Contents

1 Linear Equations **5**

1.1 Systems of Linear Equations 5

1.1.1 Linear Equations 5

1.1.2 Augmented Matrix 8

1.1.3 Elementary Row Operations 10

1.2 Gaussian Method. 11

1.2.1 Reduced Row Echelon Form 12

1.2.2 Gaussian Elimination 14

1.3 Homogeneous Equations 17

2 Matrices **19**

2.1 Matrix Notation and Terminology 19

2.2 Arithmetic Operations of Matrices 23

2.3 Inverses 31

2.4 Finding A^{-1} via row operations 36

2.4.1 Elementary Matrices 36

2.4.2 Relating Elementary Matrices and Invertible Matrices 38

2.4.3 Finding A^{-1} 39

2.5 Nonhomogeneous Systems 41

3 Determinants **45**

3.1 Determinants 45

3.1.1 Permutation & Inversion 45

3.1.2 Definition of Determinants 47

3.1.3 Useful Results 48

3.2 Co-factor Expansion. 51
3.2.1 Cofactor Expansion 51
3.2.2 Adjoint of **A** 54
3.3 Cramer's Rule 56

4 Vectors **59**
4.1 Vectors in $\mathbb{R}^n$ 59
4.2 Vector Addition & Scalar Multiplication 60
4.3 Scalar Product & Norm 62
4.3.1 Scalar Product 62
4.3.2 Norm 63
4.3.3 Projections 64
4.4 Orthogonality & The Standard Basis 65
4.4.1 Orthogonality 65
4.4.2 The Standard Basis 65

5 More about vectors in 2 and 3 dimensions **67**
5.1 The cross product 67
5.2 Lines and planes 68
5.2.1 Planes 68
5.2.2 Lines 69
5.3 Geometric View of Determinants 70
5.3.1 The volume of a parallelepiped 70

6 Linear Transformations **73**
6.1 Linear Transformations 73
6.1.1 Linear Transformations 73
6.1.2 The Standard Matrix 76
6.2 Composition and Inverse Transformations 79
6.2.1 Composition 79
6.2.2 Identity Transformation 80
6.2.3 Inverse Transformation 81
6.3 Onto and One-to-one Linear Transformations 82

6.3.1 Domain and Range 82
6.3.2 Onto 83
6.3.3 One-to-one 84
6.4 Eigenvalues and eigenvectors 87

7 Vector Spaces **91**
7.1 Real Vector Spaces 91
7.1.1 Definition 91
7.1.2 Vector Space $\mathbb{R}^n$ and Vector Space of Matrices 92
7.1.3 Zero Vector Space 95
7.1.4 Vector Spaces of Functions, Polynomials 96
7.2 Properties 98
7.3 Subspaces 100
7.4 Linear independence, basis and dimension 107
7.4.1 Linear Combinations 107
7.4.2 Span 109
7.4.3 Linear independence 111
7.4.4 Basis 114
7.4.5 Coordinate Vectors 116
7.4.6 Dimension 118

8 Rank and Nullity **121**
8.1 Null space, row space and column space 121
8.1.1 Nullspace 121
8.1.2 Column space and row space 122
8.2 Rank-nullity Formula 125
8.2.1 Rank of a matrix 125
8.2.2 Rank-nullity Formula 126

9 Inner product spaces **129**
9.1 Inner products 129
9.2 Orthogonality, projections and the Gram-Schmidt process 131
9.2.1 Orthogonal vectors 131

9.2.2 The Gram-Schmidt orthonormalization process 135
9.2.3 More on the orthogonal complement 135
9.3 Best approximations 136
9.3.1 Fourier series 137
9.3.2 Least squares solutions 139
9.4 Change of basis 141
9.5 Orthogonal matrices 144
9.6 Diagonalization 145
9.6.1 Digagonalization 145
9.6.2 Orthogonal diagonalization 149

Chapter 1

Linear Equations

Many real-life problems can be formulated in terms of linear equations (or linear inequalities) . Even if they are not linear equation (or inequalities), very often we can make use of some approximated linear equations to obtain a desired solution.

In this chapter, we focus on solving linear equation using Gaussian method. We also discuss briefly the existence of solutions for a given linear system.

1.1 Systems of Linear Equations

Let's start with a single equation $a_1x + a_2y = b$ where a_1, a_2 and b are real constants, and a_1 and a_2 are not both zero. This equation represents a straight line in the xy-plane. It is known as a linear equation in x and y. More generally, we have

1.1.1 Linear Equations

Definition 1.1.1.

A **linear equation** in the n variables $x_1, x_2, \ldots, x_n$ is an equation that can be expressed in the form

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

where $a_1, a_2, \ldots, a_n$ and b are real constants. The variables in a linear equation are sometimes called **unknowns**.

Note that for a linear equation, it does not involve any products or roots of variables. All variables occur only to the first power and do not appear as arguments for trigonometric, logarithmic, or exponential functions.

Some examples of linear equations:

$$\sqrt{2}x_1 - 4x_2 + 9.3x_3 = 100 \quad \& \quad 7u + 8v - 4w + z = 0.$$

The following are **not** linear equations in the variables:

$$7\sqrt{x_1} - x_2 + 5x_3 = 201, \quad u + 16v - 8vw + z = 37, \quad \sin(x_1) + x_2 - x_3 + x_4 = 0.$$

Definition 1.1.2.

(a) A **solution** of a linear equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ is a sequence of numbers $s_1, s_2, \ldots, s_n$ such that the equation is satisfied when we substitute $x_1 = s_1, x_2 = s_2, \cdots, x_n = s_n$.

(b) The set of all solutions of the equation is called its **solution set** or sometimes the **general solution** of the equation.

Consider the linear equation $\sqrt{2}x_1 - 4x_2 + 9.3x_3 = 100$, a solution is $x_1 = 0, x_2 = -25, x_3 = 0$. Another solution is $x_1 = \sqrt{2}, x_2 = -5/4, x_3 = 10$. To obtain its general solution, we may set $x_1 = s$ and $x_3 = t$, where both s and t are arbitrary real numbers. Then the general solution is

$$x_1 = s, x_2 = -\frac{100 - \sqrt{2}s - 9.3t}{4}, x_3 = t, \text{where } s, t \in \mathbb{R}.$$

The arbitrary numbers s and t are called **parameters**.

We have solved some simultaneous equations in our high school mathematics. Now we shall discuss more about simultaneous equations where each equation is linear. Such system of equations is formally defined as follows:

Definition 1.1.3.

(a) A **system of linear equations** or a **linear system** is a finite set of linear equations in the n variables $x_1, x_2, \ldots, x_n$.

(b) A **solution** of a system of linear equations is a sequence of numbers $s_1, s_2, \ldots, s_n$ such that $x_1 = s_1, x_2 = s_2, \cdots, x_n = s_n$ is a solution of every linear equation in the system.

(c) The set of all solutions of the linear system is called its **solution set** or sometimes the **general solution** of the linear system.

Example 1.1.4. The linear system

$$\begin{array}{rcrcrcr} 4x_1 & + & 3x_2 & - & x_3 & = & -1 \\ 2x_1 & + & x_2 & + & 3x_3 & = & 7 \end{array}$$

has a solution $x_1 = 1, x_2 = -1, x_3 = 2$. (CHECK!)

But, $x_1 = 1, x_2 = 1, x_3 = 8$ is not a solution of the linear system. (Why?)

An arbitrary system of m linear equations in n variables $x_1, x_2, \ldots, x_n$ can be expressed as:

$$\begin{array}{rcrcrcccrcl} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\ \cdot\ \ \cdot & & \cdot\ \ \cdot & & \cdot & \cdots & & & \cdot\ \ \cdot & & \cdot\ \ \cdot \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

where a_{ij} $(1 \le i \le m,\ 1 \le j \le n)$ and b are real numbers. (In general, these numbers can be complex number or number $\pmod k$ etc.)

The number a_{ij} is called the **coefficient** of the x_j in the ith linear equation. Note that the double subscripting on the coefficients is a useful device that is used to specify the location of the coefficient in the system.

The next example demonstrates that not all linear systems have solutions.

Example 1.1.5. The linear system

$$\begin{array}{rcrcr} x & + & y & = & 3 \\ 2x & + & 2y & = & 5 \end{array}$$

does not have a solution. Observe that the second linear equation gives $x + y = 2.5$ which contradicts with the first linear equation.

Definition 1.1.6.

(a) A system of linear equations that has at least one solution is called **consistent**.

(b) A system of linear equations that has no solution is said to be **inconsistent**.

Example 1.1.7. Solve the following systems of linear equations.

(a) $$\begin{array}{rcrcr} x & + & y & = & 3 \\ 2x & - & 3y & = & 16 \end{array}$$

The general solution is $x = 5, y = -2$.

(b) $$\begin{array}{rcrcr} x & + & y & = & 3 \\ 9x & + & 9y & = & 27 \end{array}$$

The general solution is $x = s, y = 3 - s$, where s is an arbitrary real number.

(c) $$\begin{array}{rcrcr} x & + & y & = & 3 \\ 2x & + & 2y & = & 5 \end{array}$$

There is no solution. The solution set is an empty set. This linear system is an inconsistent system.

From the above example, we note that for arbitrary linear systems, there are three possibilities, either

- there is no solution, or
- there is exactly one solution, or
- there are infinitely many solutions.

This seemingly basic yet important fact is recorded in the next theorem. (We defer its proof.)

Theorem 1.1.8. *Every system of linear equations has no solutions, or has exactly one solution, or has infinitely many solutions.*

1.1.2 Augmented Matrix

Definition 1.1.9.

For given positive integers m and n, an $m \times n$ **matrix** A is a rectangular array of mn numbers arranged in m horizontal rows and n vertical columns:

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

The (i,j)th entry is the term a_{ij} found in the ith row and jth column. The ith row of A, where $1 \leq i \leq m$, is

$$\begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix},$$

while the jth column, for $1 \leq j \leq n$, is

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Example 1.1.10.

(a) $\begin{bmatrix} 2 & 0 & 2 & 1 \\ 3 & -1 & 4 & 7 \\ 6 & 1 & -1 & 0 \end{bmatrix}$ is a matrix with 3 rows and 4 columns. We say this is a 3×4 matrix.

(b) $\begin{bmatrix} 2 & 0 & 2 & \pi \\ 3 & -1 & 0 & 7 \\ \frac{1}{2} & \sqrt{6} & 4 & 7 \\ 6 & 1 & -1 & 0 \end{bmatrix}$ is a matrix with 4 rows and 4 columns. We say this is a 4×4 matrix.

A general linear system of m equations in n unknowns:

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\ \cdot\ \cdot & & \cdot\ \cdot & & \cdot & \cdots & & \cdot\ \cdot & & & \cdot\ \cdot \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

can be represented by a rectangular array of numbers:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & b_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m \end{bmatrix}$$

This is called the **augmented matrix** for the linear system.

To obtain this abbreviated representation, we drop symbols '+' and '=' (and we write $-3x_j = +(-3)x_j$)), and keep the same order of the unknowns in each linear equation and the constants b_i on the right.

Example 1.1.11.

(a) The augmented matrix of the linear system

$$\begin{array}{rcrcrcr} 2x_1 & & & + & 2x_3 & = & 1 \\ 3x_1 & - & x_2 & + & 4x_3 & = & 7 \\ 6x_1 & + & x_2 & - & x_3 & = & 0 \end{array} \quad \text{is} \quad \begin{bmatrix} 2 & 0 & 2 & 1 \\ 3 & -1 & 4 & 7 \\ 6 & 1 & -1 & 0 \end{bmatrix}$$

(b) The augmented matrix of the linear system

$$\begin{array}{rrrrcr} x_1 & & & & = & 2 \\ & x_2 & & & = & 0 \\ & & x_3 & & = & 0 \\ & & & x_4 & = & 7 \end{array} \quad \text{is} \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

(c) The following augmented matrix

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

corresponds to the linear system

$$\begin{array}{rcrcrcrcr} 3x_1 & & & - & 2x_3 & & & = & 5 \\ & - & x_2 & & & + & 9x_4 & = & 6 \\ -x_1 & & & + & x_3 & + & 2x_4 & = & -7 \end{array}$$

1.1.3 Elementary Row Operations

Our aim is to solve systems of linear equations. The basic method for solving a system of linear equations is to replace the given system by a new system that has the same solution set but is easier to solve. One such technique in solving a linear system is to eliminate unknowns. We shall also look at the corresponding effect on the row of its augmented matrix. As an illustration, to solve

$$\begin{array}{rcrcr} x & + & y & = & 3 \\ 2x & - & 3y & = & 16 \end{array} \qquad \begin{bmatrix} 1 & 1 & 3 \\ 2 & -3 & 16 \end{bmatrix}$$

We may proceed as follows:

Step 1 We eliminate x by multiplying the first equation by -2 and adding it to the second equation.

$$\begin{array}{rcrcr} x & + & y & = & 3 \\ & - & 5y & = & 10 \end{array} \qquad \begin{bmatrix} 1 & 1 & 3 \\ 0 & -5 & 10 \end{bmatrix}$$

This corresponds to 'Add a multiple -2 of first row to second row' for the augmented matrix.

Step 2 Multiply second equation by $-\frac{1}{5}$.

$$\begin{array}{rcrcr} x & + & y & = & 3 \\ & & y & = & -2 \end{array} \qquad \begin{bmatrix} 1 & 1 & 3 \\ 0 & 1 & -2 \end{bmatrix}$$

For the augmented matrix, this corresponds to 'Multiply the second row throughout by $-\frac{1}{5}$.

Step 3 We eliminate y by multiplying the second equation by -1 and adding it to the first equation.

$$\begin{array}{rcrcr} x & & & = & 5 \\ & & y & = & -2 \end{array} \qquad \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -2 \end{bmatrix}$$

This corresponds to 'Add a multiple -1 of the second row to the first second row' for the augmented matrix.

Therefore, we obtain the (only) solution: $x = 5, y = -2$.

Note that we have performed some row operations on the matrices. There are three basic row operations, called **elementary row operations**:

1. Interchange two rows.

2. Multiply a row through by a non-zero constant.

3. Add a multiple of one row to another row.

Remark Elementary row operations are useful in finding solutions of systems of linear equation, distinguish singular matrices from invertible matrices, determine the inverse of an invertible matrix, and finding determinants and in simplex method for solving linear programmes.

1.2 Gaussian Method.

In order to solve a linear system, we shall apply elementary row operations on its corresponding augmented matrix instead of working on the linear systems directly. The immediate question is: What is the structure of the 'expected' augmented matrix we aim for?

1.2.1 Reduced Row Echelon Form

Consider the augmented matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

We can readily read off the solution of the corresponding linear system $x_1 = 2, x_2 = 0, x_3 = 0, x_4 = 7$.

Likewise, with less effort, we can obtain the general solution of the linear system associated with following augmented matrix

$$\begin{bmatrix} 1 & 0 & -3 & 0 & 1 \\ 0 & 1 & 8 & 0 & 5 \\ 0 & 0 & 0 & 1 & -6 \end{bmatrix}$$

as $x_1 = 1 + 3s, x_2 = 5 - 8s, x_3 = s, x_4 = -6$, where s is an arbitrary real number.

The last row of the next augmented matrix

$$\begin{bmatrix} 1 & 0 & -3 & 0 & 1 \\ 0 & 1 & 8 & 0 & 5 \\ 0 & 0 & 0 & 0 & -6 \end{bmatrix}$$

tells us that the corresponding linear system is inconsistent. (WHY?)

The above augmented matrices are in a simple form that allows us to obtain the solution set by inspection. This form is now described in

Definition 1.2.1.

A matrix in **reduced row-echelon form** must have the following properties:

1. If a row does not consist of entirely zeros, then the first non-zero number in the row is a 1. (We call this a leading 1.)
2. If there are any rows that consists 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.

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

Example 1.2.2.

Which of the following matrices are in reduced row-echelon form? row-echelon form?

(a) $\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -1 & 1/2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 1 & 8 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 & 4 & 9 \\ 0 & 1 & 0 & 3 & -2 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 0 & 0 & 9 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 1 & 0 & 0 & 9 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 7 \\ 1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 9 \\ 0 & 1 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Exercise List all types of possible 2×3 matrices that are in reduced row-echelon.

Important Note Now, once we have reduced the augmented matrix of a linear system in reduced row echelon form, it is easy to write down the solution of the linear system. For columns without leading ones, we set the corresponding variables to be free, and then express the other variables in terms of the free variables.

Example 1.2.3. Write down the solution of a linear system of equations whose augmented matrix in reduced row echelon form is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 9 \\ 0 & 1 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

1.2.2 Gaussian Elimination

For a given linear system, our aim is to obtain the reduced row echelon form of the corresponding augmented matrix in a finite number of steps by applying the elementary row operations to eliminate the unknowns systematically.

Step-by-Step Elimination Method:

Step 1. Locate the leftmost column that does not consists 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.

(*To solve the associated linear system of augmented matrix in row-echelon form , we can now perform back-substitution.*)

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. The last matrix is in reduced row-echelon form.

If we use only first five steps, the above procedure produces a row-echelon form and is called **Gaussian Elimination** . Carrying out the procedure through to the sixth step and producing a matrix in reduced row-echelon form is called **Gauss-Jordan elimination**.

Example 1.2.4. We shall solve the following linear system by Gaussian Elimination.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 8 \\ -x_1 & - & 2x_2 & + & 3x_3 & = & 1 \\ 3x_1 & - & 7x_2 & + & 4x_3 & = & 10 \end{array}$$

Solution The augmented matrix is:

$$\begin{bmatrix} 1 & 1 & 2 & 8 \\ -1 & -2 & 3 & 1 \\ 3 & -7 & 4 & 10 \end{bmatrix} \overset{R2+R1, R3+(-3)R1}{\rightarrow} \begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & -1 & 5 & 9 \\ 0 & -10 & -2 & -14 \end{bmatrix} \overset{(-1)R2}{\rightarrow} \begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & -10 & -2 & -14 \end{bmatrix}$$

$$\overset{R3+(10)R2}{\rightarrow} \begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & 0 & -52 & -104 \end{bmatrix} \overset{(-1/52)R3}{\rightarrow} \begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

The matrix is in row-echelon form,and it corresponds to the following linear system:

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 8 \\ & & x_2 & - & 5x_3 & = & -9 \\ & & & & x_3 & = & 2 \end{array}$$

By back-substituting $x_3 = 2$ into the second equation gives $x_2 = -9 + 5(2) = 1$.

Back-substituting $x_2 = 1$ and $x_3 = 2$ into the first equation gives $x_1 = 8 - (1) - 2(2) = 3$.

The linear system has a unique solution,$x_1 = 3, x_2 = 1, x_3 = 2$.

Now, we demonstrate the Gauss-Jordan Elimination to obtain the reduced row-echelon form.

Example 1.2.5. We continue from last example.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & 2x_3 & = & 8 \\ -x_1 & - & 2x_2 & + & 3x_3 & = & 1 \\ 3x_1 & - & 7x_2 & + & 4x_3 & = & 10 \end{array}$$

Solution. To obtain the reduced row-echelon form, it remains to make each entry above the leading 1 in each column containing the leading 1 zero.

$$\begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & 0 & 1 & 2 \end{bmatrix} \overset{R1+(-1)R2}{\rightarrow} \begin{bmatrix} 1 & 0 & 7 & 17 \\ 0 & 1 & -5 & -9 \\ 0 & 0 & 1 & 2 \end{bmatrix} \overset{R1+(-7)R3, R2+5R3}{\rightarrow} \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

The last matrix is in reduced row-echelon form, and the corresponding solution is $x_1 = 3, x_2 = 1, x_3 = 2$.

Example 1.2.6. Solve the linear system by Gauss-Jordan Elimination.

$$\begin{array}{rcrcrcr} & - & 2b & + & 3c & = & 1 \\ 3a & + & 6b & - & 3c & = & -9 \\ a & + & 6b & + & 4c & = & 17 \end{array}$$

Solution. (Exercise.)

Once you have mastered the systematic methods of Gaussian Elimination and Gauss-Jordan Elimination, you may vary the steps involved in specific problems to avoid fractions or to minimize computation.

In applications we usually encounter large linear systems that must be solved by computer. Most computer algorithms for solving linear systems are based on Gaussian elimination or Gauss-Jordan elimination, but the basic procedures are often modified to deal with such issues as

- Reducing roundoff errors
- Minimizing the use of computer memory space
- Solving the system with maximum speed

The next example is an application to find the coefficients of a cubic function when it is known that its graph passes through some points.

Example 1.2.7. The cubic curve $y = ax^3 + bx^2 + cx + d$ passes through $(0, 10)$, $(1, 7)$, $(3, -11)$ and $(4, -14)$. Find the coefficients a, b, c and d.

[Solution] (Exercise)

1.3 Homogeneous 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:

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & 0 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdot & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

All such systems have $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ as a solution. This solution is called the **trivial solution**. So, every homogeneous system of linear equations is consistent. There are two possibilities for its solutions:

- The system has only the trivial solution. (Unique solution)
- The system has infinitely many solutions in addition to the trivial solution. These other solutions are called **nontrivial solutions**.

Example 1.3.1. For the following homogeneous linear system

$$\begin{array}{rcrcrcl} x_1 & + & x_2 & + & 2x_3 & = & 0 \\ -x_1 & - & 2x_2 & + & 3x_3 & = & 0 \\ 3x_1 & - & 7x_2 & + & 4x_3 & = & 0 \end{array}$$

which has the same coefficients as Example 1.2.5. Its augmented matrix is

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

By Gauss-Jordan Elimination , we obtain the reduced row-echelon form

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Thus, the homogeneous system has only trivial solution $x_1 = 0, x_2 = 0, x_3 = 0$.

Since the last column of the augmented matrix of a homogeneous system remains zero when we apply elementary row operations, we could omit the last column of zeros.

Example 1.3.2. Solve the following homogeneous linear system.

$$\begin{array}{rcrcrcrcr} & & v & + & 3w & - & 2x & = & 0 \\ 2u & + & v & - & 4w & + & 3x & = & 0 \\ 2u & + & 3v & + & 2w & - & x & = & 0 \\ -4u & - & 3v & + & 5w & - & 4x & = & 0 \end{array}$$

[Solution]

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

$$\begin{array}{c} \overset{R2+(-1)R1}{\longrightarrow} \\ \overset{R3+(-3)R1}{\longrightarrow} \\ \overset{R4+(3)R1}{\longrightarrow} \end{array} \begin{bmatrix} 0 & 1 & 3 & -2 \\ 2 & 0 & -7 & 5 \\ 2 & 0 & -7 & 5 \\ -4 & 0 & 14 & -10 \end{bmatrix}$$

$$\begin{array}{c} \overset{R3+(-1)R2}{\longrightarrow} \\ \overset{R4+(2)R2}{\longrightarrow} \end{array} \begin{bmatrix} 0 & 1 & 3 & -2 \\ 2 & 0 & -7 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\overset{R1\leftrightarrow R2}{\longrightarrow} \begin{bmatrix} 2 & 0 & -7 & 5 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\overset{\frac{1}{2}R1}{\longrightarrow} \begin{bmatrix} 1 & 0 & -\frac{7}{2} & \frac{5}{2} \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Referring to the matrix in reduced row echelon form, for columns without leading ones, we set the corresponding variables to be free variables.

Thus, we have $w = s$ and $x = t$, where $s, t \in \mathbb{R}$. We express the remaining variables in terms of w and x:

$$u = \frac{7}{2}w - \frac{5}{2}x = \frac{7}{2}s - \frac{5}{2}t, \text{ and}$$
$$v = -3w + 2x = -3s + 2t$$

Chapter 2

Matrices

We have encountered matrices as augmented matrices in last section . A matrix is simply a rectangular array of numbers. Rectangular arrays of real numbers arise in many contexts other than as augmented matrices for linear systems of equations. Matrices are also applied to study another simple cryptosystem – the Hill cipher. Like numbers (or integers modulo m, refer to Chapter 1), we will study arithmetic operations of addition, subtraction and multiplication.

2.1 Matrix Notation and Terminology

For given positive integers m and n, an $m \times n$ **matrix** A is a rectangular array of mn numbers arranged in m horizontal rows and n vertical columns:

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

The (i, j)th entry (or simply ij-entry) is the term a_{ij} found in the ith row and jth column. The ith row of A, where $1 \leq i \leq m$, is

$$\begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix},$$

while the jth column, for $1 \leq j \leq n$, is

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Notation and Terminology

1. Capital letters $A, B, C, \ldots$ are used to denote matrices, and lowercase letters to denote numerical quantities. Some examples:

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 9 & 6 & 3 \end{bmatrix}, T = \begin{bmatrix} a & b & c \\ d & e & f \\ x & y & z \end{bmatrix}, Z = \begin{bmatrix} z_1 & z_2 & z_3 & z_4 \\ y_1 & y_2 & y_3 & y_4 \end{bmatrix}$$

2. The number m of rows and the number n of columns describe the **size** of a matrix. We write it as $m \times n$.

3. The matrix

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

is sometimes written as

$$[a_{ij}]_{m\times n} \text{ or simply } [a_{ij}].$$

To refer to the (i, j)th-entry of the matrix A, we use the notation $(A)_{ij}$. So, $(A)_{ij} = a_{ij}$.

For $A = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 9 & 6 & 3 & 0 \\ -2 & -4 & -8 & \sqrt{2} \end{bmatrix}$, we have $(A)_{11} = \quad , (A)_{21} = \quad$ and $(A)_{34} = \quad$.

4. Usually, we match the letter denoting a matrix with the letter denoting its entries. For a matrix B, its ij-entry is b_{ij}.

5. When $m = 1$, the matrix has only one row,i.e.

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_n \end{bmatrix}$$

We call this matrix a **row matrix** (also known as **row vector**) .

6. When $n = 1$, the matrix has only one column,i.e.

$$B = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

We call such matrix a **column matrix** (also known as **column vector**) .

Of course, for a row matrix, we may simplify the entry using only one index, i.e. $C = [c_1 \quad c_2 \quad \cdots \quad c_n]$.

7. The $m \times n$ matrix with zeros as its entries is called the **zero matrix** and we denote it by 0.

8. When $m = n$, we call A a **square matrix** of size n. (The rectangular array now looks like a square.)

   For an $n \times n$ square matrix

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

   The entries $a_{11}, a_{22}, a_{33}, \ldots, a_{nn}$ are called the **diagonal entries**. They are on the main diagonal of A.

