $c_i = \dfrac{\langle u, v_i\rangle}{\|v_i\|^2}.$
This completes the proof.

(b) In this case $\|v_1\| = \|v_2\| = \cdots = \|v_n\| = 1$. So the formula obtained in part (a) reduces to

$u = \langle u, v_1\rangle v_1 + \langle u, v_2\rangle v_2 + \cdots + \langle u, v_n\rangle v_n$.

As a consequence of the above theorem, it follows that the coordinate vector of a vector u in V relative to an orthogonal basis $S = \{v_1, v_2, \ldots v_n\}$ is

$(u)_S = \left(\frac{\langle u, v_1\rangle}{\|v_1\|^2}, \frac{\langle u, v_2\rangle}{\|v_2\|^2}, \ldots, \frac{\langle u, v_n\rangle}{\|v_n\|^2}\right)$

and relative to an orthonormal basis $S = \{v_1, v_2, \ldots v_n\}$ is

$(u)_S = (\langle u, v_1\rangle, \langle u, v_2\rangle, \ldots, \langle u, v_n\rangle)$.

Example 28. Show hat $S = \{v_1, v_2, v_3\}$ is an orthonormal basis for R^3 with the Euclidean inner product where $v_1 = (0, 1, 0), v_2 = (-4/5, 0, 3/5), v_3 = (3/5, 0, 4/5)$. Express the vector $u = (1, 1, 1)$ as a linear combination of the vectors in S, and find the coordinate vector $(u)_S$.

Solution. One can verify that

$\langle u, v_1\rangle = 1,\ \langle u, v_2\rangle = \frac{-1}{5},\ \langle u, v_3\rangle = \frac{7}{5}$.

Therefore, by Theorem 4.14 we have

$u = v_1 - \frac{1}{5}v_2 + \frac{7}{5}v_3$

that is, $(1, 1, 1) = (0, 1, 0) - \frac{1}{5}(-4/5, 0, 3/5) + \frac{7}{5}(3/5, 0, 4/5)$.

Thus, the coordinate vector of u relative to S is

$(u)_S = (\langle u, v_1\rangle, \langle u, v_2\rangle, \langle u, v_3\rangle) = (1, \frac{-1}{5}, \frac{7}{5})$.

Example 29. (a) Show that the vectors

$w_1 = (0, 2, 0), w_2 = (3, 0, 3), w_3 = (-4, 0, 4)$

form an orthogonal basis for R^3 with the Euclidean inner product, and use that basis to find an orthonormal basis by normalizing each vector.

(b) Express the vector $u = (1, 2, 4)$ as a linear combination of the orthonormal basis vectors obtained in part (a).

Solution. The given vectors form an orthogonal set since

$\langle w_1, w_2\rangle = 0, \langle w_1, w_3\rangle = 0, \langle w_2, w_3\rangle = 0.$

It follows from Theorem 4.13 that these vectors are linearly independent and hence form a basis for R^3.

Also, we obtain $\|w_1\| = 2,\ \|w_2\| = 3\sqrt{2},\ \|w_3\| = 4\sqrt{2}$.

Let $v_1 = \dfrac{w_1}{\|w_1\|} = \dfrac{(0,2,0)}{2} = (0, 1, 0)$,

$v_2 = \dfrac{w_2}{\|w_2\|} = \dfrac{(3,0,3)}{3\sqrt{2}} = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$,

$v_3 = \dfrac{w_2}{\|w_2\|} = \dfrac{(-4,0,4)}{4\sqrt{2}} = \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$.

Then $\{v_1, v_2, v_3\}$ is an orthogonal basis for R^3 with the Euclidean inner product.