9. The $n \times n$ square matrix where all the entries along the diagonal from the top left to the bottom right are 1, and 0 elsewhere, is called the **identity matrix**. It is often denoted as I_n. E.g.,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

10. The $n \times n$ square matrix where all the off diagonal entries (i.e. entries below and above the main diagonal) are 0 is called an **diagonal matrix**.

$$A = \begin{pmatrix} a & 0 & 0 \\ 0 & e & 0 \\ 0 & 0 & g \end{pmatrix} \quad B = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi \end{pmatrix}.$$

11. The $n \times n$ square matrix where all the entries below the main diagonal are 0 is called an **upper triangular matrix**.

$$A = \begin{pmatrix} a & b & c \\ 0 & e & f \\ 0 & 0 & g \end{pmatrix} \quad B = \begin{pmatrix} 1 & 3 & 5 & 0 \\ 0 & 3 & -2 & 9 \\ 0 & 0 & 0 & \pi \\ 0 & 0 & 0 & 0.8 \end{pmatrix}.$$

   In a similar way, a **lower triangular matrix** is a square matrix where all the entries above the main diagonal are 0. Eg.

$$C = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix} \quad B = \begin{pmatrix} 53 & 0 & 0 & 0 & 0 \\ -4 & 3 & 0 & 0 & 0 \\ 1/2 & 0 & \pi & 0 & 0 \\ 0 & -1 & 0.4 & \sqrt{2} & 0 \\ 1 & 3 & 5 & 7 & 9 \end{pmatrix}.$$

12. Lastly, as in the discussion of vectors, the term **scalars** refer to numerical quantities in discussing matrices (and vectors). Most of what follows could be applied to matrices whose entries are elements from $\mathbb{Z}_m$, but within the arithmetic in $\mathbb{Z}_m$.

Definition 2.1.1. Two matrices are defined to be equal if they have the same size and their corresponding entries are equal.

In matrix notation, let $A = [a_{ij}]$ and $B = [b_{ij}]$. Matrices A and B are equal if they have the same size (same number of rows and same number of columns) and $a_{ij} = b_{ij}$ for all i and j. (Here, note that if the size of both matrix is $m \times n$, then $1 \leq i \leq m$ and $1 \leq j \leq n$.)

Example 2.1.2. Solve the following matrix equation for a, b, c and d.

$$\begin{bmatrix} a-b & b+c \\ 3d+c & 2a-4d \end{bmatrix} = \begin{bmatrix} 8 & 1 \\ 7 & 6 \end{bmatrix}$$

Example 2.1.3.

(a) Let A be a 3×4 matrix whose (i, j)-th entry is defined by $(A)_{ij} = (-1)^{i+j} 2i + j$.

Write A.

(b) Let $B = [b_{ij}]$ be a 3×3 matrix where $b_{ij} = \begin{cases} i+j & \text{if } i > j \\ 0 & \text{if } i = j \\ -j & \text{if } i < j \end{cases}$

Write B.

2.2 Arithmetic Operations of Matrices

We will study the arithmetic operations for matrices.

Definition 2.2.1. If A and B are two $m \times n$ matrices, then

(a) the **sum** $A + B$ is the matrix obtained by adding entries in the same positions, and

(b) the **difference** $A - B$ is the matrix obtained by subtracting entries of B from the corresponding entries of B

NOTE that matrices of different sizes cannot be added or subtracted.

Example 2.2.2.

Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$. Then

$$A + B = \begin{pmatrix} 1+7 & 2+8 & 3+9 \\ 4+10 & 5+11 & 6+12 \end{pmatrix} = \begin{pmatrix} 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix}$$

$$A - B = \begin{pmatrix} 1-7 & 2-8 & 3-9 \\ 4-10 & 5-11 & 6-12 \end{pmatrix} = \begin{pmatrix} -6 & -6 & -6 \\ -6 & -6 & -6 \end{pmatrix}.$$

In matrix notation, if $A = [a_{ij}]$ and $B = [b_{ij}]$ have the same size, then

$$(A + B)_{ij} = (A)_{ij} + (B)_{ij} = a_{ij} + b_{ij}$$

and

$$(A - B)_{ij} = (A)_{ij} - (B)_{ij} = a_{ij} - b_{ij}$$

Definition 2.2.3. If α is a number and A is an $m \times n$ matrix, then the **scalar multiple** αA is the $m \times n$ matrix obtained by multiplying each entry of A by α.

Example 2.2.4.

Let $A = \begin{pmatrix} 1 & 3 & 5 \\ 7 & 9 & 11 \end{pmatrix}$. Then

$$2A = \begin{pmatrix} 2 & 6 & 10 \\ 14 & 18 & 22 \end{pmatrix},$$

$$-3A = \begin{pmatrix} -3 & -9 & -15 \\ -21 & -27 & -33 \end{pmatrix},$$

and

$$\frac{1}{3}A = \begin{pmatrix} \frac{1}{3} & 1 & \frac{5}{3} \\ \frac{7}{3} & 3 & \frac{11}{3} \end{pmatrix},$$

Using matrix notation, if $A = [a_{ij}]$ and α is a scalar (i.e. α is some number), then the (ij)-entry of αA is

$$(\alpha A)_{ij} = \alpha(A)_{ij} = \alpha a_{ij}.$$

Note that for $m \times n$ matrices A and B , the difference $A - B = A + (-1)B$. It is common practice to denote $(-1)B$ by $-B$.

Example 2.2.5. Let $A = \begin{pmatrix} 4 & 3 & 2 \\ 1 & 3 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 7 & 2 & 0 \\ -5 & 3 & -1 \end{pmatrix}$ and $C = \begin{pmatrix} 3 & -6 & 9 \\ 3 & 0 & 12 \end{pmatrix}$.

Evaluate $2A - B + \frac{1}{3}C$.

For matrices, when we perform arithmetic operations like sum, difference and scalar multiplication, we are basically performing similar arithmetic operations sum, difference and multiplication on numbers on 'entry'-level. Thus, we would expect properties such as associativity, commutativity and distributive to hold for such matrix operations.

Theorem 2.2.6. *Assuming the sizes of matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid.*

1. $A + B = B + A$ *(Commutative Law for addition)*
2. $(A + B) + C = A + (B + C)$ *(Associative Law for addition)*
3. $A + 0 = 0 + A = A$ *(Additive Identity)*
4. $A + (-A) = 0$ *(Additive Inverse)*
5. $0 - A = -A$
6. $\alpha(A \pm B) = \alpha A \pm \alpha B$
7. $(\alpha \pm \beta)A = \alpha A \pm \beta A$
8. $\alpha(\beta A) = (\alpha\beta)A$

Can we multiply two matrices in the similar way i.e. by multiplying corresponding entries like what we have done for addition, subtraction or scalar multiplication?

It turns out that such a definition is not very helpful for most problems. Experience has led mathematicians to the following more useful definition of matrix multiplication.

Definition 2.2.7. If A is an $m \times r$ matrix and B is an $r \times n$ matrix, then the **matrix product** (or simply the **product**) AB is the $m \times n$ matrix whose (i, j)th entry is determined by

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj} = \sum_{k=1}^{r} a_{ik}b_{kj}.$$

(This is obtained by taking the ith row of A and the jth column of B, multiplying their corresponding terms together – i.e., first term in that row with the first term in that column, etc. – and then adding up the products.)

$$AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ir} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mr} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots & & \\ b_{r1} & b_{r2} & \cdots & b_{rj} & \cdots & b_{rn} \end{pmatrix}.$$

Example 2.2.8.

(a) Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} -2 \\ 4 \\ 2 \end{pmatrix}$. Note that A is 2×3 matrix and B is 3×1, the product AB will be a 2×1 matrix.

To find the entries of AB:

$(1, 1)$-th entry:

$(2, 1)$-th entry:

$(3, 1)$-th entry:

Thus, we have $AB = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$.

(b) Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{pmatrix}$ and $C = \begin{pmatrix} -3 \\ 0 \\ 2 \\ \sqrt{3} \end{pmatrix}$. Note that A is 2×3 matrix and C is 4×1, the product AC is not defined.

(c) Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{pmatrix}$ and $D = \begin{pmatrix} -2 & 0 \\ 4 & 5 \\ 2 & 1 \end{pmatrix}$.

Find AD.

Example 2.2.9. Let $A = \begin{pmatrix} 4 & 3 & 2 \\ 1 & 3 & 1 \end{pmatrix}$, $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Then $AI_3 =$.

Example 2.2.10.

Let $A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{pmatrix}$ and $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Then $AX =$

In view of this, the following linear system of equations

$$\begin{array}{rcrcrcr} x & + & 2y & - & z & = & 9 \\ 3x & + & y & + & 4z & = & -5 \end{array}$$

can be expressed as a matrix equation

$$AX = \mathbf{b},$$

where $A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 1 & 4 \end{pmatrix}$, $X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ & $\mathbf{b} = \begin{pmatrix} 9 \\ -5 \end{pmatrix}$

The general system of linear equations

$$\begin{array}{rcrcrcccrcl} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3 \\ . & . & . & . & . & \cdots & . & . & . & . & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

is equivalent to the matrix equation $AX = \mathbf{b}$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Proposition 2.2.11. *(a) Let A be an $m \times r$ matrix and B be a column matrix, say $r \times 1$. Then the matrix product AB is an $m \times 1$ column matrix, and*

$$AB = b_1A_1 + b_2A_2 + \cdots + b_rAr,$$

where $(B)_{k1} = b_k$ and A_j is the jth column of A.

$$(AB = \begin{pmatrix} A_1 & A_2 & \cdots & A_r \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_r \end{pmatrix} = b_1A_1 + b_2A_2 + \cdots + b_rAr.)$$

(b) Let A be an $1 \times r$ row matrix and B be an $r \times n$ matrix. Then the matrix product AB is a $1 \times n$ row matrix and

$$AB = a_1B_1 + a_2B_2 + \cdots + a_rBr,$$

where $(A)_{1k} = a_k$ and B_i is the ith row of B.

$$(AB = \begin{pmatrix} a_1 & a_2 & \cdots & a_r \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_r \end{pmatrix} = a_1B_1 + a_2B_2 + \cdots + a_rBr.)$$

Proof.

Without computing the entire product, we may compute a particular row or column of a matrix product AB as follows:

Proposition 2.2.12. *Let A and B be $m \times r$ and $r \times n$ matrices respectively.*

Then

(a) the jth column of AB is the matrix product of A and the jth column of B;

$$j\text{th column of } AB = A[j\text{th column of } B].$$

(b) the ith row of AB is the matrix product of the ith row of A and the matrix B;

$$i\text{th row of } AB = [i\text{th row of } A]B.$$

Example 2.2.13. Let $A = \begin{pmatrix} 4 & 3 & 2 \\ 1 & 3 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \\ 0 & -3 & -2 \end{pmatrix}$.

Find (a) the 2nd column of AB and (b) the last row of AB.

Recall that the identity matrix I_n is the square matrix I_n of size n with (i, j)th-entry

$$(I_n)_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j \end{cases}$$

Thus,

$$I_1 = \begin{pmatrix} 1 \end{pmatrix}, I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, I_4 =$$

The next result shows that the role of identity matrices in matrix multiplication is like the number 1 in usual multiplication.

Proposition 2.2.14. *Let $A = [a_{ij}]$ be an $m \times n$ matrix. Then $AI_n = A$ and $I_m A = A$.*

[Proof.] (Exercise) You may look at the jth column of AI_n to prove $AI_n = A$.

For $I_m A = A$, what is the ith-row of of $I_m A$?

Theorem 2.2.15. *Assuming the sizes of matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid.*

1. $(AB)C = A(BC)$ *(Associative Law for multiplication);*
2. $AI_n = A$ *and* $I_m A = A$ *if* A *is* $m \times n$.
3. $A(B + C) = AB + AC$ *(Left distributive law)*
4. $(A + B)C = AC + BC$; *(Right distributive law)*
5. $A0 = 0, 0A = 0$

Remark

1. For matrix multiplication, the commutativity law does not hold. (Can you find an example to illustrate this?
2. It is not true that if $AB = 0$, then either $A = 0$ or $B = 0$. (Exercise: Find two non-zero 2×2 matrices A and B with $AB = 0$.

2.3 Inverses

In general, it is not true that for matrices A, B and C, $AB = AC$ implies $B = C$. (We can't cancel A from the matrix equation $AB = AC$.)

Example 2.3.1. Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 4 & 3 \end{pmatrix}$ and $C = \begin{pmatrix} 5 & 2 \\ 4 & 3 \end{pmatrix}$. Then

$$AB = AC = \begin{pmatrix} 4 & 3 \\ 8 & 6 \end{pmatrix} \text{ but } A \neq B.$$

We can 'cancel' A from the matrix equation $AB = AC$ provided A has 'inverse'.

Definition 2.3.2. Let A be a $n \times n$ square matrix. If there is another square matrix B such that $AB = I_n$ and $BA = I_n$, then A is said to be **invertible**, and B is called an **inverse** of A. If no such matrix B can be found, then A is said to be **singular**.

Example 2.3.3. Verify that $A = \begin{pmatrix} -1 & -2 \\ 3 & 5 \end{pmatrix}$ is invertible with $B = \begin{pmatrix} 5 & 2 \\ -3 & -1 \end{pmatrix}$

Example 2.3.4.

Consider the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where $ad - bc \neq 0$. Let the matrix B be defined by

$$B = (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

(a) Verify $AB = I$ and $BA = I$.

(b) Is A invertible?

Can an invertible matrix have more than one inverse?

Proposition 2.3.5. *If B and $\hat{B}$ are both inverses of A, then $B = \hat{B}$.*

(*Proof.* Tutorial.)

The above proposition says that the inverse of an invertible matrix A is unique. In view of this, we shall denote the inverse of A by A^{-1}. Thus, we have

$$AA^{-1} = I \text{ and } A^1 A = I$$

From Example 2.3.4, we have seen a simple formula for the inverse of 2×2 invertible matrix.

Proposition 2.3.6. *The 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. In this case the inverse A^{-1} is given by the formula*

$$A^{-1} = (ad - bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{2.1}$$

[Proof.] It remains to prove that if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible, then $ad - bc \neq 0$.

(Exercise.)

Example 2.3.7. Find the inverse of

(a) $A = \begin{pmatrix} 5 & 3 \\ 7 & 9 \end{pmatrix}$ (b) $B = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

For invertible matrices of other sizes, we shall use Gaussian method to find their inverses. This will be dealt with in the next section.

Now, we study more properties of invertible matrices.

Proposition 2.3.8. *Let A be an invertible matrix. Then we have*

1. $AB = AC \Longrightarrow B = C$ *and*
2. $BA = CA \Longrightarrow B = C$.

Proof.

Proposition 2.3.9. *Let A and B be invertible matrices. Then we have*

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

Proof.

Definition 2.3.10. Let A be a square matrix. We define the nonnegative integer powers of A to be

$$A^0 = I \quad A^n = \underbrace{AA\cdots A}(n>0)$$

Moreover, if A is invertible, then we define the negative integer powers to be

$$A^{-n} = \left(A^{-1}\right)^n = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}(n>0)$$

Parallel to real numbers, we have

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

Example 2.3.11. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Find $A^2, A^3, A^4, A^{-2}, A^{-3}$.

We also have the following laws of exponents.

Proposition 2.3.12. *Let A be an invertible matrix. Then we have*

1. *A^{-1} is invertible and $(A^{-1})^{-1} = A$.*
2. *A^n is invertible and $(A^n)^{-1} = (A^{-1})^n$.*
3. *For any nonzero scalar α, the matrix αA is invertible and $(\alpha A)^{-1} = \frac{1}{\alpha}A^{-1}$.*

Proof. (Exc. some to be discussed during lecture.)

Definition 2.3.13. For an $m \times n$ matrix A, the matrix A^T obtained by interchanging the rows and columns of A is called the **transpose** of A. Thus, the (i,j)th-entry of A^T is

$$(A^T)_{ij} = (A)_{ji}$$

Example 2.3.14. Let $A = \begin{pmatrix} 4 & 3 & 2 \\ 1 & 3 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \\ 0 & -3 & -2 \end{pmatrix}$.

Then $A^T =$ and $B^T =$.

Proposition 2.3.15. *If the sizes of the matrices are such that the stated operations can be performed, then*

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

(b) $(A \pm B)^T = A^T \pm B^T$

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

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

(e) A^T *is invertible if* A *is invertible. In this case,* $(A^T)^{-1} = (A^{-1})^T$

Proof. (Tutorial)

2.4 Finding A^{-1} via row operations

2.4.1 Elementary Matrices

In this section, we shall see that performing an elementary row operation on a matrix is the same as multiplying the matrix by another matrix namely an elementary matrix.

Definition 2.4.1. An $n \times n$ matrix is called an **elementary matrix** if it can be obtained from the $n \times n$ identity matrix I_n by performing a single elementary row operation.

Example 2.4.2.

1. Interchange of Row i and Row j:

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

2. Multiply Row i through by a non-zero constant α.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

3. Add a multiple α of Row j to Row i.

$$\begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Example 2.4.3. Which of the following are elementary matrices? If it is an elementary matrix, what is the corresponding elementary row operation?

$$(a)\begin{pmatrix} -5 & 0 \\ 0 & 1 \end{pmatrix} \quad (b)\begin{pmatrix} 0 & 1 \\ 1 & \sqrt{3} \end{pmatrix} \quad (c)\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad (d)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 9 \\ 0 & 0 & 1 \end{pmatrix} \quad (e)\begin{pmatrix} 5 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Proposition 2.4.4. *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.*

Proof (Exercise.)

Example 2.4.5. Let E be the elementary matrix obtained by multiplying Row 1 by α and A be a general 2×2 matrix. Thus, we have

$$E = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} \quad \& \quad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then

$$EA = \begin{pmatrix} \alpha a & \alpha b \\ c & d \end{pmatrix}$$

If E is an elementary matrix obtained from an elementary row operation, then E can be restored to the identity matrix by applying the inverse row operation.

Row operation on I that produces E	Row operation on E that produces I
Interchange Row i and row j	Interchange Row j and row i
Multiply row i by $\alpha \neq 0$	Multiply row i by $1/\alpha$
Add α times row i to row j	Add $-\alpha$ times row i to row j

Example 2.4.6. Let E be the elementary matrix obtained by adding to Row 1 α times of Row 2.

$$I \to E = \begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to I$$

In view of the above, we conclude that

Proposition 2.4.7. *Every elementary matrix is invertible, and the inverse is also an elementary matrix.*

2.4.2 Relating Elementary Matrices and Invertible Matrices

Theorem 2.4.8. *If A is an $n \times n$ matrix, then the following statements are equivalent, that is all true or all false.*

(a) A is invertible.

(b) $A\mathbf{x} = 0$ 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. The proof is established via the chain of implications:

$$(a) \Longrightarrow (b) \Longrightarrow (c) \Longrightarrow (d) \Longrightarrow (a)$$

• Proof of $(a) \Longrightarrow (b)$:

• Proof of $(b) \Longrightarrow (c)$: Assume the linear system $A\mathbf{x} = 0$ has only trivial solution, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$.

Using the Gauss-Jordan elimination, the augmented matrix of the linear system $A\mathbf{x} = 0$ can be reduced to the following reduced row echelon form:

$$\begin{array}{cccccc} x_1 & & & & = & 0 \\ & x_2 & & & = & 0 \\ & & \cdots & & = & 0 \\ & & & x_n & = & 0 \end{array}$$

i.e. A can be reduced to I_n by a sequence of elementary row matrices.

• Proof of $(c) \Longrightarrow (d)$:

Assume that the reduced row echelon form of A is I_n. Thus, we may perform a finite number of elementary row operations on A to get I_n. Since each elementary row operation performed on a matrix is also accomplished by pre-multiplying the matrix by an appropriate elementary matrix. We can find elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$E_k \cdots E_2 E_1 A = I_n$$

It follows from the invertibility of each elementary matrix that

$$A = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} I_n = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1}$$

So, A is expressible as a product of elementary matrices.

• Proof of $(d) \Longrightarrow (a)$:

Assume A is expressible as a product of elementary matrices. Since elementary matrices are invertible, the product of invertible matrices is invertible. **QED**

If a matrix B can be obtained from a matrix A by performing a finite sequence of elementary row operations, then we can get from B back to A by performing the inverses of these elementary row operations in reverse order. Matrices that can be obtained from one another by a finite sequence of elementary row operations are said to be **row equivalent**.

By Theorem 2.4.8, we have

Proposition 2.4.9. *An $n \times n$ matrix A is invertible if and only if it is row equivalent to the $n \times n$ identity matrix.*

2.4.3 Finding A^{-1}

In view of the following

$$E_k \cdots E_2 E_1 A = I_n,$$

we have $A^{-1} = E_k \cdots E_2 E_1$. Thus, we may determine the inverse of an invertible matrix by multiplying all the elementary matrices $E_k \cdots E_2 E_1$. Recall that pre-multiplying a matrix by an elementary matrix is the same as performing elementary row operation on that matrix. From $A^{-1} = E_k \cdots E_2 E_1 I_n$, we may obtain A^{-1} by performing the same sequence of elementary row operations, that reduces A to I_n, to I_n to obtain A^{-1}

To carry out the procedure to find the inverse of an invertible matrix A, we adjoin the identity matrix to the right side of A, and apply elementary row operations until the left side matrix is reduced to I.

$$[A|I] \;\; \text{apply elementary row operations} \;\; [I|A^{-1}]$$

This procedure is carried out for a 3×3 matrix illustrated below.

Example 2.4.10. Find the inverse of the matrix $\begin{pmatrix} 1 & 0 & 8 \\ 2 & 5 & 3 \\ 1 & 2 & 3 \end{pmatrix}$

[Solution]

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 8 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 0 & 1 \end{array}\right] \begin{array}{c} \overset{R2-2R1}{\longrightarrow} \\ \overset{R3-R1}{\longrightarrow} \end{array} \left[\begin{array}{ccc|ccc} 1 & 0 & 8 & 1 & 0 & 0 \\ 0 & 5 & -13 & -2 & 1 & 0 \\ 0 & 2 & -5 & -1 & 0 & 1 \end{array}\right] \overset{R2-2R3}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 8 & 1 & 0 & 0 \\ 0 & 1 & -3 & 0 & 1 & -2 \\ 0 & 2 & -5 & -1 & 0 & 1 \end{array}\right]$$

$$\overset{R3-2R2}{\longrightarrow} \left[\begin{array}{ccc|ccc} 1 & 0 & 8 & 1 & 0 & 0 \\ 0 & 1 & -3 & 0 & 1 & -2 \\ 0 & 0 & 1 & -1 & -2 & 5 \end{array}\right] \begin{array}{c} \overset{R1-8R3}{\longrightarrow} \\ \overset{R2+3R3}{\longrightarrow} \end{array} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 9 & 16 & -40 \\ 0 & 1 & 0 & -3 & -5 & 13 \\ 0 & 0 & 1 & -1 & -2 & 5 \end{array}\right]$$

Example 2.4.11. Is the matrix $\begin{pmatrix} 6 & 4 & 1 \\ 4 & -1 & 2 \\ 2 & 5 & -1 \end{pmatrix}$ invertible?

Example 2.4.12. Find the inverse of the following diagonal matrix, if $abcd \neq 0$.

$$\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{pmatrix}$$

Recall from the definition of invertible matrices, in order to show a matrix A is invertible, we have to find a matrix B such that

$$AB = I \text{ and } BA = I.$$

The next result shows that if we have a matrix B such that $AB = I$, then the other condition $BA = I$ holds automatically.

Proposition 2.4.13. *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 of (a). Firstly, We show that A is invertible indirectly as follows.

We show that $A\mathbf{x} = \mathbf{0}$ has only trivial solution. Suppose $\mathbf{x}^*$ is a solution. Then

$$A\mathbf{x}^* = \mathbf{0} \Longrightarrow B(A\mathbf{x}^*) = B\mathbf{0} \Longrightarrow (BA)\mathbf{x}^* = B\mathbf{0} \Longrightarrow \mathbf{x}^* = \mathbf{0}$$

By Theorem 2.4.8, A is invertible. Thus A^{-1} exists and $AA^{-1} = I$.

Multiply by B, we obtain $B(AA^{-1}) = BI$, which leads to $(BA)A^{-1} = B$. Hence $A^{-1} = B$, since $BA = I$.

Proof of (b). Similar (Exercise). **QED**

In order to verify that a matrix B is the inverse of A, it suffices to verify either $AB = I$ or $BA = I$.

2.5 Nonhomogeneous Systems

Definition 2.5.1. The linear system $A\mathbf{x} = \mathbf{b}$ is said to be a **nonhomogeneous system** if $\mathbf{b} \neq \mathbf{0}$.

Now, we establish the fundamental result in linear systems.

Theorem 2.5.2. *Every system of linear equations has no solutions, or has exactly one solution, or has infinitely many solutions.*

Proof. Let $A\mathbf{x} = \mathbf{b}$ be a linear system. There are three cases: (a) the system has no solution, (b) the system has exactly one solution, or (c) the system has infinitely many solutions.

It suffices to prove that the system has infinitely many solution, if there is more than one solution.

Suppose $A\mathbf{x} = \mathbf{b}$ has more than one solutions. Let $\mathbf{x_1}$ and $\mathbf{x_2}$ be two distinct solutions. Consider the matrix $\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{x}_2 \neq 0$.

Claim 1: $\mathbf{x}_0$ is a solution of the homogenous system $A\mathbf{x} = \mathbf{0}$.

Claim 2: For any scalar $k \in \mathbb{R}$, $\mathbf{x}_1 + k\mathbf{x}_0$ is a solution of the homogenous system $A\mathbf{x} = \mathbf{b}$. (Exercise)

Since $\mathbf{x}_0$ is nonzero and there are many choices for k, where $k \in \mathbb{R}$, the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions. **QED**

The next result provides another way of solving certain linear systems.

Proposition 2.5.3. *If A is invertible $n \times n$ matrix, then for each $n \times 1$ matrix $\mathbf{b}$, the system of equations $A\mathbf{x} = \mathbf{b}$ has exactly one solution, namely $\mathbf{x} = A^{-1}\mathbf{b}$.*

Proof

- Existence of solution

  Let $\mathbf{x}_0 = A^{-1}\mathbf{b}$. Then we have

$$A\mathbf{x}_0 = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = \mathbf{b}.$$

Thus, $\mathbf{x}_0$ is a solution of $A\mathbf{x} = \mathbf{b}$.

- Uniqueness of solution

  Suppose the system $A\mathbf{x} = \mathbf{b}$ has another solution $\mathbf{x}_1$.

  Then we have $A\mathbf{x}_1 = \mathbf{b} = A\mathbf{x}_0$. Multiplying by A^{-1}, we obtain $\mathbf{x}_1 = \mathbf{x}_0$. **QED**

Example 2.5.4. Solve the following nonhomogeneous system

$$\begin{array}{rcrcrcr} x & & & + & 8z & = & 1 \\ 2x & + & 5y & + & 3z & = & 0 \\ x & + & 2y & + & 3z & = & -1 \end{array}$$

From Example 2.4.10 , the inverse of $A = \begin{pmatrix} 1 & 0 & 8 \\ 2 & 5 & 3 \\ 1 & 2 & 3 \end{pmatrix}$ is $A^{-1} = \begin{bmatrix} 9 & 16 & -40 \\ -3 & -5 & 13 \\ -1 & -2 & 5 \end{bmatrix}$.

Thus, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = A^{-1} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 49 \\ -16 \\ -6 \end{pmatrix}$

Theorem 2.5.5. *If A is an $n \times n$ matrix, then the following statements are equivalent.*

(a) A is invertible.

(b) $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

(c) $A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.

Proof. We shall prove $(a) \Longrightarrow (b) \Longrightarrow (c) \Longrightarrow (a)$.

- (a) $\Longrightarrow$ (b): This is Proposition 2.5.3.
- (b) $\Longrightarrow$ (c): Obvious.

• (c) $\Longrightarrow$ (a):

Let $\mathbf{e}_i$ be $n \times 1$ matrix defined by

$$(\mathbf{e}_i)_{k1} = \begin{cases} 1 \text{ if } k = i \\ 0 \text{ if } k \neq i \end{cases}$$

By part (c) each linear system $A\mathbf{x} = \mathbf{e}_i$ has a solution. Let $\mathbf{x}_i$ be a solution. We form the $n \times n$ matrix B with ith- column being $\mathbf{x}_i$.

Then we have $AB = I_n$.

By Proposition 2.4.13, we conclude that A is invertible. **QED.**

Remark In general, the inverse of invertible upper (resp. lower) triangular matrix will be an upper (resp. lower) triangular matrix.

Chapter 3

Determinants

Recall that a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. This special number $ad - bc$ is known as the determinant of the 2×2 square matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. It is denoted by the symbol $\det(A)$.

The determinant is a function defined on square matrices. For a general $n \times n$ square matrix A, it is defined in a slightly more complicated way.

3.1 Determinants

3.1.1 Permutation & Inversion

We begin with some terms required to facilitate the introduction of the definition of determinant.

First, we recall that a **permutation** of the set of integers $\{1, 2, \ldots, n\}$ is an arrangement of these integers in some order without omissions or repetitions.

Example 3.1.1. List all permutations of the following sets of integers

(a) $\{1, 2, 3\}$: $(1, 2, 3)$,

(b) $\{1, 2, 3, 4\}$:

Let $(j_1, j_2, \ldots, j_n)$ be a permutation of the set $\{1, 2, \ldots, n\}$. An **inversion** is said to occur in a permutation $(j_1, j_2, \ldots, j_n)$ if a larger number precedes a smaller one.

As an example, an inversion has occurred in the permutation $(2, 3, 1)$. For this permutation, there are 2 inversions.

We are interested in the total number inversion in a permutation. The total number of inversions in a permutation $(j_1, j_2, \ldots, j_n)$ can be determined systematically as follows:

(1) Find the number of integers that are less than j_1 and that follow j_1 in the permutation;

(2) Find the number of integers that are less than j_2 and that follow j_2 in the permutation;

(3) Continue in the same way for $j_3, \ldots, j_{n-1}$.

(4) The sum of these numbers will be the total number of inversions.

Example 3.1.2. For the permutation $(2, 4, 3, 1)$:

Number of integers less than 2 and follow 2 $=$

Number of integers less than 4 and follow 4 $=$

Number of integers less than 3 and follow 3 $=$

Total number of inversions is 4.

A permutation is said to be **even** if the total number of inversions is an even integer and is said to be **odd** if the total number of inversions is an odd integer.

Example 3.1.3. Classify all permutation of $\{1, 2, 3\}$.

Permutation	Number of Inversions	Even/Odd?
$(1, 2, 3)$	0	even
$(1, 3, 2)$	1	
$(2, 1, 3)$	1	
$(2, 3, 1)$	2	
$(3, 1, 2)$	2	
$(3, 2, 1)$	3	odd

3.1.2 Definition of Determinants

Now, we return to matrices. We are almost ready to define the determinant of a square matrix.

For an $n \times n$ matrix A, we introduce the term an **elementary product** of A as the product of n entries from A, no two of which come from the same row or same column.

Example 3.1.4. Consider the 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$,

The products $a_{11}a_{22}$ and $a_{12}a_{21}$ are elementary products.

The product $a_{11}a_{21}$ is not an elementary product.

Example 3.1.5. List all elementary products of the 3×3 matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$.

Solution Each elementary product takes the form $a_{1-}a_{2-}a_{3-}$.

The column-indices in each elementary product will be a permutation of $\{1, 2, 3\}$. Thus, we have the following $3! = 6$ elementary products:

$$a_{11}a_{22}a_{33}, \quad a_{11}a_{23}a_{32}, \quad a_{12}a_{21}a_{33}, \quad a_{12}a_{23}a_{31}, \quad a_{13}a_{21}a_{32}, \quad a_{13}a_{22}a_{31}$$

As seen from the last example, for an $n \times n$ square matrix A, there are $n!$ elementary products of the form $a_{1j_1}a_{2j_2}\cdots a_{nj_n}$.

We assign a sign to each elementary product: if the permutation $(j_1, j_2, \ldots, j_n)$ is odd, the elementary product is assigned -1; if the permutation $(j_1, j_2, \ldots, j_n)$ is even, the elementary product is assigned 1. Together with the sign, we have a **signed elementary product** $\pm a_{1j_1}a_{2j_2}\cdots a_{nj_n}$.

Example 3.1.6. For the 3×3 matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, we have the following 6 signed elementary products:

$$a_{11}a_{22}a_{33} \quad -a_{11}a_{23}a_{32} \quad -a_{12}a_{21}a_{33}$$
$$a_{12}a_{23}a_{31} \quad a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}$$

Now, we are ready to define what is meant by a determinant of a matrix.

Definition 3.1.7. Let A be a square matrix. The **determinant** of A is defined to be the sum of all signed elementary products from A.

The determinant function is denoted by det or $|\ |$. i.e. $\det(A) = |A|$.

In symbols, we write $\det(A) = \sum \pm a_{1j_1} a_{2j_2} \cdots a_{nj_n}$.

Example 3.1.8.

(a) The determinant of the 3×3 matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ is

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

(b) Use the definition to obtain the determinant of the 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

Thus, for $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{pmatrix}$, we have $\det(A) = -2$ and $\det(B) = 0$.

Remark Up to this point, we have assumed that the entries of our matrices are real numbers. However, we do allow the entries of matrices to be or complex numbers or integer modulo m.

For example, in $\mathbb{Z}_{26}$, if $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, then $\det(A) = 24$ (recall that $-2 \equiv 24 \pmod{26}$).

Similarly, it can be verified that the determinant of the matrix $\begin{pmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{pmatrix}$ is 8. (If you use the formula for the determinant above, you are likely to get 60 as the answer. However, recall that $60 \equiv 8 \pmod{26}$.)

3.1.3 Useful Results

It is highly challenging to compute the determinants of matrices of larger size by definition. In this section, we shall make use of elementary row operations to simplify the calculation of the determinant of a matrix. Before we do so, we compile a list of results, many of them are obtained from definition. Some are formulae for determinants of matrices of smaller sizes or special matrices.

The following is a list of useful facts or formulae on determinants. Some of them follow readily from the definition. Other may require some effort to prove. We shall discuss briefly proofs of some results in our lecture. Students who are interested to read the proofs of the other results may refer to textbook on linear algebra.

1. Determinant of a 1×1 matrix $[a]$ is a, i.e. $|a| = a$.
2. Determinant of 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $ad - bc$, or $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.
3. $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi.$
4. det(matrix with a row or a column of zeros) $= 0$.
5. det(diagonal matrix) $=$ product of diagonal entries.
6. det(triangular matrix) $=$ product of diagonal entries.
7. $\det(I) = 1$
8. $\det(AB) = \det(A)\det(B)$
9. $\det(A^{-1}) = \frac{1}{\det(A)}$
10. $\det(EA) = -\det(A)$ if E is an elementary matrix obtained from interchange of two rows of an identity matrix. Thus, interchanging two rows of a matrix changes the sign of its determinant.
11. $\det(EA) = \det(A)$ if E is an elementary matrix corresponding to adding a multiple α of row j to row i. Thus, the elementary row operation of adding a multiple α of row j to row i does not affect the determinant of the matrix.
12. $\det(EA) = \alpha\det(A)$ if E is an elementary matrix corresponding to multiplying a row by a non-zero α. Thus, the elementary row operation that multiplies a row by a non-zero α changes the determinant of the matrix A to a multiple α of its determinant.
13. $\det(A) = \det(A^T)$
14. A is invertible if and only if $\det(A) \neq 0$.

Example 3.1.9. Find the following determinants. Is the matrix invertible?

(a) $\begin{vmatrix} 4 & 0 & 0 & 5 \\ 0 & -2 & 0 & 7 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{vmatrix}$

(b) $\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$

(c) $\begin{vmatrix} 4 & 9 & 8 & 5 \\ -5 & -2 & 0 & 7 \\ 0 & 0 & 0 & 0 \\ 3 & 4 & \pi & \frac{1}{2} \end{vmatrix}$

Example 3.1.10. We shall use elementary row operations to reduce the matrix to an upper triangular matrix and compute the following determinants.

(a) $\begin{vmatrix} 0 & 0 & 1 & 5 \\ 0 & -2 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{vmatrix}$

(b) $\begin{vmatrix} 2 & 0 & 1 & 5 \\ -2 & 1 & 3 & 0 \\ 3 & 0 & 6 & 9 \\ 0 & 2 & 0 & 4 \end{vmatrix}$

3.2 Co-factor Expansion.

In this section, we shall discuss another way of finding the determinant of an $n \times n$ matrix.

3.2.1 Cofactor Expansion

Definition 3.2.1. Let A be an $n \times n$ square matrix.

(1) The **minor of entry** a_{ij} is defined to be the determinant of the submatrix that remains after the ith-row and the jth-column are deleted from A. It is denoted by M_{ij}.

(2) The **cofactor of entry** a_{ij} is the number $(-1)^{i+j}M_{ij}$. It is denoted by C_{ij}.

Note The cofactor and the minor of an element a_{ij} differ only in signs.

Example 3.2.2. Consider the matrix

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

(a) Find the minor of a_{21}.

(b) Compute M_{41}.

(c) Find the cofactor of a_{33}.

(d) Calculate C_{14}.

How is the determinant of A related to its minors or cofactors? We start with small matrices.

Example 3.2.3. Consider the matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$.

We know that $|A| = a_{11}a_{22} + a_{12}a_{21}$. Now, we note that the cofactors of entries along first row is $C_{11} = a_{22}$ and $C_{12} = -a_{21}$. Thus, we have $|A| = a_{11}C_{11} + a_{12}C_{12}$.

Example 3.2.4. Consider the matrix 3 × 3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

From the definition of the determinant of A we have

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

Let us look at cofactors of entries along second row:

$$C_{21} = -(a_{12}a_{33} - a_{13}a_{32}), C_{22} = a_{11}a_{33} - a_{13}a_{31}, C_{23} = -(a_{11}a_{32} - a_{12}a_{31})$$

Observe that

$$|A| = a_{21}C_{21} + a_{22}C_{22} + a_{23}C_{23}.$$

Exercise: Verify that

(a) $|A| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$

(b) $|A| = a_{11}C_{11} + a_{21}C_{21} + a_{31}C_{31}$

The above examples illustrate that the determinant of a matrix A can be computed via cofactor expansion along a selected row or column as follows:

1. Select a row (or column) of A.
2. Multiply each entry of the selected row (or column) by its cofactor.
3. Add all the resulting products obtained in the last step. This number is the det(A).

Therefore, for the $n \times n$ matrix A, we have

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots a_{nj}C_{nj} = \sum_{i=1}^{n} a_{ij}Cij$$

(cofactor expansion along the jth-column) or

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots a_{in}C_{in} = \sum_{j=1}^{n} a_{ij}Cij$$

(cofactor expansion along the ith-row)

Example 3.2.5. Find the determinant of

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

(b) $B = \begin{bmatrix} 1 & 1 & 5 & 0 \\ 0 & -3 & 2 & 1 \\ 0 & 1 & 2 & 1 \\ -1 & 0 & 0 & 1 \end{bmatrix}$.

Example 3.2.6. Show that $\begin{vmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & d \end{vmatrix} = abcd.$

In the cofactor expansion along a selected row (or column), we select a row (or column) to compute the product of an entry with its corresponding cofactor.

What happens if we have selected two different rows, and compute the product of entries (of a selected row) and cofactors from different row(the other selected row)? We shall see from the next example that the sum of such products will be zero.

Example 3.2.7. Consider the matrix 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

Consider cofactors of entries along second row:

$$C_{21} = -(a_{12}a_{33} - a_{13}a_{32}), C_{22} = a_{11}a_{33} - a_{13}a_{31}, C_{23} = -(a_{11}a_{32} - a_{12}a_{31})$$

and entries along first row of A. Now, we compute

$$a_{11}C_{21} + a_{12}C_{22} + a_{13}C_{23}.$$

This is actually the determinant of the following matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Its determinant is zero.

In general, the best strategy for evaluating a determinant by cofactor expansion is to expand along a row or column having the largest number of zeros.

Example 3.2.8. Compute $\begin{vmatrix} 3 & 1 & 2 & 2 \\ 1 & 0 & -2 & 1 \\ 1 & 0 & 5 & -1 \\ 2 & 0 & 0 & 1 \end{vmatrix}$.

Cofactor expansion and row (or column) operations can sometimes be used in combination to provide an effective method for evaluating determinants.

Example 3.2.9. Find $\begin{vmatrix} 3 & 1 & 2 & 2 \\ 4 & 1 & 2 & 2 \\ 1 & 0 & 5 & -1 \\ 2 & 0 & 0 & 1 \end{vmatrix}$.

3.2.2 Adjoint of A

Definition 3.2.10. Let A be an $n \times n$ matrix, and C_{ij} be the cofactor of a_{ij}. Then the matrix

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

is called the **matrix of cofactors from** A.

The transpose of this matrix is called the **adjoint of A** and is denoted by $\text{adj}(A)$.

Example 3.2.11. Find the adjoint of $\begin{bmatrix} 1 & 5 & 0 \\ -3 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix}$.

We state following relationship between adjoint, determinant and inverse of a matrix.

Proposition 3.2.12. *Let A be an $n \times n$ square matrix. Then*

$$A \; adj(A) = \det(A)I.$$

Thus, if $\det(A) \neq 0$*, then A is invertible and*

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

On the other hand, the matrix A is singular if and only if $\det(A) = 0$.

Proof. It follows from

$$A\text{adj}(A) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{j1} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{j2} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{jn} & \cdots & C_{nn} \end{bmatrix}$$

that the ij-th entry of the product $A\text{adj}(A)$ is

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn}$$

Case $i = j$: The ij-th entry of the product $A\text{adj}(A)$ is $\det(A)$.

Case $i \neq j$: The ij-th entry of the product $A\text{adj}(A)$ is 0, since the entries and cofactors come from different rows.

In conclusion, we have $A \; adj(A) = \det(A)I$. **QED.**

Example 3.2.13. Find the inverse of $\begin{bmatrix} 1 & 5 & 0 \\ -3 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix}$ via adjoint.

3.3 Cramer's Rule

For a linear system $\mathbf{Ax} = \mathbf{b}$ whose coefficient matrix A is invertible, there is a formula for its solution. The formula is known as Cramer's rule. It is useful for studying the mathematical properties of a solution without the need for solving the system.

Theorem 3.3.1 (Cramer's Rule). *If* $\mathbf{Ax} = \mathbf{b}$ *is a system of* n *linear equations in* n *unknowns such that* $\det(A) \neq 0$, *then the system has a unique solution, namely*

$$x_j = \frac{\det(A_j)}{\det(A)}, j = 1, 2, \dots, n$$

where

$$A_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j-1} & b_1 & a_{1j+1} \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2j-1} & b_2 & a_{2j+1} \cdots & a_{2n} & \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj-1} & b_n & a_{nj+1} \cdots & a_{nn} & \end{bmatrix},$$

the matrix obtained by replacing the entries in the j*th column of* $\mathbf{A}$ *by the entries in the matrix*

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

Proof. Since $\det(A) \neq 0$, A is invertible and

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

.

Exercise Verify that the jth row of $\text{adj}(A)\mathbf{b}$ is the determinant of the matrix

$$\mathbf{A}_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j-1} & b_1 & a_{1j+1} \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2j-1} & b_2 & a_{2j+1} \cdots & a_{2n} & \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj-1} & b_n & a_{nj+1} \cdots & a_{nn} & \end{bmatrix}$$

obtained by replacing the entries in the jth column of $\mathbf{A}$ by the entries in the matrix

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

QED

Example 3.3.2. Use Cramer's Rule to solve the linear system

$$\begin{array}{rcrl} 7x_1 & - & 2x_2 & = 3 \\ 3x_1 & + & x_2 & = 5 \end{array}$$

[Solution] (Exercise)

Chapter 4

Vectors

You are probably already familiar with vectors in the plane or in 3-space. In this chapter we will will review some of this theory and also look at vectors in higher dimensions. This will be the foundation for the rest of the course.

4.1 Vectors in $\mathbb{R}^n$

Definition 4.1.1. Let n be a positive integer. With $\mathbb{R}^n$ we mean the set of all ordered n-tuples of elements from $\mathbb{R}$, that is $\mathbf{x} \in \mathbb{R}^n$ if

$$\mathbf{x} = (x_1, x_2, \ldots, x_n), \quad \text{where } x_j \in \mathbb{R}, 1 \leq j \leq n.$$

An element $\mathbf{x}$ in $\mathbb{R}^n$ is also called a *vector* in $\mathbb{R}^n$.

Example 4.1.2. The object $\mathbf{x} = (1, \pi, -5, 0, \sin 2)$ is a vector in $\mathbb{R}^5$.

A vector in $\mathbb{R}^2$ or $\mathbb{R}^3$, we think of as an object having a length and a direction and we illustrate it by drawing an arrow. A vector from the point $P_1 = (x_1, y_1, z_1)$ to the point $P_2 = (x_2, y_2, z_2)$ will be the vector $\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$. In particular, if the starting point is the origin $O = (0, 0, 0)$ and the end point $P = (x, y, z)$, then $\overrightarrow{OP} = (x - 0, y - 0, z - 0) = (x, y, z)$. Note that we can choose the starting point of a vector $\mathbf{x}$ freely. For example $\mathbf{x} = (x, y, z) = \overrightarrow{(0, 0, 0)(x, y, z)} = \overrightarrow{(x_1, y_1, z_1)(x_1 + x, y_1 + y, z_1 + z)}$ (see Figure 4.1.1).

For vectors in $\mathbb{R}^n$ where $n \geq 4$, we don't really have any geometric interpretation. It is however still useful to think in geometric terms. What we mean by this is that even if a vector in $\mathbb{R}^4$ can't be properly represented geometrically, it is still very useful for us to let our geometric intuition about $\mathbb{R}^2$ and $\mathbb{R}^3$ guide us also for $\mathbb{R}^4$. It will often lead us right anyway.

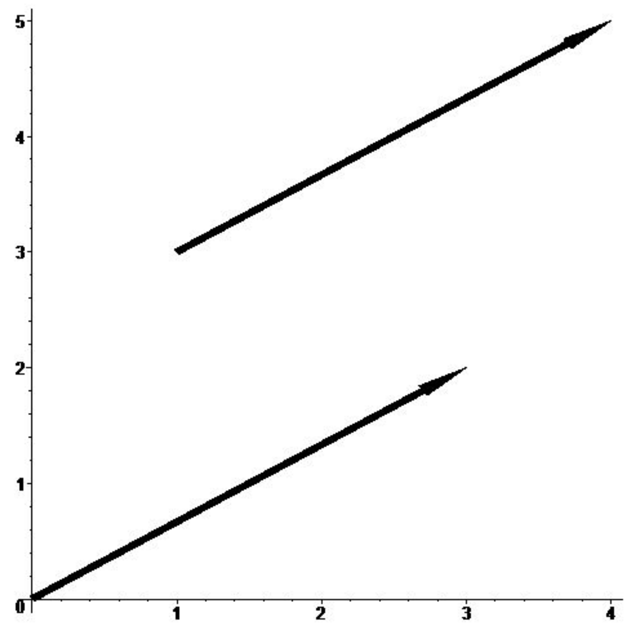

Figure 4.1.1: The vector $(3, 2)$ drawn with two different starting positions.

Note We can also represent a vector $\mathbf{x}$ in $\mathbb{R}^n$, like this

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

So, with $[\mathbf{x}]$ we mean the column matrix we get from the components of $\mathbf{x}$. If we want the corresponding row matrix, that will be $[\mathbf{x}]^T$. We will however abuse this notation a little bit and sometimes just write $\mathbf{x}$ when we really mean the column matrix $[\mathbf{x}]$. We have already done this, and there should be no confusion when we for example write $A\mathbf{x} = \mathbf{b}$ when we really mean $A[\mathbf{x}] = [\mathbf{b}]$, as a compact notation for the linear system

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

4.2 Vector Addition & Scalar Multiplication

We can add vectors, and we can multiply a vector with a scalar.

Definition 4.2.1. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be vectors in $\mathbb{R}^n$ and let

k be a scalar. Then we define the sum $\mathbf{x}+\mathbf{y}$ and the product $k\mathbf{x}$ as follows:

$$\mathbf{x}+\mathbf{y} = (x_1+y_1, x_2+y_2, \ldots, x_n+y_n), \quad k\mathbf{x} = (kx_1, kx_2, \ldots, kx_n).$$

These operations can also be interpreted geometrically in $\mathbb{R}^2$ and $\mathbb{R}^3$ as illustrated in Figure 4.2.2.

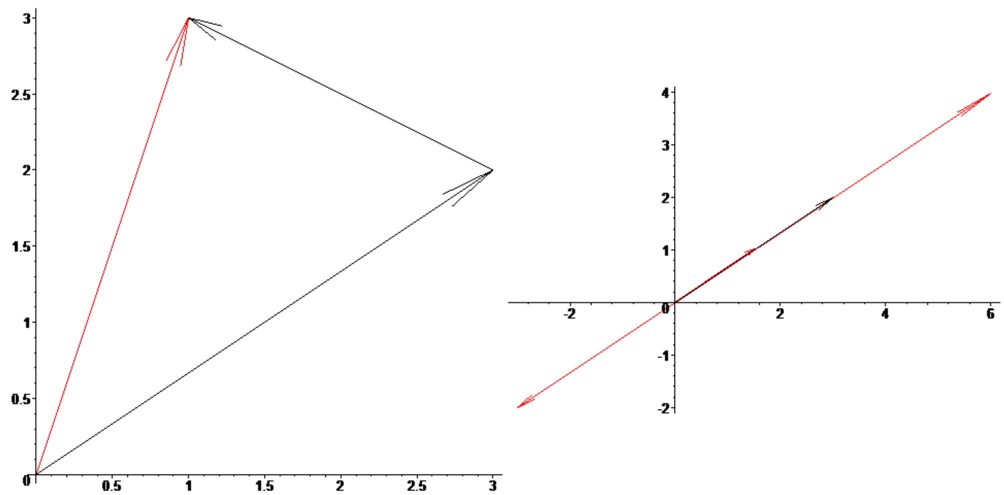

Figure 4.2.2: Geometric interpretation of vector addition and multiplication with scalar. The left picture illustrates the sum $(3,2)+(-2,1)=(1,3)$. The right picture shows $(3,2)$ together with $2\,(3,2)$, $\frac{1}{2}\,(3,2)$ and $-1\,(3,2)$.

We also have the zero vector and the additive inverse.

Definition 4.2.2. The *zero vector* in $\mathbb{R}^n$ is the vector

$$\mathbf{0} = (0, 0, \ldots, 0).$$

If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a vector in $\mathbb{R}^n$, then the additive inverse (or negative) of $\mathbf{x}$ is denoted by $-\mathbf{x}$ and is given by

$$-\mathbf{x} = (-x_1, -x_2, \ldots, -x_n).$$

The difference $\mathbf{x}-\mathbf{y}$ of two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\mathbf{x}-\mathbf{y} = \mathbf{x}+(-\mathbf{y}).$$

Vector addition and multiplication with scalar, follow these rules which are consequences of properties (such commutativity, associativity, distributivity) of real numbers.

Theorem 4.2.3. *If $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are vectors in $\mathbb{R}^n$, and if k and m are scalars, then*

1. $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$,

2. $\mathbf{x}+(\mathbf{y}+\mathbf{z}) = (\mathbf{x}+\mathbf{y})+\mathbf{z}$,

3. $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$,

4. $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$,

5. $1\mathbf{x} = \mathbf{x}$,

6. $k(m\mathbf{x}) = (km)\mathbf{x}$,

7. $(k + m)\mathbf{x} = k\mathbf{x} + m\mathbf{x}$,

8. $k(\mathbf{x} + \mathbf{y}) = k\mathbf{x} + k\mathbf{y}$.

4.3 Scalar Product & Norm

4.3.1 Scalar Product

There are also different ways of multiplying vectors. The most useful product for us will be the scalar product, or *Euclidean inner product* (also known as *dot product*).

Definition 4.3.1. If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ are vectors in $\mathbb{R}^n$, then their *scalar product* (or *Euclidean inner product*) $\mathbf{x} \cdot \mathbf{y}$, is defined by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

Note that we can also describe scalar multiplication using matrix multiplication like this;

$$[\mathbf{x} \cdot \mathbf{y}] = [x_1 y_1 + x_2 y_2 + \ldots + x_n y_n] = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = [\mathbf{x}]^T [\mathbf{y}].$$

We have the following arithmetic rules for scalar multiplication:

Theorem 4.3.2. *If $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, are vectors in $\mathbb{R}^n$ and k is any scalar, then*

(a) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.