(b) It follows from Theorem 4.14(b) that

$u = \langle u, v_1\rangle v_1 + \langle u, v_2\rangle v_2 + \cdots + \langle u, v_n\rangle v_n.$

$\langle u, v_1 \rangle = (1, 2, 4).(0, 1, 0) = 2$
$\langle u, v_2 \rangle = (1, 2, 4).\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) = \frac{5}{\sqrt{2}}$
$\langle u, v_3 \rangle = (1, 2, 4).\left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) = \frac{3}{\sqrt{2}}$.
Hence $(1, 2, 4) = 2(0, 1, 0) + \frac{5}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right) + \frac{3}{\sqrt{2}}\left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)$.

Theorem 4.15 Every nonzero finite-dimensional inner product space has an orthonormal basis.

The step-by-step construction of an orthogonal (or orthonormal) basis is called the **Gram–Schmidt process**. A summary of the steps is as given below:

To convert a basis $u_1, u_2, \ldots, u_r$ into an orthogonal basis $v_1, v_2, \ldots, v_r$, perform the following computations:

Step 1. $v_1 = u_1$

Step 2. $v_2 = u_2 - \dfrac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1$

Step 3. $v_3 = u_3 - \dfrac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \dfrac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2$

Step 4. $v_4 = u_4 - \dfrac{\langle u_4, v_1 \rangle}{\|v_1\|^2} v_1 - \dfrac{\langle u_4, v_2 \rangle}{\|v_2\|^2} v_2 - \dfrac{\langle u_4, v_3 \rangle}{\|v_3\|^2} v_3$

$\vdots \qquad \vdots$

(*continue for r steps*)

Optional Step. To convert the orthogonal basis into an orthonormal basis $\{q_1, q_2, \ldots, q_r\}$, normalize the orthogonal basis vectors.

Example 30. Assume that the vector space R^3 has the Euclidean inner product. Apply the Gram-Schmidt process to transform the basis vectors $u_1 = (1,1,1), u_2 = (0,1,1), u_3 = (0,0,1)$ into an orthogonal basis $\{v_1, v_2, v_3\}$, and then normalize the orthogonal basis vectors to obtain an orthonormal basis $\{q_1, q_2, q_3\}$.

Solution. Step 1. $v_1 = u_1 = (1,1,1)$.

Step 2. $v_2 = u_2 - \dfrac{\langle u_2, v_1\rangle}{\|v_1\|^2} v_1$

$= (0,1,1) - 2/3\ (1,1,1) = (-2/3,\ 1/3,\ 1/3)$.

Step 3. $v_3 = u_3 - \dfrac{\langle u_3, v_1\rangle}{\|v_1\|^2} v_1 - \dfrac{\langle u_3, v_2\rangle}{\|v_2\|^2} v_2$

$= (0,0,1) - 1/3(1,1,1) - \dfrac{1/3}{2/3}(-2/3,\ 1/3,\ 1/3)$

$= (0,\ -1/2,\ 1/2)$.

Thus, $v_1 = (1,1,1), v_2 = (-2/3,\ 1/3,\ 1/3), v_3 = (0,\ -1/2,\ 1/2)$ form an orthogonal basis for R^3. The norms of these vectors are $\|v_1\| = \sqrt{3},\quad \|v_2\| = \dfrac{\sqrt{6}}{3},\quad \|v_3\| = \dfrac{1}{\sqrt{2}}$.

So an orthonormal basis for R^3 is

$$q_1 = \frac{v_1}{\|v_1\|} = \left(\tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}\right), \qquad q_2 = \frac{v_2}{\|v_2\|} = \left(-\tfrac{2}{\sqrt{6}}, \tfrac{2}{\sqrt{6}}, \tfrac{2}{\sqrt{6}}\right),$$

$$q_3 = \frac{v_2}{\|v_2\|} = \left(0, -\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right).$$

Example 31. Let the vector space P_2 have the inner product $\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx$.

Apply the Gram–Schmidt process to transform the standard basis $\{1, x, x^2\}$ for P_2 into an orthogonal basis $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$.

Solution. Take $u_1 = 1, u_2 = x$, and $u_3 = x^2$.

Step 1. $v_1 = u_1 = 1$.

Step 2. We have $\langle u_2, v_1 \rangle = \int_{-1}^{1} x dx = 0$.

So $v_2 = u_2 - \dfrac{\langle u_2, v_1 \rangle}{\|v_1\|^2} v_1 = u_2 = x$.

Step 3. We have $\langle u_3, v_1 \rangle = \int_{-1}^{1} x^2 dx = 2/3$,

$\langle u_3, v_2 \rangle = \int_{-1}^{1} x^3 dx = 0$ and

$\|v_1\|^2 = \langle v_1, v_1 \rangle = \int_{-1}^{1} 1 dx = 2$.

So $v_3 = u_3 - \dfrac{\langle u_3, v_1 \rangle}{\|v_1\|^2} v_1 - \dfrac{\langle u_3, v_2 \rangle}{\|v_2\|^2} v_2 = x^2 - 1/3$.

Thus, we have obtained the orthogonal basis $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$ in which $\phi_1(x) = 1, \quad \phi_2(x) = x, \quad \phi_3(x) = x^2 - 1/3$.

Theorem 4.16 If W is a finite-dimensional inner product space, then:

(a) Every orthogonal set of nonzero vectors in W can be enlarged to an orthogonal basis for W.

(b) Every orthonormal set in W can be enlarged to an orthonormal basis for W.

4.7 Diagonalization

A square matrix A is said to be **orthogonal** if its transpose is the same as its inverse, that is, if

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

or, equivalently, if $\quad AA^T = I = A^T A.$

Example 32. Consider the matrix $\begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix}$.

Now, $A^T A = \begin{bmatrix} \frac{3}{7} & -\frac{6}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{3}{7} & \frac{6}{7} \\ \frac{6}{7} & \frac{2}{7} & -\frac{3}{7} \end{bmatrix} \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I.$

Hence $A^{-1} = A^T$ and so A is orthogonal.

Example 33. We know that the standard matrix for the rotation operator $T : R^2 \to R^2$ that moves points counterclockwise about the origin through a positive angle θ, is $\begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix}$.

The matrix is orthogonal for all choices of θ since

$$A^T A = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Similarly, we can show that the standard matrix for the rotation operators on R^3 are also orthogonal.

The standard matrices for the reflection operators on R^2 and R^3 that maps each point into its symmetric image about a fixed line or plane are orthogonal.

Observe that for the orthogonal matrices in the above example, both the row vectors and the column vectors form orthonormal sets with respect to the Euclidean innerproduct. This is a consequence of the following theorem.

Theorem 4.17 The following are equivalent for an n $\times$ n matrix A.

(a) A is orthogonal.

(b) The row vectors of A form an orthonormal set in R^n with the Euclidean innerproduct.

(c) The column vectors of A form an orthonormal set in R^n with the Euclidean innerproduct.

Proof. Let r_i be the i th row vector and c_j the j th column vector of A and let r_i^t be the i th row vector and c_j^t the j th column vector of A^T. Since transposing a matrix converts its columns to rows and rows to columns, it follows that $c_j^t = r_j$ and $r_i^t = c_i$, $\forall i, j = 1, 2, \ldots, n$.

(a) $\Leftrightarrow$ (b)

From the row-column rule for multiplication of matrices, we get the $(i,j)^{th}$ element of AA^T as $(AA^T)_{ij} = r_i c_j^t = r_i . r_j, \forall i, j = 1, 2, \ldots, n$.

So $AA^T = \begin{bmatrix} r_1 . r_1 & r_1 . r_2 & \ldots & r_1 . r_n \\ r_2 . r_1 & r_2 . r_2 & \ldots & r_2 . r_n \\ \vdots & \vdots & & \vdots \\ r_n . r_1 & r_n . r_2 & \ldots & r_n . r_n \end{bmatrix}$ (1)

It is evident from the formula (1) that AA^T=I if and only if $r_1 . r_1 = r_2 . r_2 = \cdots = r_n . r_n = 1$ and $r_i . r_j = 0$ when $i \neq j$. which are true if and only if $\{r_1, r_2, \ldots, r_n\}$ is an orthonormal set in R^n.