(b) $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$.

(c) $(k\mathbf{x}) \cdot \mathbf{y} = k(\mathbf{x} \cdot \mathbf{y})$.

(d) $\mathbf{x} \cdot \mathbf{x} \geq 0$. *Also,* $\mathbf{x} \cdot \mathbf{x} = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

Proof. We will prove part (b). The proofs of the other parts are left as exercises.
The vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ have the general form

$$\mathbf{x} = (x_1, x_2, \ldots, x_n), \quad \mathbf{y} = (y_1, y_2, \ldots, y_n), \quad \mathbf{z} = (z_1, z_2, \ldots, z_n).$$

By the definition of scalar product we have then

$$\begin{aligned}(\mathbf{x}+\mathbf{y})\cdot\mathbf{z} = (x_1+y_1, x_2+y_2, \ldots, x_n+y_n)\cdot(z_1, z_2, \ldots, z_n) = \\ (x_1+y_1)z_1 + (x_2+y_2)z_2 + \ldots + (x_n+y_n)z_n = \\ (x_1z_1 + x_2z_2 + \ldots + x_nz_n) + (y_1z_1 + y_2z_2 + \ldots + y_nz_n) = \mathbf{x}\cdot\mathbf{z} + \mathbf{y}\cdot\mathbf{z}.\end{aligned}$$

□

4.3.2 Norm

In $\mathbb{R}^3$ (or $\mathbb{R}^2$) we know that the length of the vector $\mathbf{x} = (x_1, x_2, x_3)$ (or $\mathbf{x} = (x_1, x_2)$) is given by $\sqrt{x_1^2 + x_2^2 + x_3^2}$ (or $\sqrt{x_1^2 + x_2^2}$). In both cases this is the same as $\sqrt{\mathbf{x}\cdot\mathbf{x}}$. In higher dimensions it is still useful to think about this quality as length, but we call it "norm" instead.

Definition 4.3.3. The norm $\|\mathbf{x}\|$ of a vector $\mathbf{x}$ in $\mathbb{R}^n$ is defined by

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}\cdot\mathbf{x}}.$$

Note that since $\mathbf{x}\cdot\mathbf{x} \geq 0$, taking the square root is possible and the definition makes sense.

Example 4.3.4. The norm of the vector $\mathbf{x} = (1, 0, 3, -2)$ in $\mathbb{R}^4$ is

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}\cdot\mathbf{x}} = \sqrt{1^2 + 0^2 + 3^2 + (-2)^2} = \sqrt{14}.$$

A very important result in linear algebra is the following inequality

Theorem 4.3.5 (Cauchy-Schwarz inequality)**.** *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be vectors in* $\mathbb{R}^n$*. Then*

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

We skip the proof for now, as this is a special case of Theorem 9.1.7 which will be proved independently later.

The norm also follows these rules:

Theorem 4.3.6. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be vectors in* $\mathbb{R}^n$ *and* k *a scalar. Then we have*

(a) $\|\mathbf{x}\| \geq 0$.

(b) $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

(c) $\|k\mathbf{x}\| = |k|\|\mathbf{x}\|$.

(d) $\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$.

Part (d) is known as the triangle inequality. *Can you guess why?*

Proof. We prove part (d). The other parts are left as exercises.

To prove (d), note that by using the definition of norm, properties of scalar product and the Cauchy-Schwarz inequality, we get

$$\|\mathbf{x}+\mathbf{y}\|^2 = (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y}) = \mathbf{x}\cdot\mathbf{x} + 2\mathbf{x}\cdot\mathbf{y} + \mathbf{y}\cdot\mathbf{y} \le \mathbf{x}\cdot\mathbf{x} + 2|\mathbf{x}\cdot\mathbf{y}| + \mathbf{y}\cdot\mathbf{y} \le \\ \mathbf{x}\cdot\mathbf{x} + 2\|\mathbf{x}\|\|\mathbf{y}\| + \mathbf{y}\cdot\mathbf{y} = \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.$$

Taking square roots of both sides (noting that the norms are nonnegative) we end up with

$$\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|.$$

□

4.3.3 Projections

In 2 or 3 dimensional space we have a geometric interpretation of the scalar product.

Theorem 4.3.7. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two vectors in* $\mathbb{R}^2$ *or* $\mathbb{R}^3$*, and let* θ *be the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *(here we specify* θ *to be in the interval* $[0, \pi]$*). Then*

$$\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos\theta.$$

Proof. Tutorial problem. □

Using this fact, we also get formulas for projection of vectors. Suppose $\mathbf{x}$ and $\mathbf{a}$ are vectors in 2 or 3 dimensional space. Then, there is a unique way of writing $\mathbf{x}$ as a sum $\mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1$ is a multiple of $\mathbf{a}$ (i.e. parallell to $\mathbf{a}$) and $\mathbf{x}_2$ is orthogonal to $\mathbf{a}$. The vector $\mathbf{x}_1$ is called the *orthogonal projection* of $\mathbf{x}$ onto $\mathbf{a}$ and is denoted $\text{proj}_{\mathbf{a}}\mathbf{x}$.

Theorem 4.3.8. *With notation as above, we have*

$$\text{proj}_{\mathbf{a}}\mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}$$

and

$$\mathbf{x}_2 = \mathbf{x} - \text{proj}_{\mathbf{a}}\mathbf{x}.$$

Proof. To be done at lecture. □

4.4 Orthogonality & The Standard Basis

4.4.1 Orthogonality

We just saw that in $\mathbb{R}^2$ or $\mathbb{R}^3$, if we let θ be the angle between the (nonzero) vectors $\mathbf{x}$ and $\mathbf{y}$, then

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\cdot\|\mathbf{y}\|}.$$

In particular, two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ are perpendicular if and only if $\mathbf{x}\cdot\mathbf{y} = 0$. This inspires us to make the following definition

Definition 4.4.1. Two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ are said to be *orthogonal* if $\mathbf{x}\cdot\mathbf{y} = 0$.

Example 4.4.2. The vectors $\mathbf{x} = (1, 2, 1, 2)$ and $\mathbf{y} = (3, 3, -3, -3)$ are orthogonal vectors in $\mathbb{R}^4$, since

$$\mathbf{x}\cdot\mathbf{y} = 1\cdot 3 + 2\cdot 3 + 1\cdot(-3) + 2\cdot(-3) = 0.$$

Let's note two things about orthogonal vectors:

1. In $\mathbb{R}^2$ or $\mathbb{R}^3$, two nonzero vectors are perpendicular if and only if they are orthogonal. It's useful to think about orthogonal vectors in higher dimensions in the same way, even if it makes no geometrical sense.

2. Since $\mathbf{0}\cdot\mathbf{x} = 0$, the zero vector is orthogonal to all vectors.

If we think geometrically, the following theorem comes as no surprise

Theorem 4.4.3. *If $\mathbf{x}$ and $\mathbf{y}$ are orthogonal vectors in $\mathbb{R}^n$, then*

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + |\mathbf{y}\|^2.$$

Proof. Exercise. □

4.4.2 The Standard Basis

We denote the following unit vectors in $\mathbb{R}^n$:

$$\begin{aligned}\mathbf{e}_1 &= (1, 0, \ldots, 0),\\ \mathbf{e}_2 &= (0, 1, \ldots, 0),\\ &\vdots\\ \mathbf{e}_n &= (0, 0, \ldots, 1).\end{aligned}$$

They are called **standard unit vectors** in $\mathbb{R}^n$.

Observation

1. For $1 \leq i \leq n$, each vector $\mathbf{e}_i$ has unit norm,i.e., $\|\mathbf{e}_i\| = 1$.
2. If $i \neq j$, then vectors $\mathbf{e}_i$ and $\mathbf{e}_j$ are orthogonal.
3. Moreover, for any vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$, we have

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n.$$

Definition 4.4.4. The set of vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ in $\mathbb{R}^n$ described above, is called the *standard basis* for $\mathbb{R}^n$.

We will talk more about bases later in the last topic on vector spaces, but for now let us introduce a new term to describe the last point in the above observation.

Definition 4.4.5. Suppose $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are vectors in $\mathbb{R}^n$. Then the vector $\mathbf{v}$ is said to be a **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ if

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

for some scalars $c_1, c_2, \ldots, c_k$.

Example 4.4.6. Let $\mathbf{v} = (x, y) \in \mathbb{R}^2$. It is clear that we have

$$\mathbf{v} = (x, y) = x(1, 0) + y(0, 1).$$

So, every vector in $\mathbb{R}^2$ is a linear combination of the standard vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Moreover, the scalars are also unique. (Can you prove it?)

So, each $\mathbf{v} = (x, y) \in \mathbb{R}^2$ has a unique representation with respect to the set of standard unit vectors.

Example 4.4.7. Every vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ is a linear combination of the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$.

Example 4.4.8. The vector $(1, 2, 5)$ is a linear combination of $(1, 3, 0)$, (0,2,7) , $(3, 0, 8)$ and $(0, 0, 1)$, because

$$(1, 2, 5) = -2(1, 3, 0) + 4(0, 2, 7) + (3, 0, 8) - 31(0, 0, 1)$$

Note that we can also express $(1, 2, 5)$ as follows:

$$(1, 2, 5) = 4(1, 3, 0) - 5(0, 2, 7) - (3, 0, 8) + 48(0, 0, 1)$$

The scalars may not be unique in general.

Chapter 5

More about vectors in 2 and 3 dimensions

5.1 The cross product

Let $\mathbf{x} = (x_1, x_2, x_3)$ and $\mathbf{y} = (y_1, y_2, y_3)$ be two vectors in $\mathbb{R}^3$. The cross product $\mathbf{x} \times \mathbf{y}$ of $\mathbf{x}$ and $\mathbf{y}$ is defined as follows:

$$\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2, -x_1y_3 + x_3y_1, x_1y_2 - x_2y_1)$$

Using the notation, $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$ and $\mathbf{k} = (0, 0, 1)$, we may make use of the determinant to express $\mathbf{x} \times \mathbf{y}$ as:

$$\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2)\mathbf{i} + (-x_1y_3 + x_3y_1)\mathbf{j} + (x_1y_2 - x_2y_1)\mathbf{k} = \det\left(\begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}\right)$$

Some useful properties are

1. $\|\mathbf{x} \times \mathbf{y}\| = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \sin\theta$, where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.
2. If $\mathbf{x}$ and $\mathbf{y}$ are parallel, then $\mathbf{x} \times \mathbf{y} = \mathbf{0}$.
3. $\mathbf{x} \times \mathbf{y} = -(\mathbf{x} \times \mathbf{y})$.
4. $\mathbf{x} \times \mathbf{y}$ is orthogonal (perpendicular) to both vectors $\mathbf{x}$ and $\mathbf{y}$
5. For vectors $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{y} = (y_1, y_2, y_3)$ and $\mathbf{z} = (z_1, z_2, z_3)$, we have

$$\mathbf{x} \cdot (\mathbf{y} \times \mathbf{z}) = \det\left(\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}\right).$$

5.2 Lines and planes

5.2.1 Planes

A plane in $\mathbb{R}^3$ can be described in several ways. The most common being the following.

1. Specify three points in the plane, not all lying along the same line, *or*
2. specify one point in the plane, and the orientation of the plane by specifying a normal vector for the plane.

With a *normal vector* above, we simply mean a vector that is perpendicular to the plane. From this information it is then possible to derive an equation for the plane, that is an equation in x, y and z such that (x_0, y_0, z_0) lies in the plane if and only if the equation is satisfied with $x = x_0$, $y = y_0$, $z = z_0$.

We will now explain how such an equation for a plane Π can be derived in the second case, when we know a point $P_0(x_0, y_0, z_0)$ in Π and we know a normal vector $\mathbf{n} = (a, b, c)$ for Π (see Fig. 5.2.1).

Let, $P(x, y, z)$ be an arbitrary point. Since $\mathbf{n}$ in perpendicular to Π, we see that P is in Π if and only if $\overrightarrow{P_0P}$ is perpendicular to $\mathbf{n}$, that is if and only if $\mathbf{n} \cdot \overrightarrow{P_0P} = 0$. Evaluating this dot product using the vector componentns, we see that (x, y, z) lies in the plane Π if and only if

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0. \tag{5.1}$$

If we prefer we can multiply the above factors together, and by setting $d = -ax_0 - by_0 - cz_0$, the above equation takes the form

$$ax + by + cz + d = 0. \tag{5.2}$$

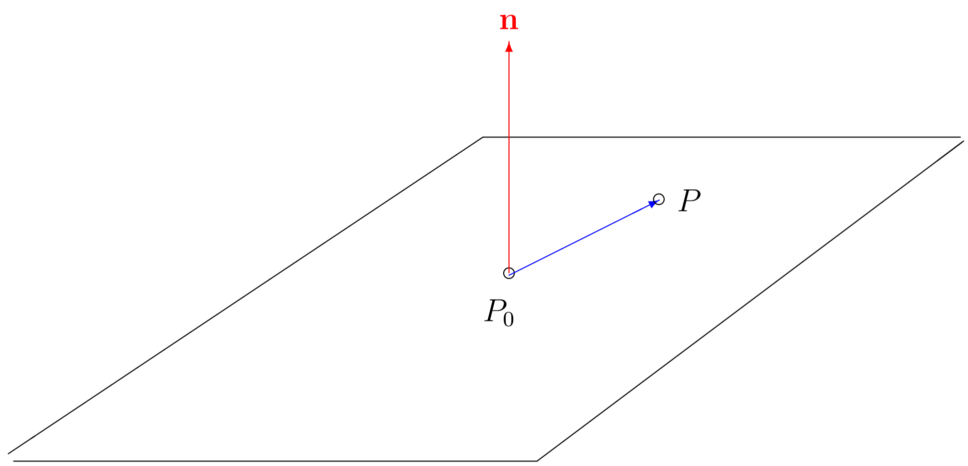

Figure 5.2.1: The plane determined by the point P_0 and the normal vector $\mathbf{n}$.

The equation (5.1) is referred to as an equation for the plane in *point-normal form* while (5.2) is an equation for the plane in *standard form*.

Example 5.2.1. Find an equation for the plane that contains the three points $(1, 0, 1)$, $(1, 2, 3)$ and $(3, 2, 1)$.

Solution: To be done at lecture.

5.2.2 Lines

There are different ways to describe a line in two or three dimensional space, but the one most useful for our purpose is the parametric representation. This has the advantage of being similar in both two and three dimensional space. To derive the parametric form of a line, we need two pieces of information, a point on the line and a direction. Suppose P_0 is a point on the line ℓ and suppose $\mathbf{v}$ is a vector parallell to ℓ. Then a point P lies on ℓ if and only if for some scalar t,

$$\overrightarrow{OP} = \overrightarrow{OP_0} + t\mathbf{v}.$$

Hence, if we're in three dimensional space and if $P_0 = (x_0, y_0, z_0)$, $\mathbf{v} = (a, b, c)$ and $P = (x, y, z)$, then the above equation takes the form

$$\begin{aligned} x &= x_0 + at, \\ y &= y_0 + bt, \quad t \in \mathbb{R} \\ z &= z_0 + ct. \end{aligned}$$

For lines in two dimensional space, we get the same type of parametric equation, but with one less coordinate.

Example 5.2.2. Find a parametric equation for the line passing through the points $(1, 2, 3)$ and $(3, 2, 1)$.

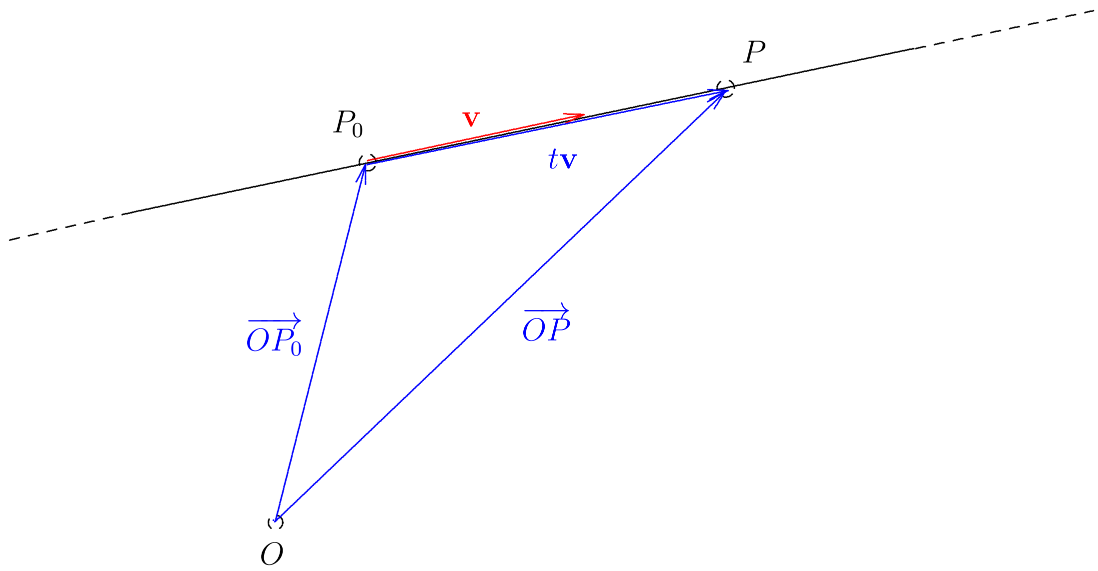

Figure 5.2.2: The line parallell to $\mathbf{v}$ passing through P_0.

Example 5.2.3. Find a parametric equation for the line that passes through $(1, 0, 0)$ and which is parallell to both the plane $x + y + z = 0$ and the plane $x + y - z = 2$.

5.3 Geometric View of Determinants

Restricting ourselves to $\mathbb{R}^3$, we shall prove that the volume of the parallelepiped determined by three vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ (see Figure 5.3.3), is given by the modulus of the determinant of the matrix we get with these three vectors as rows (or columns).

Now, for the above arguments to mean anything to us, we should prove that the modulus of the determinant of a 3×3 matrix, is the same as the volume determined by its row vectors.

5.3.1 The volume of a parallelepiped

Consider three nonzero vectors $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{y} = (y_1, y_2, y_3)$ and $\mathbf{z} = (z_1, z_2, z_3)$. Let

$$M = \begin{bmatrix} \begin{bmatrix}\mathbf{x}\end{bmatrix} \begin{bmatrix}\mathbf{y}\end{bmatrix} \begin{bmatrix}\mathbf{z}\end{bmatrix} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{bmatrix}.$$

We have

$$\det(M) = x_1 \begin{vmatrix} y_2 & z_2 \\ y_3 & z_3 \end{vmatrix} - x_2 \begin{vmatrix} y_1 & z_1 \\ y_3 & z_3 \end{vmatrix} + x_3 \begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix} = \mathbf{x}\cdot(\mathbf{y} \times \mathbf{z}). \tag{5.3}$$

Let us now calculate the volume of the parallelepiped determined by $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ and see that we get the modulus of the above expression.

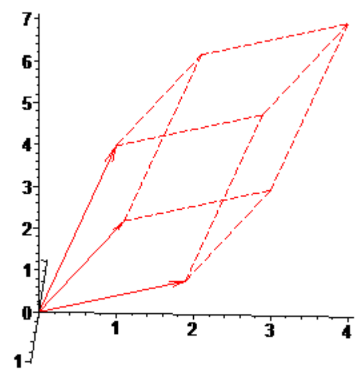

Figure 5.3.3: The parallelepiped determined by three vectors.

Consider the parallelogram determined by $\mathbf{y}$ and $\mathbf{z}$ as the base of the parallelepiped. (Here, we assume that $\mathbf{y}$ and $\mathbf{z}$ are not parallel.) The area of this is given by

$$A = \|\mathbf{y} \times \mathbf{z}\|.$$

To get the volume V, we must find the height h, then

$$V = Ah.$$

To find h we note that if $\mathbf{a}$ is a unit vector perpendicular to both $\mathbf{y}$ and $\mathbf{z}$, then h will be the length of the projection of x orthogonally on $\mathbf{a}$, that is

$$h = \|\mathbf{x}\| \cos\theta = |\mathbf{x} \cdot \mathbf{a}|,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{a}$.

Since $\mathbf{y} \times \mathbf{z}$ is perpendicular to both $\mathbf{y}$ and $\mathbf{z}$, a possible choice for $\mathbf{a}$ is

$$\mathbf{a} = \frac{\mathbf{y} \times \mathbf{z}}{\|\mathbf{y} \times \mathbf{z}\|}.$$

Then we have

$$\begin{aligned} h = |\mathbf{x} \cdot \mathbf{a}| &= |\mathbf{x} \cdot \left(\frac{\mathbf{y} \times \mathbf{z}}{\|\mathbf{y} \times \mathbf{z}\|} \right) | \\ &= | \frac{\mathbf{x} \cdot (\mathbf{y} \times \mathbf{z})}{\|\mathbf{y} \times \mathbf{z}\|} | = \frac{|\det(M)|}{\|\mathbf{y} \times \mathbf{z}\|} \end{aligned}$$

Combining the above with equation (5.3), we get

$$V = Ah = |\det(M)| = |\mathbf{x}\cdot(\mathbf{y} \times \mathbf{z})|$$

as required.

A useful consequence
Vectors $\mathbf{x}$,$\mathbf{y}$ and $\mathbf{z}$ are coplanar (i.e., they lie on the same plane) if and only if $\mathbf{x}\cdot(\mathbf{y} \times \mathbf{z}) = 0$.

(This is useful to check whether three vectors are coplanar.)

Remark The corresponding result holds also in $\mathbb{R}^2$, the area of the parallelogram determined by two vectors is the same as the modulus of the corresponding determinant (can you prove this?). For 4 and higher dimensions, we can interpret the determinant as some kind of multidimensional "volume" as well.

Chapter 6

Linear Transformations

6.1 Linear Transformations

One reason that the "strange" definition of matrix multiplication is useful is what we have already seen, that it helps us in the study of linear systems. A related concept is that of linear transformations.

6.1.1 Linear Transformations

We recall that a function $F : \mathbb{R}^n \to \mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x} \in \mathbb{R}^n$ a unique vector $F(\mathbf{x}) \in \mathbb{R}^m$.

Some basic terminology

1. The **domain** of the function F is $\mathbb{R}^n$ and the **codomain** of F is $\mathbb{R}^m$.
2. The vector $F(\mathbf{x})$ is called the image of $\mathbf{x}$ under F.
3. The set of all images $F(\mathbf{x})$ as $\mathbf{x}$ runs throughout the domain of F is called the **range** of F.

Example 6.1.1. Consider the function $F : \mathbb{R}^3 \to \mathbb{R}^2$ where $F(x_1, x_2, x_3) = (x_1 + x_2, x_3^2)$.

(a) What is the domain of F?

(b) What is the codomain of F?

(c) Determine the images of $(1, 0, 0)$, $(0, 1, 0)$ and $(1, 1, 0)$ under F.

(d) Find all vectors $\mathbf{x} \in \mathbb{R}^3$ such that $F(\mathbf{x}) = (0, 4)$.

(e) Is there a vector $\mathbf{x} \in \mathbb{R}^3$ such that $F(\mathbf{x}) = (0, -4)$?

(Note: For this function F, its range is $\{(y_1, y_2) \in \mathbb{R}^2 | y_2 \geq 0\}$.)

Since addition and scalar multiplication are defined on both domain and codomain, the next question we could ask is which functions preserve addition and scalar multiplication. This leads to the definition of linear transformation.

Linear transformations appear widely in many areas such as geometry, cryptography, coding, social sciences. We shall soon see that they are closely related to matrix multiplication.

Definition 6.1.2. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a *linear* transformation if and only if the following conditions both hold:

1. $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,
2. $T(c\mathbf{x}) = cT(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$, $c \in \mathbb{R}$.

Note The first condition says that T preserves addition, while the second condition says that T preserves scalar multiplication.

Example 6.1.3. [Reflection about the x-axis] Consider the reflection of a point $\mathbf{x} = (x, y) \in \mathbb{R}^2$ about the x-axis. It can be described by the the function $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x \\ -y \end{bmatrix}$$

Is T a linear transformation?

[Solution] We have to check the 2 conditions are satisfied.

• T preserves addition.

Take any two vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ in $\mathbb{R}^2$. Then

$$T(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}) = T(\begin{bmatrix} x + x_1 \\ y + y_1 \end{bmatrix}) = (\begin{bmatrix} x + x_1 \\ -(y + y_1) \end{bmatrix})$$

$$T(\begin{bmatrix} x \\ y \end{bmatrix}) + T(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}) = (\begin{bmatrix} x \\ -y \end{bmatrix}) + (\begin{bmatrix} x_1 \\ -y_1 \end{bmatrix}) = (\begin{bmatrix} x + x_1 \\ -y - y_1 \end{bmatrix}) = (\begin{bmatrix} (x + x_1) \\ -(y + y_1) \end{bmatrix})$$

Thus, $T(\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}) = T(\begin{bmatrix} x \\ y \end{bmatrix}) + T(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix})$.

• T preserves scalar multiplication.

Take any vector $\begin{bmatrix} x \\ y \end{bmatrix}$ in $\mathbb{R}^2$ and any scalar $c \in \mathbb{R}$. Then

$$T(c\begin{bmatrix} x \\ y \end{bmatrix}) = T(\begin{bmatrix} cx \\ cy \end{bmatrix}) = \begin{bmatrix} cx \\ -cy \end{bmatrix} = c\begin{bmatrix} x \\ -y \end{bmatrix} = cT(\begin{bmatrix} x \\ y \end{bmatrix})$$

Therefore, T is a linear transformation.

Example 6.1.4. Is the function $F : \mathbb{R}^3 \to \mathbb{R}^2$ where $F(x_1, x_2, x_3) = (x_1 + x_2, x_3^2)$ defined in Example 6.1.1 a linear transformation?

[Solution] (If at least one of the 2 conditions is not satisfied, then F is not a linear transformation.)

$F(2(0, 0, 1)) =$

$2F(0, 0, 1) =$

Since $F(2(0, 0, 1)) \neq 2F(0, 0, 1)$, the second condition is not satisfied for all vectors in $\mathbb{R}^3$. Thus, F is not a linear transformation.

Remark In $\mathbb{R}^2$ and $\mathbb{R}^3$, the following geometrical transformations are linear transformations: rotations, reflections, projection, dilations and contraction.