(a) $\Leftrightarrow$ (c)

From the row-column rule for multiplication of matrices, we get the $(i,j)^{th}$ element of AA^T as $(AA^T)_{ij} = r_i^t c_j = c_i . c_j, \forall i, j = 1, 2, \ldots, n$.

So $AA^T = \begin{bmatrix} c_1 . c_1 & c_1 . c_2 & \ldots & c_1 . c_n \\ c_2 . c_1 & c_2 . c_2 & \ldots & c_2 . c_n \\ \vdots & \vdots & & \vdots \\ c_n . c_1 & c_n . c_2 & \ldots & c_n . c_n \end{bmatrix}$ (1)

It is evident from the formula (1) that AA^T=I if and only if $c_1 . c_1 = c_2 . c_2 = \cdots = c_n . c_n = 1$ and $c_i . c_j = 0$ when $i \neq j$.

which are true if and only if $\{c_1, c_2, \ldots, c_n\}$ is an orthonormal set in R^n.

Theorem 4.18(a) The transpose of an orthogonal matrix is orthogonal.

(b) The inverse of an orthogonal matrix is orthogonal.

(c) A product of orthogonal matrices is orthogonal.

(d) If A is orthogonal, then det(A) = 1 or det(A) = -1.

Proof. (a) Let A be an orthogonal matrix. Then by definition

$\mathrm{AA}^T = \mathrm{I} = \mathrm{A}^T\mathrm{A}$.

Since $(\mathrm{A}^T)^T$=A, the above equation can be written as

$(\mathrm{A}^T)^T\mathrm{A}^T = \mathrm{I} = \mathrm{A}^T(\mathrm{A}^T)^T$.

Hence $(\mathrm{A}^T)^{-1}$=$(\mathrm{A}^T)^T$ and so A^T is orthogonal. Thus transpose of every orthogonal matrix is orthogonal.

(b) Let A be an orthogonal matrix. Then by definition, A is invertible and

$\mathrm{A}^{-1} = \mathrm{A}^T$. Then $(\mathrm{A}^{-1})^{-1} = (\mathrm{A}^T)^{-1}$.

We know if A is invertible, then its transpose is also invertible and $(\mathrm{A}^T)^{-1}$=$(\mathrm{A}^{-1})^T$.

Hence $(\mathrm{A}^{-1})^{-1} = (\mathrm{A}^T)^{-1} = (\mathrm{A}^{-1})^T$.

Hence A^{-1} is orthogonal. Thus inverse of every orthogonal matrix is orthogonal.

(c) Let A and B be two orthogonal matrices of same order. Since A and B are orthogonal,

$A^{-1} = A^T$ and $B^{-1} = B^T$.

Since A and B are invertible, $|A| \neq 0$ and $|B| \neq 0$. Hence $|AB| = |A||B| \neq 0$ and $|BA| = |B||A| \neq 0$.

Thus both AB and BA are nonsingular,

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

$(BA)^{-1} = A^{-1}B^{-1} = A^T B^T = (BA)^T$.

Thus both AB and BA are orthogonal. Thus product of orthogonal matrices is orthogonal.

(d) Let A be an orthogonal matrix.

A is orthogonal $\Rightarrow AA^T = I$

$\Rightarrow \det(AA^T) = \det(I)$

$\Rightarrow \det(A)\ \det(A^T) = 1$

$\Rightarrow \det(A)\ \det(A) = 1$ [since $\det(A)=\det(A^T)$]

$\Rightarrow [\det(A)^2] = 1$

$\Rightarrow \det(A) = \pm\ 1$.

Theorem 4.19 If A is an $n \times n$ matrix, then the following are equivalent.

(a) A is orthogonal.

(b) Ax=x for all x in R^n.

(c) Ax . Ay = x . y for all x and y in R^n.

The above theorem has a useful geometric interpretation when considered from the view point of matrix transformations: If A is an orthogonal matrix and $T_A : R^n \to R^n$ is multiplication by A, then we will call T_A an **orthogonal operator** on R^n.

4.8 Orthogonal Diagonalization

We know that two square matrices A and B are said to be **similar** if there is an invertible matrix P such that $P^{-1}AP = B$.

If A and B are square matrices, then we say that B is **orthogonally similar** to A if there is an orthogonal matrix P such that $B = P^TAP$.

Note that if B is orthogonally similar to A, then it is also true that A is orthogonally similar to B since we can express A as

$A = Q^TBQ$ by taking $Q = P^T$. This being the case we will say that A and B are orthogonally similar matrices if either is orthogonally similar to the other.

If A is orthogonally similar to some diagonal matrix, say $P^TAP = D$, then we say A is **orthogonally diagonalizable** and that P **orthogonally diagonalizes** A.

Suppose that A is orthogonally diagonalizable.

Then $P^TAP = D$ (1), where P is an orthogonal matrix and D is a diagonal matrix. Multiplying the left side of (1) by P, the right side by P^T, and then using the fact that $PP^T = P^TP = I$, we can rewrite this equation as $A = PDP^T$. (2)

Now transposing both sides of this equation and using the fact that a diagonal matrix is the same as its transpose we obtain

$A^T = (PDP^T)^T = PD^TP^T = PDP^T = A$

so A must be symmetric if it is orthogonally diagonalizable. The following theorem shows that every orthogonally diagonalizable is symmetric and every symmetric matrix is diagonalizable.

Theorem 4.20 If A is an n $\times$ n matrix with real entries, then the following are equivalent.

(a) A is orthogonally diagonalizable.

(b) A has an orthonormal set of n eigenvectors.

(c) A is symmetric.

Proof. (a) $\Rightarrow$ (b) Since A is orthogonally diagonalizable, there is an orthogonal matrix P such that $\mathrm{P}^{-1}\mathrm{AP}$ is diagonal. As shown in Formula (2) in the proof of Theorem 4.8, the n column vectors of P are eigenvectors of A. Since P is orthogonal, these column vectors are orthonormal, so A has n orthonormal eigenvectors.

(b) $\Rightarrow$ (a) Assume that A has an orthonormal set of n eigenvectors $\{p_1, p_2, \ldots, p_n\}$. As shown in the proof of Theorem 4.8, the matrix P with these eigenvectors as columns diagonalizes A. Since these eigenvectors are orthonormal, P is orthogonal and thus orthogonally diagonalizes A.

(a) $\Rightarrow$ (c) In the proof that (a) $\Rightarrow$ (b) we showed that an orthogonally diagonalizable n $\times$ n matrix A is orthogonally diagonalized by an n $\times$ n matrix P whose columns form an orthonormal set of eigenvectors of A. Let D be the diagonal matrix $\mathrm{D} = \mathrm{P}^T\mathrm{AP}$ from which it follows that $\mathrm{A} = \mathrm{PDP}^T$.

Thus, $\mathrm{A}^T = (\mathrm{PDP}^T)^T = \mathrm{PD}^T\mathrm{P}^T = \mathrm{PDP}^T = \mathrm{A}$

which shows that A is symmetric.

(c) $\Rightarrow$ (a) The proof of this part is beyond the scope of the syllabus.

Theorem 4.21 If A is a symmetric matrix with real entries, then:

(a) The eigenvalues of A are all real numbers.

(b) Eigenvectors from different eigenspaces are orthogonal.

The above theorem yields the following procedure for orthogonally diagonalizing a symmetric matrix.

Orthogonally Diagonalizing an $n \times n$ Symmetric Matrix

Step 1. Find a basis for each eigenspace of A.

Step 2. Apply the Gram–Schmidt process to each of these bases to obtain an orthonormal basis for each eigenspace.