(You could check geometrically or algebraically that they preserve addition and scalar multiplication.)

Proposition 6.1.5. *A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if $T(c\mathbf{x} + d\mathbf{y}) = cT(\mathbf{x}) + dT(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and for all scalars $c, d \in \mathbb{R}$.*

[Proof.] (Exercise.)

Observe that the reflection about the x-axis in Example 6.1.3 can be expressed in terms of matrix multiplication:

$$T(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Let us represent vectors in $\mathbb{R}^n$ as column vectors. We shall see that an $n \times m$ matrix A defines a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Theorem 6.1.6. *Let A be an $m \times n$ matrix. Then the function $T_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by*

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

is a linear transformation.

[Proof.] The 2 conditions in the definition of a linear transformation follow from properties of matrix multiplication. □

Example 6.1.7. Consider the projection $P : \mathbb{R}^3 \to \mathbb{R}^3$ that projects a vector in $\mathbb{R}^3$ onto the xz-plane. Is P a linear transformation?

[Solution] The projection P can be described easily:

Let $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$. Then

$$P(\mathbf{x}) = \begin{bmatrix} x \\ 0 \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

By Theorem 6.1.6, the projection P is a linear transformation.

6.1.2 The Standard Matrix

The next question facing us is

Can every linear transformation T be expressed as a T_A for some matrix A?

Theorem 6.1.8. *The transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if $T = T_A$ for an $m \times n$ matrix A.*

Proof. Note that the 'If' part is just Theorem 6.1.6.

It suffices to prove the 'Only if' part, i.e., If T is a linear transformation, then there is an $m \times n$ matrix A such that $T = T_A$.

To prove this, we will construct a matrix A such that

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

Let us first observe however that condition that T preserves addition also means that for k vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$, we have

$$T(\mathbf{v}_1 + \mathbf{v}_2 + \ldots + \mathbf{v}_k) = T(\mathbf{v}_1) + T(\mathbf{v}_2) + \ldots + T(\mathbf{v}_k),$$

(you can prove this using induction) and combining this with the condition that T preserves scalar multiplication,we get for scalars $c_1, \ldots, c_k$ that

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \ldots + c_kT(\mathbf{v}_k). \tag{6.1}$$

The matrix A is constructed as follows. Let

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Then we let

$$A = [T(\mathbf{e}_1)|T(\mathbf{e}_2)| \cdots |T(\mathbf{e}_n)],$$

meaning that the jth column of A is the column vector $T(\mathbf{e}_j)$.

With

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

we get by using Proposition 2.2.11 and Equation (6.1) (in that order) that

$$\begin{aligned} A\mathbf{x} = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \ldots + x_nT(\mathbf{e}_n) = \\ = T\left((x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n\right) = T(\mathbf{x}) \end{aligned}$$

Therefore, we have $T = T_A$. □

Note that each linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ defines a unique matrix A via the set of unit vectors $\mathbf{e}_j \in \mathbb{R}^n$.

Definition 6.1.9. For a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, the standard matrix, denoted by $[T]$, of T is the following matrix

$$[T] = [T(\mathbf{e}_1)|T(\mathbf{e}_2)| \cdots |T(\mathbf{e}_n)], \tag{6.2}$$

whose jth column is the column vector $T(\mathbf{e}_j)$.

The standard matrix provides us with a method to determine the rule of a linear transformation easily.

Example 6.1.10. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that rotates a vector the angle θ clockwise about the origin. Determine $[T]$.

[Solution] We'll use equation (6.2) to find $[T]$.

Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. It is then evident that

$$T(\mathbf{e}_1) = \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix}, \quad \text{and } T(\mathbf{e}_2) = \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}.$$

Hence

$$[T] = \begin{bmatrix} \begin{bmatrix} T(\mathbf{e}_1) \end{bmatrix} \begin{bmatrix} T(\mathbf{e}_2) \end{bmatrix} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

And, we have

$$T(\begin{bmatrix} x \\ y \end{bmatrix}) = [T] \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{bmatrix}.$$

Example 6.1.11. (Dilations and contractions)

Find the standard matrix of the linear transformation that dilates a vector in $\mathbb{R}^3$ by a factor 2.

[Solution]

Example 6.1.12. Find the standard matrix for the linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$ defined by the rotation about the positive x-axis by an angle θ (anticlockwise).

(Tutorial)

Exercise 6.1.13. Let ℓ be a straight line in the xy-plane, going through the origin and making angle θ with the x-axis. Find the standard matrix for the linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ that projects a vector $\mathbf{v}$ orthogonally on ℓ (see Figure 6.1.1).

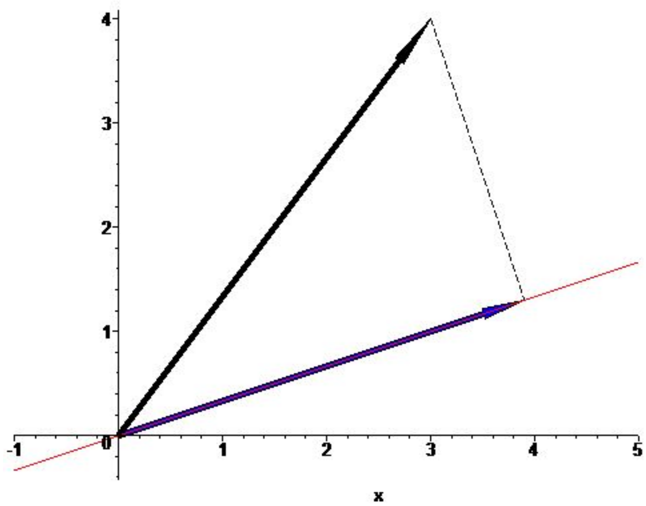

Figure 6.1.1:

6.2 Composition and Inverse Transformations

6.2.1 Composition

We will form a new linear transformation from some linear transformations.

Suppose now that $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^m \to \mathbb{R}^k$, are two linear transformations. We could then form the composition $T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^k$, defined as

$$T_2 \circ T_1(\mathbf{x}) = T_2(T_1(\mathbf{x})).$$

Question: Is $T_2 \circ T_1$ a linear transformation?

Note that

$$(T_2 \circ T_1)(\mathbf{x}) = T_2(T_1(\mathbf{x})) = T_2([T_1]\mathbf{x}) = [T_2]([T_1]\mathbf{x}) = ([T_2][T_1])\mathbf{x}.$$

Hence, noting that $[T_2]$ is an $k \times m$-matrix and $[T_1]$ an $m \times n$-matrix, we see that $T_2 \circ T_1$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^k$ whose standard matrix is the $k \times n$-matrix $[T] = [T_2][T_1]$.

Example 6.2.1. Projecting a vector $\mathbf{x} \in \mathbb{R}^2$ orthogonally on the y-axis followed by rotating the projected vector about the origin by 90^{O} in the clockwise direction is a linear transformation.

What is the standard matrix of this linear transformation?

[Solution] (Exercise.)

Remarks

(a) In general, $T_1 \circ T_2 \neq T_2 \circ T_1$. (This follows since matrix multiplication is **not** commutative.)

Can you find an example to illustrate $T_1 \circ T_2 \neq T_2 \circ T_1$?

(b) It also follows from matrix multiplication that $\circ$ is associative on linear transformations, i.e,

$$(T_1 \circ T_2) \circ T_3 = T_1 \circ (T_2 \circ T_3)$$

6.2.2 Identity Transformation

Definition 6.2.2. The **identity transformation** is the transformation $I : \mathbb{R}^n \to \mathbb{R}^n$, which sends every vector $\mathbf{x} \in \mathbb{R}^n$ to itself.

Notes

(a) I is a linear transformation and its standard matrix is the identity matrix I_n of size n. (Tutorial.)

(b) $T \circ I = T$ where $T : \mathbb{R}^n \to \mathbb{R}^m$.

(c) $I \circ S = S$ where $S : \mathbb{R}^m \to \mathbb{R}^n$.

6.2.3 Inverse Transformation

Definition 6.2.3. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be **invertible** if there is a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$S \circ T = I \quad \text{and } T \circ S = I.$$

The linear transformation S is said to be an inverse of T.

In terms of standard matrices, we have

$$[S][T] = I_n \text{ and } [T][S] = I_n.$$

Proposition 6.2.4. *(Uniqueness of Inverse) Suppose the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ invertible. Then T has exactly one inverse, which is denoted by T^{-1}.*

[Proof.] Since each linear transformation corresponds to its standard matrix, the result follows from the uniqueness of matrix inverse of $[T]$. □

Notes

(a) $T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation if and only if $[T]$ is invertible (i.e., $\det([T]) \neq 0$).

(b) The standard matrix $[T^{-1}]$ of the inverse matrix T^{-1} is given by $[T]^{-1}$, i.e.,

$$[T^{-1}] = [T]^{-1}.$$

Example 6.2.5. (Reflection about x-axis.)

The reflection R of a vector $\mathbf{x} \in \mathbb{R}^2$ about the x-axis whose standard matrix is

$$[R] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

is invertible and its inverse transformation R^{-1} is itself.

(This is geometrically obvious.)

Note: $[R]^{-1} = [R]$, $(R \circ R)(\mathbf{x}) = \mathbf{x}$.

6.3 Onto and One-to-one Linear Transformations

6.3.1 Domain and Range

Definition 6.3.1. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then $\mathbb{R}^n$ is also called the *domain* of T, $\mathbb{R}^m$ is called the *codomain* of T, and the set

$$\{\mathbf{w} \in \mathbb{R}^m : \mathbf{w} = T(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^n\},$$

is called the *range* of T.

In other words

- The domain is the set of vectors you can plug into the transformation T.
- The codomain is a set containing all the transformed vectors.
- The range is the set of transformed vectors. This may be equal to the codomain, but it can also be a smaller set contained in the codomain.

Example 6.3.2. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation that projects a vector orthogonally on the xy-plane. Here, $\mathbb{R}^3$ is both the domain and the codomain, and our geometric intuition tells us that the range consists of all vectors in the xy-plane (why?). Suppose however that we don't see this. Is there any way for us to *calculate* the range? Let's try:

We know that a vector $\mathbf{w} \in \mathbb{R}^3$ belongs to the range of T, if and only if there is a vector $\mathbf{x}$ in $\mathbb{R}^3$ such that

$$T(\mathbf{x}) = \mathbf{w}.$$

Writing this in component form, we see that $\mathbf{w} = (w_1, w_2, w_3)$ belongs to the range of T if and only if for some $\mathbf{x} = (x, y, z)$,

$$T(x, y, z) = (w_1, w_2, w_3).$$

Using that $T(x, y, z) = (x, y, 0)$, we see that $\mathbf{w} = (w_1, w_2, w_3)$ belongs to the range of T if and only if the linear system

$$\begin{aligned} x \quad\quad\quad &= w_1 \\ y \quad\; &= w_2 \\ 0 &= w_3 \end{aligned}$$

can be solved. It is however trivial to see that this system is consistent if and only if $w_3 = 0$, that is if and only if $\mathbf{w}$ lies in the xy-plane.

We will discuss a better way to determine the range of a linear transformation in the topic on vector spaces.

6.3.2 Onto

We have already noted that the range may or may not be the whole codomain. We have a special word for this:

Definition 6.3.3. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a transformation. We say that T is *surjective* or *onto*, if the range is equal to $\mathbb{R}^m$.

Example 6.3.4.

1. Rotations and reflections are onto.
2. The projection of a vector in $\mathbb{R}^3$ to the xy-plane is not onto.
3. The linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$T(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$$

   is not onto.

Theorem 6.3.5. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then the following statements are equivalent:*

(a) T is onto.

(b) The linear system $[T]\mathbf{x} = \mathbf{w}$ is consistent for any choice of $\mathbf{w} \in \mathbb{R}^n$.

Proof. Tutorial. □

Remark Now, a vector $\mathbf{w}$ in the range of T satisfies

$$\mathbf{w} = T(\mathbf{x}),$$

for some vector $\mathbf{x}$ in $\mathbb{R}^3$. Using equation (6.2) we get

$$\mathbf{w} = \begin{bmatrix} T(\mathbf{e}_1) \; T(\mathbf{e}_2) \; T(\mathbf{e}_3) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + x_3 T(\mathbf{e}_3).$$

In other words, $\mathbf{w}$ is in the range of T if and only if $\mathbf{w} = x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) + x_3 T(\mathbf{e}_3)$ for some scalars x_1, x_2 and x_3. Another way to say that is that $\mathbf{w}$ is a linear combination of the vectors $\{T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)\}$, we'll say more about linear combinations in the topic on Vectors spaces).

Now, its geometrically intuitive that if the three vectors $\{T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)\}$ are all parallel to some common plane, then any linear combination $x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + x_3T(\mathbf{e}_3)$ will also be parallel to that plane. On the other hand, if the three vectors are *not* in the same plane, then any other vector $\mathbf{w}$ in $\mathbb{R}^3$ can be represented as some linear combination of these. If that geometric picture is clear, then we have some intuitive understanding that 'T is onto if and only if $\{T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)\}$ are not all parallel to some common plane'. But the three vectors are parallel to a common plane, if and only if the parallelepiped determined by these has volume zero, that is if and only if $\det([T]) = 0$.

(Recall that the volume of the parallelepiped determined by three vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ is given by the modulus of the determinant of the matrix we get with these three vectors as columns (or rows).)

Another way to look at the determinant is that $|\det([T])|$ is the factor by which volumes are scaled by T, meaning that a region with volume V will be transformed under T to a region with volume $|\det([T])|V$.

6.3.3 One-to-one

Another important property of a transformation is whether several different vectors can have the same transformation or not. For example, if T is the projection on the xy-plane, then both the vectors $(1, 1, 1)$ and $(1, 1, 2)$ transform to $(1, 1, 0)$. On the other hand, if T is reflection about the xy-plane, then the *only* vector that transforms to (w_1, w_2, w_3), is $(w_1, w_2, -w_3)$.

Definition 6.3.6. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a transformation. We say that T is *injective* or *one-to-one*, if the implication

$$T(\mathbf{x}_1) = T(\mathbf{x}_2) \Rightarrow \mathbf{x}_1 = \mathbf{x}_2$$

holds.

You should convince yourself that the above definition is equivalent to saying that a transformation is one-to-one if two different vectors cannot transform to the same vector. In view of what was just said about the projection and reflection operators about the xy-plane, we realize that the projection is not one-to-one, while the reflection is.

The next result provides a good way to check whether a given linear transformation is one to one.

Theorem 6.3.7. *The following statements for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ are equivalent:*

(a) T is one to one.

(b) $T(\mathbf{x}) = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$.

Proof. Tutorial. □

For a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ (domain and codomain have the same dimension) we have the simple fact that T is one-to-one if and only if T is onto:

Theorem 6.3.8. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then the following statements are equivalent*

(a) T is onto.

(b) T is one-to-one.

(c) $[T]$ is invertible.

Proof. We'll prove first that (a) ⇔ (c) and then that (b)⇔(c).

Proof that (a) ⇔ (c): To say that T is onto is by definition equal to saying that for every $\mathbf{w} \in \mathbb{R}^n$, there is an $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{w}$. This in turn is equivalent to saying that the linear system $[T]\mathbf{x} = \mathbf{w}$ is consistent for any choice of $\mathbf{w} \in \mathbb{R}^n$. However, by Theorem 2.5.5, this is equivalent to $[T]$ being invertible.

Proof that (b) ⇔ (c): To say that T is one-to-one is by definition equal to saying that for $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$,

$$T(\mathbf{x}_1) = T(\mathbf{x}_2) \Rightarrow \mathbf{x}_1 = \mathbf{x}_2.$$

Writing this in matrix form, rearranging terms and using that $AB - AC = A(B - C)$, the above can be written as

$$[T](\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0} \Rightarrow \mathbf{x}_1 - \mathbf{x}_2 = \mathbf{0},$$

or compactly, with $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$:

$$[T]\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}.$$

So, we see that T being one-to-one is equivalent to the homogeneous system $[T]\mathbf{x} = \mathbf{0}$ having only the trivial solution, but we know from Theorem 2.4.8 that this is equivalent to $[T]$ being invertible. □

Example 6.3.9. Determine if the linear transformation T, from $\mathbb{R}^3$ to $\mathbb{R}^3$, defined by

$$\begin{aligned} w_1 &= -x_1 + 3x_2 + 2x_3 \\ w_2 &= 2x_1 + 4x_3 \\ w_3 &= x_1 + 3x_2 + 6x_3 \end{aligned}$$

is onto and/or one-to-one.

Solution: Theorem 6.3.8 tells us that $[T]$ is onto and one-to-one, if $\det([T]) \neq 0$ and that otherwise it is neither. Since

$$[T] = \begin{bmatrix} -1 & 3 & 2 \\ 2 & 0 & 4 \\ 1 & 3 & 6 \end{bmatrix},$$

we can evaluate the determinant. Doing that we'll get $\det([T]) = 0$, so T is neither onto nor one-to-one.

Note that it is essential in the above theorem that the domain and the codomain are the same. For example, this makes $[T]$ a square matrix and we can only talk about a matrix being invertible if it is square. Also (a) and (b) will not be equivalent if the domain and codomain are different.

Exercise 6.3.10. Determine if the linear transformation T, from $\mathbb{R}^2$ to $\mathbb{R}^3$, defined by

$$\begin{aligned} w_1 &= 2x_1 + x_2 \\ w_2 &= -x_2 \\ w_3 &= x_1 \end{aligned}$$

is onto and/or one-to-one.

(Tutorial.)

6.4 Eigenvalues and eigenvectors

Definition 6.4.1. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. We say that a nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$ is an *eigenvector* of T with eigenvalue λ, if

$$T(\mathbf{x}) = \lambda \mathbf{x}. \tag{6.3}$$

If $A = [T]$, the standard matrix of T, then the equation (6.3) can be written as

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

A nonzero vector satisfying this, is also called an eigenvector of A, with eigenvalue λ.

Example 6.4.2. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation that projects a vector on the xy-plane. We realize that if $\mathbf{x}$ already lies in the xy-plane ($\mathbf{x} = (s, t, 0)$), then $T(\mathbf{x}) = \mathbf{x}$, so $\mathbf{x}$ is an eigenvector with eigenvalue 1. On the other hand, if $\mathbf{x}$ is perpendicular to the xy-plane $\mathbf{x} = (0, 0, t)$, then $T(\mathbf{x}) = 0 = 0\mathbf{x}$ so $\mathbf{x}$ is an eigenvector with eigenvalue 0.

The above argument tells us that nonzero vectors parallel to or perpendicular to the xy-plane, are eigenvectors with eigenvalue 1 or 0 respectively. We also realize that any other vector will not be an eigenvector since transforming another vector will change its direction.

It is however not always possible to find eigenvectors by geometric reasoning like we just did. We will now look at an algebraic method of finding eigenvectors and eigenvalues of A. We start by rewriting the equation

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

as

$$(\lambda I - A)\mathbf{x} = \mathbf{0}. \tag{6.4}$$

This is a homogeneous linear system in $\mathbf{x}$, with coefficient matrix $(\lambda I - A)$. Since we said that eigenvectors are nonzero, we are looking for nontrivial solutions to this system, but Theorem 2.4.8 tells us that these exist if and only if $(\lambda I - A)$ is not invertible, i.e. if and only if

$$\det(\lambda I - A) = 0.$$

Given A, the above equation is a polynomial equation in λ and it is called the *characteristic equation* of A. Hopefully we can solve the characteristic equation, and once we've done that we can plug the different solutions λ into (6.4) and solve the corresponding linear system. The nonzero solutions will be the eigenvectors of A corresponding to that eigenvalue.

Now let's do the same example again but with our new method:

Example 6.4.3. Find the eigenvectors and eigenvalues of the transformation from the previous example.

Solution: The transformation has standard matrix

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

Since

$$\lambda I - A = \begin{bmatrix} \lambda - 1 & 0 & 0 \\ 0 & \lambda - 1 & 0 \\ 0 & 0 & \lambda \end{bmatrix},$$

the characteristic equation

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

becomes

$$\begin{vmatrix} \lambda - 1 & 0 & 0 \\ 0 & \lambda - 1 & 0 \\ 0 & 0 & \lambda \end{vmatrix} = 0,$$

i.e.

$$(\lambda - 1)^2 \lambda = 0.$$

We immediately see that the solutions are $\lambda = 1$ and $\lambda = 0$.

To find the eigenvectors corresponding to the eigenvalue $\lambda = 1$, we substitute this value into (6.4) and get the linear system

$$x_3 = 0.$$

The solution on parametric form then becomes

$$\begin{aligned} x_1 &= s, \\ x_2 &= t, \quad s, t \in \mathbb{R}, \\ x_3 &= 0. \end{aligned}$$

and the nonzero solutions will be the eigenvectors corresponding to the eigenvalue $\lambda = 1$.

To find the eigenvectors corresponding to the eigenvalue $\lambda = 0$, we substitute into (6.4) again and get

$$\begin{aligned} x_1 &= 0, \\ x_2 &= 0. \end{aligned}$$

The solution in parametric form is now

$$\begin{aligned} x_1 &= 0, \\ x_2 &= 0, \\ x_3 &= t, \quad t \in \mathbb{R}, \end{aligned}$$

and the nonzero solutions will be the eigenvectors corresponding to $\lambda = 0$.

We have recovered the same solution as the one we obtained by a geometric argument.

Example 6.4.4. Find eigenvectors and the corresponding eigenvalues of the matrix

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

Solution

Theorem 6.4.5. *(Eigenvalues of triangular matrices) If A is an $n \times n$* triangular matrix*, then the eigenvalues of A are the* entries on the main diagonal *of A.*

Proof. If A is a triangular matrix, then the matrix $\lambda I - A$ is also triangular, and we have $\det(\lambda I - A)$ is the product of diagonal entries of $(\lambda I - A)$, which is $(\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn})$.

Therefore, the eigenvalues of A are $a_{11}, a_{22}, \dots, a_{nn}$. □

Besides geometry, eigenvectors which are nonzero vectors being mapped into scalar multiples of themselves under a linear transformation arise in the study of

- Google page rank
- Physics/Engineering: vibrations, quantum mechanics
- Genetics, Population dynamics
- Economics
- Probability (Markov Chain)

Chapter 7

Vector Spaces

We have already defined vectors in $\mathbb{R}^n$ and matrices. We have also examined some basic properties. We are now going to study the fundamental structure of $\mathbb{R}^n$ and the set of $m \times n$ matrices.

In many applications in mathematics, the sciences, and engineering, the notion of a vector space arises. This idea is merely a carefully constructed generalization of $\mathbb{R}^n$. In studying the properties and structure of a vector space, we can study not only $\mathbb{R}^n$, in particular, but many other important vector spaces.

7.1 Real Vector Spaces

Many other important examples of vector spaces appear in numerous areas of mathematics. The advantage of the definition of a vector space is that it deals only with the algebraic behaviour of the elements in a vector space. Whether we view a vector in $\mathbb{R}^3$ as a point, a directed line segment or a 3×1 matrix, the algebraic behaviour is the same. These features that are common in such mathematical objects are abstracted in the notion of a vector space.

7.1.1 Definition

Definition 7.1.1. A **real vector space** $< \mathcal{V}, \oplus, \otimes >$ is a non-empty set $\mathcal{V}$ together with two operations $\oplus$ and $\otimes$ satisfying the following axioms:

1. If $\mathbf{u}$ and $\mathbf{v}$ are elements in $\mathcal{V}$, then $\mathbf{u} \oplus \mathbf{v}$ is in $\mathcal{V}$ (i.e., $\mathcal{V}$ is closed under the operation $\oplus$).

   (a) $\mathbf{u} \oplus \mathbf{v} = \mathbf{v} \oplus \mathbf{u}$ for all elements $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{V}$. ($\oplus$ is commutative)

   (b) $\mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w}) = (\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w}$ for all elements $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in $\mathcal{V}$. ($\oplus$ is associative)

   (c) There is an element $\mathbf{0}$ in $\mathcal{V}$ such that

$$\mathbf{u} \oplus \mathbf{0} = \mathbf{u}, \text{ for all } \mathbf{u} \text{ in } \mathcal{V}.$$

   (d) For each $\mathbf{u}$ in $\mathcal{V}$, there is an element $-\mathbf{u}$ in $\mathcal{V}$ such that

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

2. If $\mathbf{u}$ is any element of $\mathcal{V}$ and k is any real number, then $k \otimes \mathbf{u}$ is in $\mathcal{V}$ (i.e., $\mathcal{V}$ is closed under the operation $\otimes$).

   (a) $k \otimes (\mathbf{u} \oplus \mathbf{v}) = k \otimes \mathbf{u} \oplus k \otimes \mathbf{v}$ for all real numbers k and for all elements $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{V}$.

   (b) $(k + l) \otimes \mathbf{u} = k \otimes \mathbf{u} \oplus l \otimes \mathbf{u}$ for all real numbers k and l and all elements $\mathbf{u}$ in $\mathcal{V}$.

   (c) $k \otimes (l \otimes \mathbf{u}) = (k \cdot l) \otimes \mathbf{u}$ for all real numbers k and l and all elements $\mathbf{u}$ in $\mathcal{V}$.

   (d) $1 \otimes \mathbf{u} = \mathbf{u}$ for all elements $\mathbf{u}$ in $\mathcal{V}$.

The elements of $\mathcal{V}$ are called **vectors**; the real numbers are called **scalars**. The operation $\oplus$ is called the **vector addition**; the operation $\otimes$ is called the **scalar multiplication**. The vector $\mathbf{0}$ is called the **zero vector**. The vector $-\mathbf{u}$ is called the **negative** of $\mathbf{u}$.

7.1.2 Vector Space $\mathbb{R}^n$ and Vector Space of Matrices

Now, let us start with some familiar vector spaces.

Example 7.1.2. The set $\mathbb{R}^n$ together with the usual vector addition $+$ and usual scalar multiplication $\cdot$ is a vector space $< \mathbb{R}^n, +, \cdot >$. Sometimes, we may simply say that $\mathbb{R}^n$ is a vector space when we refer to the usual vector addition and scalar multiplication.

Example 7.1.3. The set $\mathbb{M}(m, n)$ of all $m \times n$ matrices with real entries, together with the operations matrix addition $+$ and scalar multiplication $\cdot$ is a vector space $< \mathbb{M}(m, n), +, \cdot >$.

Example 7.1.4. Consider the set $\mathcal{W}$ of all ordered triples of real numbers of the form $(0, y, z)$ and define the addition $\oplus$ as the usual vector addition and the scalar $\otimes$ as the usual scalar multiplication:

$$(0, y, z) \oplus (0, y', z') = (0, y + y', z + z')$$
$$k \otimes (0, y, z) = (0, ky, kz)$$

Is $< \mathcal{W}, \oplus, \otimes >$ a vector space?

[Solution] We shall check that whether all the axioms listed in the definition for a real vector space are satisfied for the above operations on $\mathcal{W}$.

1. Operation $\oplus$:

- Closure:

- Commutative:

- Associative:

- Zero vector for $\oplus$:

- Negative vector:

2. Operation $\otimes$:

- Closure:

(a) $k \otimes (\mathbf{u} \oplus \mathbf{v}) = k \otimes \mathbf{u} \oplus k \otimes \mathbf{v}$ for all real numbers k and for all elements $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{W}$.

(b) $(k + l) \otimes \mathbf{u} = k \otimes \mathbf{u} \oplus l \otimes \mathbf{u}$ for all real numbers k and l and all elements $\mathbf{u}$ in $\mathcal{W}$.

(c) $k \otimes (l \otimes \mathbf{u}) = (k \cdot l) \otimes \mathbf{u}$ for all real numbers k and l and all elements $\mathbf{u}$ in $\mathcal{W}$.

(d) $1 \otimes \mathbf{u} = \mathbf{u}$ for all elements $\mathbf{u}$ in $\mathcal{W}$.

Example 7.1.5. Consider the set $\mathbb{R}^3$ and define the addition $\oplus$ as the usual vector addition but the scalar multiplication $\otimes$ by

$$k \otimes (x, y, z) = (kx, y, z).$$

Is $< \mathbb{R}^3, +, \otimes >$ a vector space?

7.1.3 Zero Vector Space

Example 7.1.6. (Zero Vector Space)
Consider the set $\mathbf{V} = \{*\}$ and define the addition $\oplus$ and scalar multiplication $\otimes$ as follows:

$$* \oplus * = *$$
$$k \otimes * = *.$$

Is $< \mathbf{V}, \oplus, \otimes >$ a vector space?

Remark The zero vector of the above vector space is $*$. This vector space has only the zero vector $*$. It is thus known as the **zero vector space**. The symbol $*$ can be replaced by any other symbol you like. Very often, we simply use $\mathbf{0}$ to represent the only vector of the zero vector space.

Question Is there a real vector space with only two vectors? In other words, let $\mathbf{V} = \{0, 1\}$, can you define addition $\oplus$ and scalar multiplication $\otimes$ such that $< \mathbf{V}, \oplus, \otimes >$ is a vector space?

7.1.4 Vector Spaces of Functions, Polynomials

In many areas of mathematics, we will encounter mathematical objects with operations. We shall see that sets of real valued functions become vector spaces under the familiar addition and expected scalar multiplication of functions.

Example 7.1.7. (A vector space of real-valued functions)
Consider the set $\mathcal{V}$ of real-valued functions defined on the set $\mathbb{R}$.

For two such functions $\mathbf{f} : \mathbb{R} \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R} \to \mathbb{R}$, and $k \in \mathbb{R}$, define the addition $\oplus$ and scalar multiplication $\otimes$ as follows:

$\mathbf{f} \oplus \mathbf{g} : \mathbb{R} \to \mathbb{R}$ such that $(\mathbf{f} \oplus \mathbf{g})(x) = \mathbf{f}(x) + \mathbf{g}(x)$ (pointwise addition)

$k \otimes \mathbf{f} : \mathbb{R} \to \mathbb{R}$ such that $(k \otimes \mathbf{f})(x) = k\,(\mathbf{f}(x))$ (pointwise scalar multiplication)

Then, $< \mathcal{V}, \oplus, \otimes >$ is a vector space.

Among real-valued functions, the set of polynomials (with real coefficients) provides another important example of a real vector space.

Recall that a polynomial is a function in x that is expressed in the form

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_{n-1}, a_n$ are real numbers. If $a_n \neq 0$, then $p(x)$ is a known as a polynomial of degree n.

Note that a polynomial of degree $< n$ can always be expressed as

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where a_n (and maybe other coefficients) are zero.

Among all polynomials, there is a special polynomial whose coefficients $a_0, a_1, \ldots, a_{n-1}, a_n$ are all zero. This is the zero polynomial and it has no degree.

Example 7.1.8. Consider the set $\mathbf{P}_n$ of all polynomials of degree $\leq n$ together with the zero polynomial. We define the addition and scalar multiplication as follows:

If $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ and $q(x) = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + b_0$, then

$$p(x) \oplus q(x) = (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \cdots + (a_1 + b_n)x + (a_0 + b_0).$$

If $k \in \mathbb{R}$, then

$$k \otimes p(x) = (ka_n)x^n + (ka_{n-1})x^{n-1} + \cdots + (ka_1)x + ka_0.$$

Then $< \mathbf{P}_n, \oplus, \otimes >$ is a vector space.

(*We shall leave the verification as a tutorial question.*)

The zero polynomial is the corresponding zero vector and the negative of $p(x)$ is

$$-p(x) = -a_n x^n - a_{n-1}x^{n-1} - \cdots - a_1 x - a_0.$$

7.2 Properties

Now we are about to discuss the properties of all vector spaces without referring to any particular vector space. A 'vector' is just an element of a vector space. For example, when we refer to the vector space $\mathbf{P}_n$ in Example 7.1.8, a nonzero vector will be a polynomial of degree $\leq n$.

Proposition 7.2.1. *Suppose* $< \mathcal{V}, \oplus, \otimes >$ *is a vector space.*

1. *The zero vector* $\mathbf{0}$ *of* $\mathcal{V}$ *is unique. That is, if there is a vector* $\mathbf{0}'$ *of* $\mathcal{V}$ *satisfying Axiom 1(c), i.e.,*

$$\mathbf{u} \oplus \mathbf{0}' = \mathbf{u}, \text{ for all } \mathbf{u} \text{ in } \mathcal{V},$$

   then $\mathbf{0}' = \mathbf{0}$.

2. *Every vector* $\mathbf{u}$ *in* $\mathcal{V}$ *has a unique negative vector* $-\mathbf{u}$ *in* $\mathcal{V}$. *That is, if there is a vector* $\mathbf{u}' \in \mathcal{V}$ *satisfying Axiom 1(d), i.e.,*

$$\mathbf{u} \oplus \mathbf{u}' = \mathbf{0} = \mathbf{u}' \oplus \mathbf{u},$$

   then $\mathbf{u}' = -\mathbf{u}$.

[Proof.]

Proposition 7.2.2. *If* $< \mathcal{V}, \oplus, \otimes >$ *is a vector space, then*

1. $0 \otimes \mathbf{v} = \mathbf{0}$, *for every* $\mathbf{v} \in \mathcal{V}$.

2. $k \otimes \mathbf{0} = \mathbf{0}$, *for every* $k \in \mathbb{R}$.

3. *If* $k \otimes \mathbf{v} = \mathbf{0}$, *then* $k = 0$ *or* $\mathbf{v} = \mathbf{0}$.

4. $(-1) \otimes \mathbf{v} = -\mathbf{v}$, *for every* $\mathbf{v} \in \mathcal{V}$.

Proof. To prove these properties, we use the axioms of a vector space (Definition 7.1.1).

Proof of 1. Let $\mathbf{v} \in \mathcal{V}$. Note that

$$\begin{aligned} 0 \otimes \mathbf{v} &= (0+0) \otimes \mathbf{v} \\ &= 0 \otimes \mathbf{v} \oplus 0 \otimes \mathbf{v} \quad (\text{ use Axiom2}(b)) \end{aligned}$$

By Axiom 1(d) of Definition 7.1.1, the vector $0 \otimes \mathbf{v}$ has a negative vector $-(0 \otimes \mathbf{v})$. Adding $-(0 \otimes \mathbf{v})$ to both sides, we get

$$0 \otimes \mathbf{v} \oplus (-(0 \otimes \mathbf{v})) = (0 \otimes \mathbf{v} \oplus 0 \otimes \mathbf{v}) \oplus (-(0 \otimes \mathbf{v}))\,.$$

For the left hand side, we use Axiom 1(d) of Definition 7.1.1 to get

$$0 \otimes \mathbf{v} \oplus (-(0 \otimes \mathbf{v})) = \mathbf{0}.$$

For the right hand side, we apply the associativity (Axiom 1(b)) of $\oplus$ followed by Axioms 1(c,d), and get

$$(0 \otimes \mathbf{v} \oplus 0 \otimes \mathbf{v}) \oplus (-(0 \otimes \mathbf{v})) = 0 \otimes \mathbf{v} \oplus (0 \otimes \mathbf{v} \oplus (-(0 \otimes \mathbf{v}))\ ) = 0 \otimes \mathbf{v} \oplus \mathbf{0} = 0 \otimes \mathbf{v}.$$

Therefore, we have $\mathbf{0} = 0 \otimes \mathbf{v}$.

Proof of 2.

$\square$

7.3 Subspaces

We have seen that the vector space $\mathcal{W}$ in Example 7.1.4, which is indeed a subset of the set $\mathbb{R}^3$, under the same operations. We shall call such a subset a subspace. This is formally defined below.

Definition 7.3.1. Let $< \mathcal{V}, \oplus, \otimes >$ be a vector space and $\mathcal{W}$ be a nonempty subset of $\mathcal{V}$. If $\mathcal{W}$ is a vector space with respect to the same operations $\oplus$ and $\otimes$, then $\mathcal{W}$ is said to be a **subspace** of $\mathcal{V}$.

To avoid cumbersome notation, very often we have omitted stating explicitly the usual operations addition and scalar multiplication on $\mathbb{R}^3$ (or matrices) when we refer $\mathbb{R}^3$ (or the set of matrices) as a vector space.

Example 7.3.2. From examples discussed in the last section, we are able to list some examples of subspaces.

(a) The set $\mathcal{W}$ of all ordered triples of real numbers of the form $(0, y, z)$ is a subspace $\mathbb{R}^3$.

(b) The set $\mathbf{P}_n$ of all polynomials of degree $\leq n$ is a subspace of the set $\mathbf{P}_m$ of all polynomials of degree $\leq m$, where $m > n$.

(c) The set $\mathbf{P}_n$ of all polynomials of degree $\leq n$ is a subspace of the set $\mathcal{V}$ of real-valued functions defined on the set $\mathbb{R}$.

Recall that in Example 7.1.4, to verify $\mathcal{W}$ is a vector space we check all axioms in Definition 7.1.1 are satisfied. In doing do, we have noticed that some axioms are 'inherited' from $\mathcal{V}$, and thus they do not need to be verified valid again. Inherited axioms are: 1(a), 1(b), 2 (a), (b), (c) and (d).

(Why are Axioms 1(c) and 1 (d) not 'inherited' axioms?)

What remain to be verified are: $\mathcal{W}$ is closed under the two operations, and axioms 1(c) and 1(d) are satisfied. However, it is not difficult to see that Axioms 1(c) and 1(d) follow from the closure of $\mathcal{W}$ under addition and scalar multiplication. This is recorded in the next useful result.

Theorem 7.3.3. *If $\mathcal{W}$ is a set of one or more vectors from a vector space $\mathcal{V}$, then $\mathcal{W}$ is a subspace of $\mathcal{V}$ if and only if $\mathcal{W}$ is closed under the operations addition and scalar multiplication, i.e., the following conditions hold.*

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

(b) If $k \in \mathbb{R}$ is any scalar and $\mathbf{u} \in \mathcal{W}$, then $k \otimes \mathbf{u} \in \mathcal{W}$.

Proof. ($\Rightarrow$): This follows readily from definitions of a vector space and subspace.

($\Leftarrow$): Assume that $\mathcal{W}$ is closed under addition and scalar multiplication.

Firstly, we note the vectors in $\mathcal{W}$, being vectors in $\mathcal{V}$, will satisfy all axioms, except axioms 1(c) and 1(d).

• To verify axiom 1(c) means that we have to prove that $\mathcal{W}$ contains the zero vector, i.e., $\mathbf{0} \in \mathcal{W}$.

Given that $\mathcal{W}$ is not empty, we can choose a vector $\mathbf{u} \in \mathcal{W}$. As $< \mathcal{V}, \oplus, \otimes >$ is a vector space and $\mathbf{u} \in \mathcal{V}$, we have $\mathbf{0} = 0 \otimes \mathbf{u}$, by Proposition 7.2.2 (Part 1).

Since $\mathcal{W}$ is closed under scalar multiplication, the vector $0 \otimes \mathbf{u} \in \mathcal{W}$. Thus, $\mathbf{0} \in \mathcal{W}$.

• To verify axiom 1(d) means that we have to prove that $\mathcal{W}$ contains the negative vector $-\mathbf{u}$ for every $\mathbf{u} \in \mathcal{W}$.

Let $\mathbf{u} \in \mathcal{W}$. By Proposition 7.2.2 (Which part?), we have $-\mathbf{u} = (-1) \otimes \mathbf{u}$.

Using the fact that $\mathcal{W}$ is closed under scalar multiplication, we conclude that $-\mathbf{u} \in \mathcal{W}$.

□

Note that that if $\mathcal{W}$ is a subspace of a vector space $\mathcal{V}$, then the zero vector $\mathbf{0} \in \mathcal{W}$.

Therefore, to check whether a given subset $\mathcal{W}$ is a subspace of a vector space $\mathcal{V}$, we shall check the the following three conditions are satisfied:

(i) $\mathbf{0} \in \mathcal{W}$. (This also ensures that $\mathcal{W} \neq \phi$.)

(ii) $\mathcal{W}$ is closed under addition.

(iii) $\mathcal{W}$ is closed under scalar multiplication.

Example 7.3.4. A line through the origin of $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$.

Solution We represent the subset of points on a given line passing through the origin as

$$L = \{(x, y) \in \mathbb{R}^2 \mid ax + by = 0 \}$$

where a and b are some real constants. We shall verify that L is non-empty and it is closed under addition and scalar multiplication.

- $\mathbf{0} \in L$.

- L is closed under addition.

- L is closed under scalar multiplication.

Example 7.3.5. Is the line $L_1 = \{(x, x+1) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ a subspace of $\mathbb{R}^2$?

Solution

Example 7.3.6. Is the subset $L_2 = \{(x, 3x) \in \mathbb{R}^2 \mid x \in \mathbb{Z}\}$ a subspace of $\mathbb{R}^2$?

Solution

Example 7.3.7. Is the half plane $\mathcal{W} = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0\}$ a subspace of $\mathbb{R}^2$?

Solution

Example 7.3.8. Consider the subset $\mathcal{W}$ of 2×2 matrices where

$$\mathcal{W} = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$$

Is the subset $\mathcal{W}$ a subspace of $\mathbb{M}(2, 2)$?

Solution

- Does $\mathcal{W}$ contain the 2×2 zero matrix?

- Is $\mathcal{W}$ is closed under addition?

- Is $\mathcal{W}$ is closed under scalar multiplication?

Example 7.3.9. Is the subset $\mathcal{W}$ of 2×2 invertible matrices a subspace of $\mathbb{M}(2, 2)$?

Solution

Example 7.3.10. Consider the subset $\mathcal{W}$ of 2×2 matrices where

$$\mathcal{W} = \left\{ x \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \mid x, y \in \mathbb{R} \right\}.$$

Is the subset $\mathcal{W}$ a subspace of $\mathbb{M}(2,2)$?

Solution

- Does $\mathcal{W}$ contain the $2\times$ zero matrix $\mathbf{0}$?

- Is $\mathcal{W}$ is closed under addition?

Let $x \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \in \mathcal{W}$ and $x' \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y' \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \in \mathcal{W}$. Then

$$\begin{aligned} &\left(x \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \right) + \left(x' \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y' \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \right) \\ &= (x + x') \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + (y + y') \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \in \mathcal{W} \end{aligned}$$

- Is $\mathcal{W}$ is closed under scalar multiplication?

Let $x \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \in \mathcal{W}$ and $k \in \mathbb{R}$. Then

$$k \left(x \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \right) = (kx) \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + (ky) \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \in \mathcal{W}$$

Therefore, $\mathcal{W}$ a subspace of $\mathbb{M}(2,2)$.

Proposition 7.3.11. *Let $< \mathcal{V}, \oplus, \otimes >$ be a vector space. The subset $\mathcal{W} = \{\mathbf{0}\}$, containing only the zero vector $\mathbf{0}$, is a subspace.*

Proof. *We leave the proof as a Tutorial question.* □

The above result tells us that every vector space $< \mathcal{V}, \oplus, \otimes >$ has two subspaces, namely the zero subspace $< \{\mathbf{0}\}, \oplus, \otimes >$ and itself $< \mathcal{V}, \oplus, \otimes >$. These two subspaces are called trivial subspaces of $< \mathcal{V}, \oplus, \otimes >$.

Let us return to systems of linear equations. Let $\mathbf{A}$ be an $m \times n$ matrix and $\mathbf{b}$ an $n \times 1$ column matrix. Consider the linear system $\mathbf{Ax} = \mathbf{b}$ of m linear equations in n unknowns. A vector $\mathbf{x}_0 \in \mathbb{R}^n$ is termed a **solution vector** whenever $\mathbf{Ax}_0 = \mathbf{b}$.

Proposition 7.3.12. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix. Consider the homogeneous linear system* $\mathbf{Ax} = \mathbf{0}$ *of* m *linear equations in* n *unknowns. Then the set of solution vectors is a subspace of* $\mathbb{R}^n$.

We shall call the above subspace the **solution space** of the homogeneous system.

Proof. Let $\mathcal{W}$ be the set of solution vectors.

- $\mathbf{0} \in \mathcal{W}$:

- $\mathcal{W}$ is closed under addition:

- $\mathcal{W}$ is closed under scalar multiplication:

□

Question If $\mathbf{b} \neq \mathbf{0}$, is the solution set of the system $\mathbf{Ax} = \mathbf{b}$ of m linear equations in n unknowns a subspace of $\mathbb{R}^n$? (Tutorial Question)

Example 7.3.13. Find the solution space of the linear system:

(a)

$$\begin{bmatrix} 1 & 2 & -3 \\ -2 & -3 & 6 \\ 0 & 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Answer: $\{(0, 0, 0)\}$

(b)

$$\begin{bmatrix} 1 & 2 & -3 \\ -2 & -3 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Answer: $\{(9t, -3t, t) | t \in \mathbb{R}\}$

(c)

$$\begin{bmatrix} 1 & 2 & -3 \\ -2 & -4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Answer: $\{(-2s + 3t, s, t) | s, t \in \mathbb{R}\}$

(d)

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Answer: $\mathbb{R}^3$

7.4 Linear independence, basis and dimension

In this section we will see how we can add some structure to a vector space. For example, In $\mathbb{R}^n$ the vectors have standard coordinates which are closely related to the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ in the sense that each vector $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ can be expressed as follows:

$$(x_1, x_2, \ldots, x_n) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n.$$

We will generalize the concept of a basis in a way that allows us to consider other spaces in $\mathbb{R}^n$ (for example subspaces) and see how a basis allows us to introduce a (new) coordinate system on that space.

Before we do that, we must however examine what it is we really need from a coordinate representation. For example, we want *every* vector in our space to have a coordinate representation. The corresponding needed property of a basis will be that it *spans* our vector space and we will examine this concept in Section 7.4. The other thing we need from coordinate representations is uniqueness - a vector should only have *one* coordinate representation in a given coordinate system. The property of a basis that we need for that is *linear independence* which will be investigated in Section 7.4.3.

7.4.1 Linear Combinations

We start with some terminology.

Definition 7.4.1. Let $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a set of vectors in a vector space V. A *linear combination* of $\mathcal{S}$ is an expression of the form

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r,$$

where k_1, k_2,..., k_r, are scalars.

We say $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ if

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r,$$

for some scalars k_1, k_2,..., k_r.

Example 7.4.2.

(a) In $\mathbb{R}^3$, let $\mathbf{v}_1 = (1, 0, 2)$ and $\mathbf{v}_2 = (5, 1, -3)$. The vector $\mathbf{v} = (3, 2, -20)$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ as $\mathbf{v} = -7\mathbf{v}_1 + 2\mathbf{v}_2$.

(b) In $\mathbb{M}(2, 3)$, let $\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}$, $\mathbf{A}_2 = \begin{bmatrix} -1 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix}$ and $\mathbf{A}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. The matrix $\mathbf{A} = \begin{bmatrix} 5 & 0 & 7 \\ -2 & 3 & 0 \end{bmatrix}$ is a linear combination of $\mathbf{A}_1, \mathbf{A}_2$ and $\mathbf{A}_3$ since we have $\mathbf{A} = 3\mathbf{A}_1 - 2\mathbf{A}_2 + 5\mathbf{A}_3$.

Example 7.4.3. In $\mathbb{R}^3$, let $\mathbf{v}_1 = (1, 3, 2)$ and $\mathbf{v}_2 = (-1, 0, 1)$.

(a) Is $(1, 9, 8)$ a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$?

(b) Is $(0, 0, 1)$ a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$?

7.4.2 Span

Now, consider the set of *all* linear combinations of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$.

Definition 7.4.4. Let $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a set of vectors in a vector space. The *set spanned by* $\mathcal{S}$ or simply the *span of* $\mathcal{S}$ is the set of all linear combinations of $\mathcal{S}$ and it is denoted span$\mathcal{S}$. In other words, $\mathbf{v} \in \text{span}\mathcal{S}$ if and only if there are scalars k_1, k_2,..., k_r, such that

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r.$$

Example 7.4.5. $\mathbb{R}^n$ is spanned by $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$. We write

$$\mathbb{R}^n = \text{span}\ \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}.$$

Example 7.4.6. Determine if $\mathcal{S} = \{(1, 2, 3), (1, 1, -1), (-1, 0, 5), (-1, 2, 13)\}$ spans $\mathbb{R}^3$. If it doesn't, describe the set spanned by $\mathcal{S}$.

Example 7.4.7. Consider the subspace $\mathcal{W}$ of $\mathbb{M}(2, 2)$ in Example 7.3.8, where

$$\mathcal{W} = \left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \mid a, b \in \mathbb{R} \right\}.$$

We note that

$$\mathcal{W} = \text{span} \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Example 7.4.8. In Example 7.3.13 (c), the solution space of the homogeneous system

$$\begin{bmatrix} 1 & 2 & -3 \\ -2 & -4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is the set $\mathcal{W} = \left\{ \begin{bmatrix} -2s+3t \\ s \\ t \end{bmatrix} | s,t \in \mathbb{R} \right\}$. Note that for each $s, t \in \mathbb{R}$, we have

$$\begin{bmatrix} -2s+3t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}.$$

Thus, $\mathcal{W} = \text{span} \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Theorem 7.4.9. *If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of vectors in a vector space $\mathcal{V}$, then $W = \text{span}\mathcal{S}$ is a subspace of $\mathcal{V}$.*

Proof. Tutorial problem. □

Example 7.4.10. Consider the subspace $\mathcal{W}$ of 2×2 matrices in Example 7.3.10 where

$$\mathcal{W} = \left\{ x \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix} + y \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \mid x, y \in \mathbb{R} \right\}.$$

We note that $\mathcal{W} = \text{span} \left\{ \begin{bmatrix} 1 & 3 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 5 & 1 \end{bmatrix} \right\}$.

Proposition 7.4.11. *If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a set of n vectors in $\mathbb{R}^n$, then $\text{span}\mathcal{S} = \mathbb{R}^n$ if and only if*

$$\left| \begin{bmatrix} \mathbf{v}_1 \; \mathbf{v}_2 \; \cdots \; \mathbf{v}_n \end{bmatrix} \right| \neq 0,$$

where $\begin{bmatrix} \mathbf{v}_1 \; \mathbf{v}_2 \; \cdots \; \mathbf{v}_n \end{bmatrix}$ is the matrix whose ith column is the column vector $\mathbf{v}_i$.

Proof. Exercise. □

7.4.3 Linear independence

Definition 7.4.12. The set $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ consisting of vectors in a vector space, is said to be *linearly independent* if the system

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r = \mathbf{0},$$

only has the trivial solution $k_1 = k_2 = \ldots = k_r = 0$. Otherwise, $\mathcal{S}$ is said to be *linearly dependent*.

Example 7.4.13. Consider the vector space $\mathbb{R}^4$, determine if $\{(1,2,3,4),(1,-1,2,2),(-2,5,-3,-2)\}$ is linearly dependent or linearly independent.

Example 7.4.14. Consider the vector space $\mathbb{R}^n$, the set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is linearly independent.

Example 7.4.15. Consider the vector space $\mathbb{M}(2,3)$, let

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} -1 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{A}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Is the set $S = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$ linearly independent?

From the solution to the example above, we may realize how to prove the following proposition.

Proposition 7.4.16. *If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a set of n vectors in $\mathbb{R}^n$, then $\mathcal{S}$ is linearly independent if and only if*

$$\left| \begin{bmatrix} \mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n \end{bmatrix} \right| \neq 0.$$

Proof. Exercise. □

Another useful way of looking at linearly independent/dependent vectors, is given by the following theorem

Theorem 7.4.17. *A set of two or more vectors is linearly dependent if and only if (at least) one of the vectors is a linear combination of the others.*

Proof. Tutorial problem. □

Remark We realize some useful facts which follow from the above theorem or definition.

- For a vector space $\mathcal{V}$, if a set S contains the zero vector, then S is linearly dependent.
- For a vector space $\mathcal{V}$, if a set $S = \{\mathbf{v}\}$ contains only a non-zero vector $\mathbf{v}$, then S is linearly independent.
- A set of *two* vectors, is linearly dependent if and only if one of the vectors is a scalar multiple of the other.

  In $\mathbb{R}^2$ or $\mathbb{R}^3$, this means that a set of two nonzero vectors is linearly dependent if and only if they are parallel to some common line.
- For a set of three vectors in $\mathbb{R}^3$, one vector is a linear combination of the other two, if and only if it lies in the plane spanned by these. Hence, three vectors in $\mathbb{R}^3$ are linearly dependent if and only if they all are parallel to some common plane.

Linear independence of functions

Definition 7.4.18. Suppose I is an open interval. With $C^n(I)$ we mean the set of all functions $f : I \to \mathbb{R}$ which have continuous derivatives on I up to and including order n.