Step 3. Form the matrix P whose columns are the vectors constructed in Step 2. This matrix will orthogonally diagonalize A, and the eigenvalues on the diagonal of $\mathrm{D} = \mathrm{P}^T\mathrm{AP}$ will be in the same order as their corresponding eigenvectors in P.

Example 34. Find an orthogonal matrix P that diagonalizes

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

The characteristic equation of A is det(λI - A)

$$= \det \begin{bmatrix} \lambda - 4 & -2 & -2 \\ -2 & \lambda - 4 & -2 \\ -2 & -2 & \lambda - 4 \end{bmatrix} = (\lambda - 2)^2(\lambda - 8) = 0.$$

Thus, the distinct eigenvalues of A are $\lambda = 2$ and $\lambda = 8$.

It can be shown that $u_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ form a basis for the eigenspace corresponding to $\lambda = 2$.

Applying the Gram–Schmidt process to $\{u_1, u_2\}$ yields the following orthonormal eigenvectors

$$v_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \text{ and } v_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}.$$

Similarly, the eigenspace corresponding to $\lambda = 8$ has

$u_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as a basis. Applying the Gram–Schmidt process to $\{u_3\}$ (i.e., normalizing u_3) yields

$$v_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}.$$

Finally, using v_1, v_2, v_3 as column vectors, we obtain

$$\text{P}=\begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

which orthogonally diagonalizes A. One can check that

$$\text{P}^T\text{AP}= \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{bmatrix}.$$

4.8.1 Spectral Decomposition

If A is a symmetric matrix that is orthogonally diagonalized by $\text{P} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix}$ and if $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A corresponding to the unit eigenvectors $u_1, u_2, \dots, u_n$, then we know that D = P^TAP , where D is a diagonal matrix with the eigenvalues in the diagonal positions. It follows from this that the matrix A can be expressed as

$$A=PDP^T= \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \dots \\ u_n^T \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 u_1 & \lambda_2 u_2 & \dots & \lambda_n u_n \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \dots \\ u_n^T \end{bmatrix}$$

$$= \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T.$$

Hence we obtain the formula $A=\lambda_1 u_1 u_1^T+\lambda_2 u_2 u_2^T+\cdots+\lambda_n u_n u_n^T$, which is called a **spectral decomposition** of A.

Note that each term of the spectral decomposition of A has the form $\lambda u u^T$, where u is a unit eigenvector of A in column form, and λ is an eigenvalue of A corresponding to u. Since u has size n $\times$ 1, it follows that the product uu^T has size n $\times$ n. It can be proved that uu^T is the standard matrix for the orthogonal projection of R^n on the subspace spanned by the vector u. Accepting this to be so, the spectral decomposition of A tells that the image of a vector **x** under multiplication by a symmetric matrix A can be obtained by projecting **x** orthogonally on the lines (one-dimensional subspaces) determined by the eigenvectors of A, then scaling those projections by the eigenvalues, and then adding the scaled projections.

Example 35. Find the spectral decomposition of the matrix

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

The characteristic equation of A is det(λI - A)

$$= \det \begin{bmatrix} \lambda - 1 & -2 \\ -2 & \lambda + 2 \end{bmatrix} = (\lambda + 3)(\lambda - 2) = 0.$$

Hence $\lambda = -3, \lambda = 2$ are the eigen values.

Solving we get the basis for the eigenspace corresponding to the eigenvector $\lambda = -3$ is $\{v_1\} = \left\{ \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\}$ and the eigenspace corresponding to the eigenvector $\lambda = 2$ is $\{v_2\} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$.

Normalizing the vectors v_1 and v_2, we get the unit basis vectors as $u_1 = \dfrac{v_1}{\|v_1\|} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$ and $u_2 = \dfrac{v_2}{\|v_2\|} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$.

Hence the spectral decomposition of A is

$$\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T$$

$$= (-3) \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix} + (2) \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

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