In a tutorial problem we have seen that $C^n(I)$ is a subspace of the space $F(I)$ consisting of all real valued functions on I.

In general it can be difficult to figure out whether a set of functions in $C^n(I)$ is linearly independent or not, but the following tool can be useful.

Definition 7.4.19. Let $f_1, f_2, \ldots, f_n$ be a collection of functions in $C^{n-1}(I)$. With the *Wronsikan*, $W(f_1, f_2, \ldots, f_n)(x)$, we mean the determinant

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

Note that, given $f_1, f_2, \ldots, f_n$, $W(f_1, f_2, \ldots, f_n)(x)$ is a real valued function of x. If this function is not indentically zero, then the functions form a linearly independent set.

Theorem 7.4.20. *Let I be an open interval and $S = \{f_1, f_2, \ldots, f_n\}$ a set of functions in $C^{n-1}(I)$. If the Wronskian $W(f_1, f_2, \ldots, f_n)(x)$ is not identically zero on I, then S is a linearly independent set in $C^{n-1}(I)$.*

Example 7.4.21. Show that $f_1(x) = 1$, $f_2(x) = \sin x$, $f_3(x) = \cos x$, forms a linearly independent set in $C^2(infty, infty)$.

Done during lecture

Note: Dont confuse Theorem 7.4.20 with it's converse. If the wronskian is identically zero, we can *not* conclude that the set is linearly dependent. It might be linearly dependent or it might be linearly independent. The theorem gives us no information and we have to use some other way to find out which.

7.4.4 Basis

In this section we shall see how the concepts of spanning and linear independence provide us with a way of introducing (new) coordinates on a vector space. This will also help us to make a proper definition of the dimension of a space.

Sets of vectors that both span a given space and are linearly dependent, are so important for this theory that they deserve a special name.

Definition 7.4.22. Let $\mathcal{V}$ be a vector space and $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ a set of vectors in $\mathcal{V}$. If $\mathcal{B}$ spans $\mathcal{V}$ and is linearly independent, then $\mathcal{B}$ is said to be a *basis* for $\mathcal{V}$.

Example 7.4.23. Standard basis of $\mathbb{R}^n$.

The simplest possible example of a basis would be the standard basis

$$\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\} = \{(1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0), (0, 0, 0, \ldots, 1)\},$$

of $\mathbb{R}^n$. It is quite obvious (why?) that $\mathcal{B}$ spans $\mathbb{R}^n$ and is linearly independent, so $\mathcal{B}$ is a basis for $\mathbb{R}^n$.

Example 7.4.24. The set

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

is a basis of $\mathbb{M}(2, 3)$.

Example 7.4.25. Show that $\mathcal{B} = \{(1, 1), (0, 2)\}$ is a basis for $\mathbb{R}^2$.

[Solution] We need to show that $\mathcal{B}$ spans $\mathbb{R}^2$ and that $\mathcal{B}$ is linearly independent.

The above examples tell us that a vector space can have more than one bases. The following are bases for $\mathbb{R}^2$:

(a) $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$,

(b) $\mathcal{B}_1 = \{(1, 1), (0, 2)\}$,

(c) $\mathcal{B}_2 = \{(\pi, 0), (2/7, \sqrt{2})\}$.

Example 7.4.26. Consider the following 2×3 matrices

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} -1 & 0 & 2 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } \mathbf{A}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

In Example 7.4.15, we have shown that the set $S = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\}$ is linearly independent in $\mathbb{M}(2,3)$. Now, by the definition of spanning set span S, we see that the set S is a basis for the subspace span S.

Example 7.4.27. Consider the solution space $\mathcal{W} = \left\{ \begin{bmatrix} -2s+3t \\ s \\ t \end{bmatrix} | s, t \in \mathbb{R} \right\}$ in Example 7.4.8, the set $\mathcal{B} = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\}$ (which contains 2 vectors) is linearly independent since neither vector is a scalar multiple of the other. Therefore, $\mathcal{B}$ is a basis of $\mathcal{W}$.

Example 7.4.28. Let W be the subspace of $\mathbb{R}^3$ consisting of the plane

$$x + 3y - z = 0.$$

Find a basis for W.

(Tutorial.)

7.4.5 Coordinate Vectors

It also comes as no surprise, that each vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ can be written as a linear combination of the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ in exactly one way. The (unique) coefficients are x_1, $x_2, \ldots, x_n$. Note that these coefficients are also known as the coordinates of $\mathbf{x}$ in a coordinate system where $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ are unit vectors (vectors of norm one) pointing along the coordinate axes. This inspires the definition of coordinate vector with respect to a basis. For the definition to make sense, we need the property of uniqueness of coefficients in a linear combination of basis vectors. This important property of a basis is captured by the following theorem.

Theorem 7.4.29. *Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a basis for the vector space $\mathcal{V}$. Then, for every vector $\mathbf{v}$ in $\mathcal{V}$, there exist* unique *scalars k_1, k_2,...,k_r, such that*

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r.$$

Another way to state the theorem above, is that if $\mathcal{B}$ is a basis for $\mathcal{V}$, then each vector in $\mathcal{V}$ can be written as a linear combination of the vectors in $\mathcal{B}$ in *exactly one* way. The proof of the theorem is given below. Note that we simply use the definitions we've learned.

Proof. Since $\mathcal{B}$ is a basis for $\mathcal{V}$ we know that $\mathcal{B}$ spans $\mathcal{V}$ and that $\mathcal{B}$ is linearly independent.

Since $\mathcal{B}$ spans $\mathcal{V}$ and since $\mathbf{v} \in \mathcal{V}$, there exist scalars k_1, k_2,...,k_r, such that

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r. \tag{7.1}$$

It remains to show that the scalars k_1, k_2,...,k_r are unique. To do this we assume that there also exist scalars c_1, c_2,..., c_r such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_r\mathbf{v}_r, \tag{7.2}$$

and our goal is then to show that $c_1 = k_1$, $c_2 = k_2$,..., $c_r = k_r$. For this we use linear independence. Note that subtracting equations (7.1) and (7.2) from each other, gives us

$$(k_1 - c_1)\mathbf{v}_1 + (k_2 - c_2)\mathbf{v}_2 + \ldots + (k_r - c_r)\mathbf{v}_r = \mathbf{0}.$$

Since $\mathcal{B}$ is linearly independent, this means that

$$k_1 - c_1 = k_2 - c_2 = \ldots = k_r - c_r = 0$$

and the uniqueness follows. □

Definition 7.4.30. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a basis for the vector space V and let $\mathbf{v}$ be a vector in V. The *coordinate vector* of V, relative to $\mathcal{B}$ is denoted $(\mathbf{v})_{\mathcal{B}}$ and is given by

$$(\mathbf{v})_{\mathcal{B}} = (k_1, k_2, \ldots, k_r),$$

where

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r.$$

Note that Theorem 7.4.29 guarantees that the coordinate vector exists and is unique. Also note that the coordinate vector $(\mathbf{v})_{\mathcal{B}}$ above, is a vector in $\mathbb{R}^r$.

Example 7.4.31. In Example 7.4.25, we have shown that $\mathcal{B} = \{(1, 1), (0, 2)\}$ is a basis for $\mathbb{R}^2$. Find $(\mathbf{v})_{\mathcal{B}}$ where $\mathbf{v} = (2, 3)$.

Example 7.4.32. In a previous example we have found a basis $\mathcal{B}$ for the subspace W of $\mathbb{R}^3$ consisting of the plane

$$x + 3y - z = 0.$$

Verify that $\mathbf{v} = (1, 2, 7) \in W$ and find $(\mathbf{v})_B$.

Let us also note that we can perform arithmetic on the coordinate vectors instead of the original vectors if we prefer.

Proposition 7.4.33. *Let $\mathcal{B}$ be a basis for the vector space $\mathcal{V}$. Then*

1. $(\mathbf{u} + \mathbf{v})_{\mathcal{B}} = (\mathbf{u})_{\mathcal{B}} + (\mathbf{v})_{\mathcal{B}}$ *for all* $\mathbf{u}, \mathbf{v} \in V$.
2. $(k\mathbf{v})_{\mathcal{B}} = k(\mathbf{v})_{\mathcal{B}}$ *for all* $\mathbf{v} \in \mathcal{V}$ *and all scalars* k.

Proof. Tutorial problem. □

7.4.6 Dimension

Definition 7.4.34. Let $\mathcal{V}$ be a vector space. If there exists a basis for $\mathcal{V}$ consisting of a finite number of vectors, we say that $\mathcal{V}$ is *finite dimensional*. Otherwise we say that $\mathcal{V}$ is *infinite dimensional*.

A basis is also the key to defining the dimension of a vector space. Intuitively, we can probably guess that a basis of a vector space always has a fixed number of vector (a basis of a space that forms a plane always has two vectors, a basis of a space that forms a line always consists of one vector) and then it would be natural to define the dimension of the space as the number of vectors in a basis. However, before we do that we must *know* that a basis for a given vector space always has the same number of vectors. We will record this in a proposition and prove it.

Proposition 7.4.35. *Let $\mathcal{V}$ be a finite dimensional vector space. Then all bases for $\mathcal{V}$ will consist of the same number of vectors.*

In proving the proposition, we will make use of several results that have independent interest. We will formulate these as lemmas.

Lemma 7.4.36. *A homogeneous linear system with more unknowns than equations, has infinitely many solutions.*

Proof. The augmented matrix to a linear system with more unknowns than equations, has more columns than rows. This means that after reduction to reduces row echelon form, there will be more columns than leading ones. Since a homogeneous system is always consistent, we get a general solution containing at least one parameter, which means that we have infinitely many solutions. □

Lemma 7.4.37. *If $\mathcal{V}$ is a vector space with a basis consisting of n vectors, then any set of vectors in $\mathcal{V}$ consisting of more than n vectors, is linearly dependent.*

Proof. Assume that $\mathcal{V}$ has the basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and let $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a set of vectors in $\mathcal{V}$ with $m > n$. We wish to prove that $\mathcal{S}$ is linearly dependent, and to do that we want to show that the equation

$$k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \ldots + k_m\mathbf{u}_m = \mathbf{0}, \tag{7.3}$$

has nontrivial solutions. By Proposition 7.4.33, the above equation is equivalent to

$$k_1(\mathbf{u}_1)_{\mathcal{B}} + k_2(\mathbf{u}_2)_{\mathcal{B}} + \ldots + k_m(\mathbf{u}_m)_{\mathcal{B}} = (\mathbf{0})_{\mathcal{B}}. \tag{7.4}$$

If

$$\begin{aligned}
(\mathbf{u}_1)_{\mathcal{B}} &= (u_{11}, u_{12}, \ldots, u_{1n}),\\
(\mathbf{u}_2)_{\mathcal{B}} &= (u_{21}, u_{22}, \ldots, u_{2n}),\\
&\vdots\\
(\mathbf{u}_m)_{\mathcal{B}} &= (u_{m1}, u_{m2}, \ldots, u_{mn}),
\end{aligned}$$

we see that the equation (7.4) turns into the linear system

$$\begin{aligned}
u_{11}k_1 + u_{21}k_2 + \ldots + u_{m1}k_m &= 0,\\
u_{12}k_1 + u_{22}k_2 + \ldots + u_{m2}k_m &= 0,\\
&\vdots\\
u_{1n}k_1 + u_{2n}k_2 + \ldots + u_{mn}k_m &= 0.
\end{aligned}$$

Since $m > n$, Lemma 7.4.36 above, tells us that the system has infinitely may solutions. Hence there exist nontrivial solutions to (7.3), so $\mathcal{S}$ is linearly dependent. □

And now we can prove the proposition.

Proof of Proposition 7.4.35. We use proof by contradiction. Suppose we have two different bases for V which don't have the same number of vectors. Then one basis consists of a larger number of vectors than the other, and by Lemma 7.4.37, that is a linearly dependent set. However, a basis is by definition linearly independent, so we have a contradiction. □

Now we can safely define the dimension of a space.

Definition 7.4.38. The *dimension* of a finite dimensional vector space, is the number of vectors in a basis for that space.

Example 7.4.39.

(a) The dimension of $\mathbb{R}^n$ is n.

(b) The dimension of $\mathbb{M}(m, n)$ is mn.

(c) The dimension of the vector space of polynomials of degree at most n is $n + 1$.

Chapter 8

Rank and Nullity

8.1 Null space, row space and column space

An $m \times n$ matrix A, will give us some interesting subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$. One of them is the solution space of the homogeneous system $A\mathbf{x} = \mathbf{0}$ which is a subspace of $\mathbb{R}^n$ (Proposition 7.3.12). The solution space also has another name.

8.1.1 Nullspace

Definition 8.1.1. Let A be an $m \times n$ matrix. The set of solutions to $A\mathbf{x} = \mathbf{0}$ (which is a subspace of $\mathbb{R}^n$), is called the *null space* of A. We denote it by null($\mathbf{A}$).

The dimension of the null($\mathbf{A}$) is called the *nullity* of $\mathbf{A}$, and is denoted by nullity($\mathbf{A}$).

Example 8.1.2. Find a basis for the null space of

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

Solution: Solving the homogeneous system $A\mathbf{x} = \mathbf{0}$ we get the augmented matrix

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 0 \\ 0 & 3 & 1 & -1 & 0 \\ 2 & 1 & 5 & 9 & 0 \end{bmatrix},$$

which after row reduction becomes

$$\begin{bmatrix} 1 & 0 & 7/3 & 14/3 & 0 \\ 0 & 1 & 1/3 & -1/3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The solution in parametric form is

$$\begin{aligned}
x_1 &= -\frac{7}{3}s - \frac{14}{3}t = -7s' - 14t', \\
x_2 &= -\frac{1}{3}s + \frac{1}{3}t = -s + t, \\
x_3 &= s = 3s', \\
x_4 &= t = 3t',
\end{aligned}$$

where $s = 3s'$, $t = 3t'$, or with column vector notation:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = s' \begin{bmatrix} -7 \\ -1 \\ 3 \\ 0 \end{bmatrix} + t' \begin{bmatrix} -14 \\ 1 \\ 0 \\ 3 \end{bmatrix}.$$

From this we see that all solutions are linear combinations of the two vectors $\mathbf{u}_1 = (-7, -1, 3, 0)$ and $\mathbf{u}_2 = (-14, 1, 0, 3)$, so $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2\}$ spans the null space. Furthermore, it is evident from the last two components of $\mathbf{u}_1$ and $\mathbf{u}_2$ that $\mathcal{B}$ is linearly independent (why?). Hence, $\mathcal{B}$ is a basis for the null space of A. The nullity of $\mathbf{A}$ is 2, nullity$(\mathbf{A}) = 2$.

8.1.2 Column space and row space

The column vectors and row vectors of a matrix also span spaces of special interest.

Definition 8.1.3. Let A be an $m \times n$ matrix with column vectors $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ and row vectors $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$. Then the subspace of $\mathbb{R}^m$ spanned by $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ is called the *column space* of A, and the subspace of $\mathbb{R}^n$ spanned by $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$ is called the *row space* of A. We will denote the column space and row space of A by $C(A)$ and $R(A)$ respectively.

The row space and column space are interesting because of how they are affected by elementary row operations on A.

Proposition 8.1.4. *If A and B are row equivalent, then $R(A) = R(B)$.*

The proof is omitted. You can look it up in a linear algebra textbook if you like.

The above result provides a simple way to obtain a basis of $R(A)$, as illustrated in the following example.

Example 8.1.5. Find a basis for the subspace W of $\mathbb{R}^4$ spanned by

$$\{(1, 2, 3, 4), (0, 3, 1, -1), (2, 1, 5, 9)\}.$$

Solution: If put these vectors as rows in a matrix $\mathbf{A}$,

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 1 & -1 \\ 2 & 1 & 5 & 9 \end{bmatrix},$$

we have $W = R(\mathbf{A})$. After row reduction, we get the matrix

$$B = \begin{bmatrix} 1 & 0 & 7/3 & 14/3 \\ 0 & 1 & 1/3 & -1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

It is clear (why?) that the first two row vectors $\{(1, 0, 7/3, 14/3), (0, 1, 1/3, -1/3)\}$ are linearly independent and span $R(B)$, hence they form a basis for $R(B)$. But since $R(B) = R(A) = W$, $\{(1, 0, 7/3, 14/3), (0, 1, 1/3, -1/3)\}$ is a basis for W.

For the column space, things are a little bit different. Row operation will change the column space, for example it is clear that the two matrices

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix},$$

are row equivalent but have different column spaces. However, row operations will preserve the relations between columns in a sense best described by an example.

Example 8.1.6. Let

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

As we have seen in previous examples, the reduced row echelon form of A is

$$B = \begin{bmatrix} 1 & 0 & 7/3 & 14/3 \\ 0 & 1 & 1/3 & -1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let us denote the column vectors of A, by $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$ and $\mathbf{c}_4$, and the column vectors of B, by $\mathbf{c}'_1$, $\mathbf{c}'_2$, $\mathbf{c}'_3$ and $\mathbf{c}'_4$. It is easy to see (why?) that $\{\mathbf{c}'_1, \mathbf{c}'_2\}$ is a basis for $C(B)$. Furthermore, it is also easy to see that

$$\mathbf{c}'_3 = \frac{7}{3}\mathbf{c}'_1 + \frac{1}{3}\mathbf{c}'_2, \quad \text{and } \mathbf{c}'_4 = \frac{14}{3}\mathbf{c}'_1 - \frac{1}{3}\mathbf{c}'_2.$$

The interesting thing here, is that we have exactly the same relations for the columns of A, namely

$$\mathbf{c}_3 = \frac{7}{3}\mathbf{c}_1 + \frac{1}{3}\mathbf{c}_2, \quad \text{and } \mathbf{c}_4 = \frac{14}{3}\mathbf{c}_1 - \frac{1}{3}\mathbf{c}_2.$$

You should check by direct calculation that the above is actually true.

In view of the observations above, the following result may be expected.

Proposition 8.1.7. *Suppose* $\mathbf{A}$ *and* $\mathbf{B}$ *are row equivalent. Then column vectors of* $\mathbf{A}$ *are linearly independent if and if column vectors of* $\mathbf{B}$ *are linearly independent.*

Proof. Column vectors of $\mathbf{A}$ are linearly independent

$\iff$ $\mathbf{Ax} = \mathbf{0}$ has only trivial solution.

$\iff$ $\mathbf{Bx} = \mathbf{0}$ has only trivial solution, since $\mathbf{A}$ and $\mathbf{B}$ are row equivalent.

$\iff$ Column vectors of $\mathbf{B}$ are linearly independent.

□

As a consequence of the above proposition, we have

Corollary 8.1.8. *Suppose* $\mathbf{R}$ *is a row echelon form of* $\mathbf{A}$*. Then column vectors of* $\mathbf{A}$ *that correspond to column vectors with leading 1 in* $\mathbf{R}$ *are linearly independent.*

We record the following result which is a summary of the above discussion.

Theorem 8.1.9. *Suppose* $\mathbf{R}$ *is a row echelon form of* $\mathbf{A}$*. Then*

(i) column vectors of $\mathbf{A}$ *that correspond to column vectors with leading 1 in* $\mathbf{R}$ *forms a basis for the column space of* $\mathbf{A}$*.*

(ii) nonzero row vectors of $\mathbf{R}$ *form a basis of the row space of* $\mathbf{A}$*.*

Example 8.1.10. Let $\mathbf{v}_1 = (1, 3, 1)$, $\mathbf{v}_2 = (0, 4, 2)$ and $\mathbf{v}_3 = (-1, 1, 1)$. Find a basis $\mathcal{B}$ for the space W spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Also, determine if $\mathbf{v} = (-2, 6, 4)$ is in W. If yes, find $(\mathbf{v})_{\mathcal{B}}$.

8.2 Rank-nullity Formula

8.2.1 Rank of a matrix

Being subspaces of $\mathbb{R}^n$ or $\mathbb{R}^m$, the row space $R(\mathbf{A})$ and the column space $C(\mathbf{A})$ of an $m \times n$ matrix $\mathbf{A}$ are finite dimensional.

Definition 8.2.1. Let $\mathbf{A}$ be an $m \times n$ matrix.

(a) The dimension of the row space $R(\mathbf{A})$ of $\mathbf{A}$ is called *row rank* of $\mathbf{A}$.

(b) The dimension of the column space $C(\mathbf{A})$ of $\mathbf{A}$ is called *column rank* of $\mathbf{A}$.

Proposition 8.2.2. *Let* $\mathbf{A}$ *be a matrix. Then the row rank of* $\mathbf{A}$ *equals the column rank of* $\mathbf{A}$*.*

Proof. Suppose $\mathbf{R}$ is a row echelon form of $\mathbf{A}$. The row space of $\mathbf{A}$ and $\mathbf{R}$ are the same. The nonzero row vectors of R form a basis for the row space of R. Thus, the row rank of $\mathbf{A}$ is the number of nonzero rows in $\mathbf{R}$. This number is the same as the number of leading 1's in $\mathbf{R}$.

On the other hand, using Theorem 8.1.9, the column rank of $\mathbf{A}$ is the number of leading 1's in $\mathbf{R}$.

So, the row rank of $\mathbf{A}$ and the column rank of $\mathbf{A}$ are equal. □

The common number is known as the *rank* of $\mathbf{A}$, denoted by $\operatorname{rank}(\mathbf{A})$.

Example 8.2.3. Determine the rank of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 1 & -1 \\ 2 & 1 & 5 & 9 \end{bmatrix}$$

.

Solution

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 1 & -1 \\ 2 & 1 & 5 & 9 \end{bmatrix} \sim \mathbf{R} = \begin{bmatrix} 1 & 0 & 7/3 & 14/3 \\ 0 & 1 & 1/3 & -1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, $\operatorname{rank}(\mathbf{A}) = 2$.

8.2.2 Rank-nullity Formula

Is there any relationship between the null space, the row space and the column space of a matrix **A**?

Since the vector **Ax** is a linear combination of the columns of **A** (with $x_1, \ldots, x_n$ as coefficients), the definition of column space immediately gives us the following:

Proposition 8.2.4. *A linear system* $\mathbf{Ax} = \mathbf{b}$ *is consistent if and only if* **b** *is in the column space of* **A**.

Recall that we may treat a matrix **A** as a linear transformation T, and that a vector **b** is in the range of T if and only if $\mathbf{Ax} = \mathbf{b}$ is consistent. That brings us to the following relationship.

Proposition 8.2.5. *Suppose* **A** *is an* $m \times n$ *matrix. Let* $T : \mathbb{R}^n \to \mathbb{R}^m$ *be the linear transformation with standard matrix* **A**. *Then the range of* T *is the column space of* **A**.

In one of our tutorials, we have proved the following about the general solution to a linear system.

Proposition 8.2.6. *For a given linear system* $\mathbf{Ax} = \mathbf{b}$, *suppose* $\mathbf{y}_0$ *is a solution of* $\mathbf{Ax} = \mathbf{b}$. *Then* **y** *is a solution* $\mathbf{Ax} = \mathbf{b}$ *if and only if*

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{y}_h$$

where $\mathbf{y}_h$ *is some solution of the homogeneous system* $\mathbf{Ax} = \mathbf{0}$.

Using the terminology introduced in this section, $\mathbf{y}_h$ lies in the null space of **A**. Since null(**A**) is finite dimensional, we can find a basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ of null(A).

Thus, the general solution of $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{y} = \mathbf{y}_0 + k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_p\mathbf{u}_p, \text{ where } k_1, k_2, \ldots, k_p \in \mathbb{R}$$

Example 8.2.7. Consider the linear system $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 3 & 1 & -1 \\ 2 & 1 & 5 & 9 \end{bmatrix} \& \mathbf{b} = \begin{bmatrix} 4 \\ 3 \\ 5 \end{bmatrix}.$$

Note that $\mathbf{y}_0 = (2, 1, 0, 0)$ is a solution of $\mathbf{Ax} = \mathbf{b}$.

The reduced row echelon form **R** of **A** is

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 7/3 & 14/3 \\ 0 & 1 & 1/3 & -1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From Example 8.1.2, a basis of null(A) is $\mathcal{B} = \{(-7, -1, 3, 0), (-14, 1, 0, 3)\}$.

Thus, the general solution $\mathbf{y}$ of $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{y} = (2, 1, 0, 0) + (-7, -1, 3, 0)s + (-14, 1, 0, 3)t, \text{where } s, t \in R$$

In column vector notation,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -7 \\ -1 \\ 3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -14 \\ 1 \\ 0 \\ 3 \end{bmatrix}.$$

Remark The nullity of $\mathbf{A}$ is the number of free parameters in the general solution of the homogeneous system $\mathbf{Ax} = \mathbf{0}$.

We conclude this section with the following result depicting a relationship between dimensions of null space, row space and column space of a matrix $\mathbf{A}$.

Theorem 8.2.8. *Let A be an $m \times n$ matrix. Then*

$$\mathit{rank}(\mathbf{A}) + \mathit{nullity}(\mathbf{A}) = n.$$

Proof. The homogeneous system $\mathbf{Ax} = \mathbf{0}$ contains n unknowns and m equations. Consider the reduced row echelon form $\mathbf{A}$. The number of free parameters is $n - \text{rank}(\mathbf{A})$.

On the other hand, the nullity is the number of free parameters. Thus, we have

$$\text{nullity}(\mathbf{A}) = n - \text{rank}(\mathbf{A}).$$

□

Chapter 9

Inner product spaces

9.1 Inner products

In a general vector space, we can add vectors and multiply with scalars, but we have so far nothing that corresponds to the scalar product of vectors. In this section we will define what is meant by an inner product, which will also provide us with a way of introducing a norm of vectors.

Definition 9.1.1. Suppose $\mathcal{V}$ is a vector space. An *inner product* on $\mathbf{v}$ is an operation that to each pair of vectors $\mathbf{u}, \mathbf{v} \in \mathcal{V}$, associates a scalar $\langle \mathbf{u}, \mathbf{v} \rangle$ and for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$ and all scalars k, satisfies:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$.
3. $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$.
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

A vector space equipped with an inner product, is called an *inner product space*.

Some other properties follow from this definition.

Example 9.1.2. If $\mathcal{V}$ is an inner product space with inner product $\langle ., . \rangle$, then $\langle \mathbf{0}, \mathbf{v} \rangle = 0$ for every $\mathbf{v} \in \mathcal{V}$. To see this, we observe that by the property of the zero vector and using property 2 above, we have

$$\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{0} + \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{0}, \mathbf{v} \rangle + \langle \mathbf{0}, \mathbf{v} \rangle.$$

This is just an equation with scalars, and adding the scalar $-\langle \mathbf{0}, \mathbf{v} \rangle$ to both sides, we get

$$0 = \langle \mathbf{0}, \mathbf{v} \rangle.$$

Definition 9.1.3. Suppose $\mathcal{V}$ is an inner product space with inner product $\langle .,. \rangle$. The *norm* $\|\mathbf{u}\|$ of a vector $\mathbf{u} \in \mathcal{V}$ is then defined as

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Example 9.1.4. If in the vector space $\mathbb{R}^n$, we take $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y}$, then the properties 1–4 listed in the definition of an inner product, are satisfied since these are exactly the properties (a)–(d) listed in Theorem 4.3.2. Hence, the scalar product is an inner product, also referred to as *the Euclidean inner product*.

Example 9.1.5. Weighted Euclidean inner products.

Example 9.1.6. Let $C([a, b])$ denote the set of real valued continuous fuctions on the interval $[a, b]$. It is straightforward to verify that with pointwise addition and scalar multiplication, this is a vector space. Also, with

$$\langle f, g \rangle = \int_a^b f(x)g(x)\,dx,$$

we will verify (during lecture) that $\langle .,. \rangle$ is an inner product on $[a, b]$.

In particular, the norm $\|f\|$ of a function in $C([a, b])$ is given by

$$\|f\| = \sqrt{\langle f, f \rangle} = \left(\int_a^b (f(x))^2\,dx \right)^{1/2}.$$

One important result is the following.

Theorem 9.1.7 (Cauchy-Schwarz inequality)**.** *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors in an inner product space* $\mathcal{V}$*. Then*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|.$$

Proof. If $\mathbf{u} = \mathbf{0}$ then both sides are zero so the statement obviously holds. If $\mathbf{u} \neq \mathbf{0}$ we'll make the following observation. Using the properties of inner products, we have for all scalars t that

$$0 \leq \langle t\mathbf{u} + \mathbf{v}, t\mathbf{u} + \mathbf{v} \rangle = t^2 \langle \mathbf{u}, \mathbf{u} \rangle + 2t\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle.$$

With $a = \langle \mathbf{u}, \mathbf{u} \rangle$, $b = 2\langle \mathbf{u}, \mathbf{v} \rangle$ and $c = \langle \mathbf{v}, \mathbf{v} \rangle$, this becomes

$$at^2 + bt + c \geq 0.$$

The above quadratic polynomial being nonnegative, means that it has either no roots or one repeated root. Noting that $\mathbf{u} \neq \mathbf{0}$ means that $a = \langle \mathbf{u}, \mathbf{u} \rangle > 0$ and that the nonexistence of two different roots means that $b^2 - 4ac \leq 0$, we get

$$4\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq 4\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle.$$

Cancelling the four and then taking square roots (note that $\langle \mathbf{u}, \mathbf{u} \rangle$ and $\langle \mathbf{v}, \mathbf{v} \rangle$ are nonnegative) we get

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle},$$

which can be stated as

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|.$$

□

And the following rules for the norm are still valid:

Theorem 9.1.8. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be vectors in an inner product space and let* k *a scalar. Then we have*

(a) $\|\mathbf{x}\| \geq 0$.

(b) $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$.

(c) $\|k\mathbf{x}\| = |k|\|\mathbf{x}\|$.

(d) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The proof is left as an exercise.

9.2 Orthogonality, projections and the Gram-Schmidt process

9.2.1 Orthogonal vectors

We define orthogonality in an inner product space in a similar way as we did in $\mathbb{R}^n$.

Definition 9.2.1. We say that two vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space, are *orthogonal* if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Note that Example 9.1.4 tells us that the zero vector is orthogonal to all vectors in an inner product space.

Definition 9.2.2. A set S in an inner product space is said to be an *orthogonal set* if any pair of two different vectors from S are orthogonal. If also, each vector in S has norm 1, then S is said to be an *orthonormal set*.

Example 9.2.3. Consider the inner product space $C([-\pi, \pi])$ with the inner product

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x)\, dx.$$

Show that the set

$$S = \{1/\sqrt{2}, \sin x, \cos x, \sin 2x, \cos 2x, \ldots\},$$

is an orthonormal set in this space.

For orthogonal vectors, the following generalization of Pythagoras' Theorem holds.

Theorem 9.2.4. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal vectors in an inner product space, then*

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Once you have an orthogonal set of nonzero vectors spanning some space, this will also be a linearly independent set and hence a basis for the space. This follows from the following result.

Lemma 9.2.5. *An orthogonal set* $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ *in an inner product space, is linearly independent if and only if each* $\mathbf{v}_j \neq \mathbf{0}$.

Proof. We have already seen that any set containing the zero vector is linearly dependent so it remains to see that if $\mathbf{v}_j \neq \mathbf{0}$ for each j, then S is linearly independent. Considering the equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r = \mathbf{0}, \tag{9.1}$$

for each j, $1 \leq j \leq r$, we can take the inner product of both sides with $\mathbf{v}_j$ and get

$$\langle k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \ldots + k_r\mathbf{v}_r, \mathbf{v}_j\rangle = \langle \mathbf{0}, \mathbf{v}_j\rangle,$$

and using properties 2 and 3 from the definition of an inner product, together with the result of Example 9.1.4, we get

$$k_1\langle\mathbf{v}_1, \mathbf{v}_j\rangle + k_2\langle\mathbf{v}_2, \mathbf{v}_j\rangle + \ldots + k_r\langle\mathbf{v}_r, \mathbf{v}_j\rangle = 0.$$

Since the set S is orthogonal, $\langle\mathbf{v}_i, \mathbf{v}_j\rangle = 0$ for each $i \neq j$, so the left hand side simplifies to

$$k_j\langle\mathbf{v}_j, \mathbf{v}_j\rangle = 0,$$

and since, by assumption, $\mathbf{v}_j \neq \mathbf{0}$, property 4 of inner products tells us that $\langle\mathbf{v}_j, \mathbf{v}_j\rangle \neq 0$, so we get $k_j = 0$. Since this holds for all j we see that the equation (9.1) has only the trivial solution, i.e. S is linearly independent. □

Most of the time we are interested in having an orthonormal basis for an inner product space. Such a basis is usually refered to as as *ON-basis*. Once we have an orthogonal or an orthonormal basis for a space, it is easy to find the coordinates of other vectors relative to this basis.

Theorem 9.2.6. *Suppose $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for an inner product space $\mathcal{V}$. Then for any $\mathbf{v} \in \mathcal{V}$,*

$$\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{v}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \ldots + \frac{\langle \mathbf{v}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n. \tag{9.2}$$

Of course, if B is an ON-basis the above formula simplifies to

$$\mathbf{v} = \langle \mathbf{v}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{v}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \ldots + \langle \mathbf{v}, \mathbf{v}_n \rangle \mathbf{v}_n.$$

If instead of being an orthogonal basis for $\mathcal{V}$, $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for *a subspace* $\mathcal{W}$ of $\mathcal{V}$, then the same formula gives what we know as the projection on the space. Before making this claim exact, we will introduce the concept of orthogonal complements and a theorem about projections.

We befin with orthogonal complements, the name given to the set of vectors orthogonal to all vectors in a subspace.

Definition 9.2.7. Suppose $\mathcal{V}$ is an inner product space and suppose $\mathcal{W}$ is a subspace of $\mathcal{V}$. Then the set of vectors in $\mathcal{V}$ which are orthogonal to all vectors in $\mathcal{W}$ is called *the orthogonal complement* of $\mathcal{W}$ and is denoted $\mathcal{W}^\perp$.

With set notation we have

$$\mathcal{W}^\perp = \{\mathbf{v} \in \mathcal{V} : \langle \mathbf{v}, \mathbf{w} \rangle = 0, \text{ for all } \mathbf{w} \in \mathcal{W}\}.$$

Theorem 9.2.8. *Let $\mathcal{V}$ be an inner product space and $\mathcal{W}$ a finite dimensional subspace of $\mathcal{V}$. Then for any vector $\mathbf{v} \in \mathcal{V}$ there exist unique vectors $\mathbf{w}_1$, $\mathbf{w}_2$ such that*

$$\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2, \quad \textit{where } \mathbf{w}_1 \in \mathcal{W}, \mathbf{w}_2 \in \mathcal{W}^\perp.$$

For a proof of the above theorem, you can look up [1, Theorem 6.3.4]. The theorem also allows us to define what we mean by the projection on a subspace.

Definition 9.2.9. With notation as in the above theorem, the vector $\mathbf{w}_1$ is called the *orthogonal projection of* $\mathbf{v}$ *onto* $\mathcal{W}$ and is denoted $\operatorname{proj}_{\mathcal{W}} \mathbf{v}$, while $\mathbf{w}_2$ is called the *component of* $\mathbf{v}$ *orthogonal to* $\mathcal{W}$ and is denoted $\operatorname{proj}_{\mathcal{W}^\perp} \mathbf{v}$.

With this notation we can write

$$\mathbf{v} = \operatorname{proj}_{\mathcal{W}} \mathbf{v} + \operatorname{proj}_{\mathcal{W}^\perp} \mathbf{v}. \tag{9.3}$$

We also have a formula for the projection. It looks like the formula in Theorem 9.2.6 although now the basis B is a basis for a subspace of $\mathcal{V}$ instead of $\mathcal{V}$ itself.

Theorem 9.2.10. *If $\mathcal{V}$ is an inner product space, $\mathcal{W}$ is a subspace of $\mathcal{V}$ and $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and orthogonal basis for $\mathcal{W}$. Then, for any $\mathbf{v} \in \mathcal{V}$,*

$$\text{proj}_{\mathcal{W}}\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{v}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \ldots + \frac{\langle \mathbf{v}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n, \tag{9.4}$$

while

$$\text{proj}_{\mathcal{W}^\perp}\mathbf{v} = \mathbf{v} - \text{proj}_{\mathcal{W}}\mathbf{v}. \tag{9.5}$$

Proof. Suppose

$$\mathbf{w}_1 = \frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{v}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \ldots + \frac{\langle \mathbf{v}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n, \tag{9.6}$$

and

$$\mathbf{w}_2 = \mathbf{v} - \mathbf{w}_1. \tag{9.7}$$

By Theorem 9.2.8 and Definition 9.2.9, What we need to prove is that

1. $\mathbf{w}_1 \in \mathcal{W}$,
2. $\mathbf{w}_2 \in \mathcal{W}^\perp$,
3. $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$.

Point 3 above follows immedietately from Equation (9.7), and since (9.6) states that $\mathbf{w}_1 \in \operatorname{span} B$, Point 1 follows since B is a basis for $\mathcal{W}$.

It remains for us to verify Point 2. In order to do so, we prove that $\mathbf{w}_2$ is orthogonal to every vector in B. Then it follows (tutorial problem) that indeed $\mathbf{w}_2$ is orthogonal to every vector in $\mathcal{W}$. So, evaluating the relevant inner products we get

$$\langle \mathbf{w}_2, \mathbf{v}_j \rangle = \langle \mathbf{v} - \mathbf{w}_1, \mathbf{v}_j \rangle = \left\langle \mathbf{v} - \left(\frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{v}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \ldots + \frac{\langle \mathbf{v}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n \right), \mathbf{v}_j \right\rangle =$$
$$\langle \mathbf{v}, \mathbf{v}_j \rangle - \left(\frac{\langle \mathbf{v}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\langle \mathbf{v}_1, \mathbf{v}_j \rangle + \frac{\langle \mathbf{v}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\langle \mathbf{v}_2, \mathbf{v}_j \rangle + \ldots + \frac{\langle \mathbf{v}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2}\langle \mathbf{v}_n, \mathbf{v}_j \rangle \right).$$

Since by assumption B is an orthogonal basis, i.e. $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$, most terms in the expression above dissapears, and we are left with

$$\langle \mathbf{v} - \text{proj}_{\mathcal{W}}\mathbf{v}, \mathbf{v}_j \rangle = \langle \mathbf{v}, \mathbf{v}_j \rangle - \frac{\langle \mathbf{v}, \mathbf{v}_j \rangle}{\|\mathbf{v}_j\|^2}\langle \mathbf{v}_j, \mathbf{v}_j \rangle = \langle \mathbf{v}, \mathbf{v}_j \rangle - \frac{\langle \mathbf{v}, \mathbf{v}_j \rangle}{\|\mathbf{v}_j\|^2}\|\mathbf{v}_j\|^2 = 0.$$

□

Example 9.2.11. Consider the inner product space $C([-\pi, \pi])$ with inner product

$$\langle f, g \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx,$$

and let $\mathcal{W}$ be the subspace spanned by $\{1/\sqrt{2}, \cos\theta, \sin\theta, \cos 2\theta, \sin 2\theta\}$. Find $\operatorname{proj}_{\mathcal{W}} f$, where

$$f(x) = \begin{cases} x + \pi & \text{for } -\pi \leq x < 0, \\ -x + \pi & \text{for } 0 \leq \pi \leq \pi. \end{cases}$$

Note that to apply the formulas (9.2) and (9.4) above we need a basis that is orthogonal (and preferably orthonormal as this makes the formulas simpler). Hence, an important question for us is if there always exists such a basis and in that case how we can find one. The procedure described in the next section solves that problem for us, at least for finite dimensional spaces.

9.2.2 The Gram-Schmidt orthonormalization process

Suppose we have a basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for an inner product space $\mathcal{V}$. Then we can construct an ON-basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, by following a procedure known as the *Gram-Schmidt process.*

1. Let $\mathbf{v}_1 = \mathbf{u}_1/\|\mathbf{u}_1$ and let $\mathcal{W}_1 = \operatorname{span}\{\mathbf{v}_1\}$. Note that we have $\|\mathbf{v}_1\| = 1$, so $\{\mathbf{v}_1\}$ is an ON-basis for $\mathcal{W}_1$.
2. Let $\mathbf{v}_2' = \operatorname{proj}_{\mathcal{W}_1^\perp} \mathbf{u}_2$, $\mathbf{v}_2 = \mathbf{v}_2'/\|\mathbf{v}_2'\|$ and $\mathcal{W}2 = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Note that $\mathbf{v}_2'$ and hence also $\mathbf{v}_2$ is orthogonal to $\mathbf{v}_1$, so $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an ON-basis for $\mathcal{W}_2$.
3. Let $\mathbf{v}_3' = \operatorname{proj}_{\mathcal{W}_2^\perp} \mathbf{u}_3$, $\mathbf{v}_3 = \mathbf{v}_3'/\|\mathbf{v}_3'\|$ and $\mathcal{W}3 = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Note that $\mathbf{v}_3'$ and hence also $\mathbf{v}_3$ is orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$, so $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an ON-basis for $\mathcal{W}_3$.
4. Repeat the process until you have a set of n orthonormal vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. Since this is a linearly independent set in the n-dimensional space $\mathcal{V}$, it is also a basis for $\sqsubseteq$.

If we consider an arbitrary n-dimensional inner product space $\mathcal{V}$ ($n \geq 1$), by definition it has a basis consisting of n vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$. The Gram-Schmidt process described above will then produce an orthonormal basis, i.e. we can summarize some of our discussion in the following theorem.

Theorem 9.2.12. *For a finite dimensional nonzero inner product space, there always exists an orthonormal basis.*

A consequence of the theorem is that we can always construct the projection given by Definition 9.2.9.

9.2.3 More on the orthogonal complement

In this section we will derive a few further results about the orthogonal complement of a subspace.

Theorem 9.2.13. *Suppose $\mathcal{V}$ is an inner product space and suppose $\mathcal{W}$ is a subspace of $\mathcal{V}$. Then*

- *$\mathcal{W}^{\perp}$ is a subspace of $\mathcal{V}$.*
- *$(\mathcal{W}^{\perp})^{\perp} = \mathcal{W}$.*

Proof. The proof of is left as a tutorial problem (the steps are outlined there). □

Since by the above theorem, $\mathcal{W}$ and $\mathcal{W}^{\perp}$ are orthogonal complements of one another, we simply say that they are *orthogonal complements*. An interesting fact regarding matrices is that the row space and null space of a matrix are orthogonal complements.

Theorem 9.2.14. *If A is an matrix with n columns, then the row space and nullspace of A are orthogonal complements in $\mathbb{R}^n$ with respect to the Euclidean inner product.*

Proof. See [1, Theorem 6.2.6]. □

Example 9.2.15. Let $\mathcal{W}$ be the subspace of $\mathbb{R}^3$ spanned by the vectors $\mathbf{v}_1 = (1, 2, 3)$, $\mathbf{v}_2 = (3, 2, 1)$. Find a basis for $\mathcal{W}^{\perp}$.

9.3 Best approximations

An important reason that projections are useful, is that they provide best approximations. Suppose $\mathcal{W}$ is a subspace of an inner product space $\mathcal{V}$ and we have a vector $\mathbf{v} \in \mathcal{V}$. It is a common situation that we want to find the vector in $\mathcal{W}$ that is "closest" to $\mathbf{v}$ and in one sense this will be the orthogonal projection of $\mathbf{v}$ on $\mathcal{W}$.

Theorem 9.3.1. *If $\mathcal{W}$ is a finite dimensional subspace of an inner product space $\mathcal{V}$, and if $\mathbf{v} \in \mathcal{V}$, then $\text{proj}_{\mathcal{W}}\mathbf{v}$ is the best approximation of $\mathbf{v}$ in $\mathcal{W}$ in the sense that*

$$\|\mathbf{v} - \text{proj}_{\mathcal{W}}\,\mathbf{v}\| < \|\mathbf{v} - \mathbf{w}\|,$$

for any $\mathbf{w} \in \mathcal{W}$ different from $\text{proj}_{\mathcal{W}}\,\mathbf{v}$.

Proof. Take any $\mathbf{w} \in \mathcal{W}$. We have

$$\mathbf{v} - \mathbf{w} = (\mathbf{v} - \text{proj}_{\mathcal{W}}\,\mathbf{v}) + (\text{proj}_{\mathcal{W}}\,\mathbf{v} - \mathbf{w}).$$

Now, since both $\text{proj}_{\mathcal{W}}\,\mathbf{v}$ and $\mathbf{w}$ are in $\mathcal{W}$, we also have $(\text{proj}_{\mathcal{W}}\,\mathbf{v} - \mathbf{w}) \in \mathcal{W}$. On the other hand $\mathbf{v} - \text{proj}_{\mathcal{W}}\,\mathbf{v} = \text{proj}_{\mathcal{W}^{\perp}}\,\mathbf{v}$ is in $\mathcal{W}^{\perp}$, so by Pythagoras generalized theorem (Theorem 9.2.4), we have

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v} - \text{proj}_{\mathcal{W}}\,\mathbf{v}\|^2 + \|\text{proj}_{\mathcal{W}}\,\mathbf{v} - \mathbf{w}\|^2.$$

Now, if $\mathbf{w}$ is different from $\text{proj}_{\mathcal{W}}\,\mathbf{v}$, the second term above is positive, so

$$\|\mathbf{v} - \text{proj}_{\mathcal{W}}\,\mathbf{v}\|^2 < \|\mathbf{v} - \mathbf{w}\|^2.$$

Taking square roots and noting that $\|\mathbf{z}\| \geq 0$ for any $\mathbf{z} \in \mathcal{V}$, we conclude that

$$\|\mathbf{v} - \operatorname{proj}_{\mathcal{W}} \mathbf{v}\| < \|\mathbf{v} - \mathbf{w}\|.$$

□

We will now look at two common applications of this theorem. Fourier series and least squares solutions.

9.3.1 Fourier series

Let us again consider the space $C([-\pi, \pi])$ of functions continuous on $[-\pi, \pi]$ (it is of course possible to consider other intervals as well) with inner product

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) g(x)\, dx.$$

Let us also denote,

$$g_0(x) = \frac{1}{\sqrt{2}}, \quad g_k(x) = \cos kx, \quad k = 1, 2, \ldots,$$

$$h_k(x) = \sin kx, \quad k = 1, 2, \ldots,$$

and

$$a_k = \langle f, g_k \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx\, dx, \quad k = 0, 1, 2, \ldots,$$

$$b_k = \langle f, h_k \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx\, dx, \quad k = 1, 2, \ldots,$$

and also

$$\mathcal{T}_n = \operatorname{span}\{g_0, h_1, g_1, \ldots, h_n, g_n\}.$$

In Example 9.2.3 we saw that all the functions $\{g_0, h_1, g_1, \ldots\}$ are orthogonal, so that makes $\{g_0, h_1, g_1, \ldots, h_n, g_n\}$ an ON-basis for $\mathcal{T}_n$, which thus is an $2n+1$ dimensional subspace of $C([-\pi, \pi])$. A function in $\mathcal{T}_n$ is called a *trigonometric polynomial* of degree n.

In Example 9.2.11 we calculated the orthogonal projection $T_2(x)$ of the function

$$f(x) = \begin{cases} x + \pi & \text{for } -\pi \leq x < 0, \\ -x + \pi & \text{for } 0 \leq \pi \leq \pi. \end{cases}$$

onto $\mathcal{T}_2$. We got the expression for $T_2(x)$ from the projection formula

$$\begin{aligned} T(x) = \operatorname{proj}_{\mathcal{W}} f(x) &= \\ &\langle f, g_0 \rangle g_0(x) + \langle f, h_1 \rangle h_1(x) + \langle f, g_1 \rangle g_1(x) + \langle f, h_2 \rangle h_2(x) + \langle f, g_2 \rangle g_2(x) = \\ &\frac{a_0}{2} + b_1 \sin x + a_1 \cos x + b_2 \sin 2x + a_2 \cos 2x. \end{aligned}$$

By Theorem 9.3.1, the function $T(x)$ is the best approximation in $\mathcal{T}_2$ of $f(x)$ on $[-\pi, \pi]$ in the sense that among all functions $T \in \mathcal{T}_2$, $T_2(x)$ minimizes the norm,

$$\|f - T\| = \sqrt{\langle f - T, f - T\rangle} = \left(\frac{1}{\pi}\int_{-\pi}^{\pi}(f(x) - T(x))^2\,dx\right)^{1/2},$$

end hence it also minimizes the *mean square error*

$$\int_{-\pi}^{\pi}(f(x) - T(x))^2\,dx.$$

More generally, if f is any function in $C([-\pi, \pi])$, then of all the trigonometric polynomials of degree n, i.e. of all functions in $\mathcal{T}_n$, the projection T_n of f onto $\mathcal{T}_n$ is the best approximation of f on $[-\pi, \pi]$ among all functions in $\mathcal{T}_n$, in the sense that it minimizes the mean square error. Doing the same calcualation as above, we get

$$T_n(x) = \sum_{k=0}^{n}\langle f, g_k\rangle g_k(x) + \sum_{k=1}^{n}\langle f, h_k\rangle h_k(x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kx + \sum_{k=1}^{n} b_k \sin kx.$$

You might recognize the expressions above. The coefficients a_k and b_k are called the *Fourier coefficients* of f and the sum above is a partial sum of the *Fourier series*

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + \sum_{k=1}^{\infty} b_k \sin kx.$$

Since the space $\mathcal{T}_{n+1}$ includes all functions from $\mathcal{T}_n$ (and more), and since the projection minimizes the mean square error among all functions in the space, we can conclude that the mean square error when we approximate f with its projection on $\mathcal{T}_{n+1}$, can't be bigger than if we project on $\mathcal{T}_n$. In fact, we would expect the error to become smaller. Although outside the scope of this course, the following facts are relevant.

- If $f \in C([-\pi, \pi])$, and T_n is the projection onto $\mathcal{T}_n$ of f, then the error $\|f - T_n\|$ tends to zero as n tends to infinity.

- Although the statement above tells us that the mean square error in the approximation tends to zero as n approaches infinity, we can't conclude that $|f(x) - T_n(x)|$ tends to zero for all $x \in [-\pi, \pi]$ (why not?). In fact, it can be proved that there exist continuous functions on $[-\pi, \pi]$ for which $|f(0) - T_n(0)| \to \infty$ as $n \to \infty$.

- A positive result though is that if $f \in C^1([-\pi, \pi])$ (meaning that f has a continuous first derivative), then it is possible to prove that $|f(x) - T_n(x)|$ tends to zero for all $x \in [-\pi, \pi]$.

9.3.2 Least squares solutions

Our other application of the best approximation theorem (Theorem 9.3.1) is least squares solutions. In Tutorial 1, we considered the linear system

$$\begin{array}{ccccccc} x & + & y & + & 2z & = & a, \\ x & & & + & z & = & b, \\ 2x & + & y & + & 3z & = & c. \end{array}$$

The corresponding augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & a \\ 1 & 0 & 1 & b \\ 2 & 1 & 3 & c \end{array}\right],$$

and after Gauss-Jordan elimination this becomes

$$\left[\begin{array}{ccc|r} 1 & 0 & 1 & b \\ 0 & 1 & 1 & a-b \\ 0 & 0 & 0 & c-b-a \end{array}\right],$$

so we see that the system is consistent if and only if $c - b - a = 0$.

Now, suppose that the linear system above describes a physical system, where a, b and c are measured data from which we wish to calculate x_1, x_2 and x_3. According to our theoretical model, only a, b and c satisfying the condition $c-b-a = 0$ can occur as physical data, but since our measurements have limited accuracy we are unlikely to have the condition exactly satisfied and then the system is inconsistent. In this situation, it is relevant to look for approximate solutions instead.

In general, consider a linear system

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

Due to for example measurement errors, the system might be inconsistent althogh it should be consistent. Can we still find an approximate solution? What we would like is to find an approximate solution $\mathbf{x}$ such that even if we can't make $A\mathbf{x} = \mathbf{b}$, we will at least try to *minimize* $\|A\mathbf{x} - \mathbf{b}\|$. A vector $\mathbf{x}$ that does to this is called a *least squares solution* and it is possible to find this by a projection argument explained below.

If our system $A\mathbf{x} = \mathbf{b}$ has m equations and n unknowns, it means that A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$ and we consider $\mathbf{x} \in \mathbb{R}^n$. First, we make the important observation that for any $\mathbf{x} \in \mathbb{R}^n$, the vector $A\mathbf{x}$ *is a linear compbination of the columns of* A, which is evident from the calculation

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \ldots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Hence, as $\mathbf{x}$ varies over all vectors in $\mathbb{R}^n$, $A\mathbf{x}$ varies over all possible linear combinations of the columns, that is, $A\mathbf{x}$ varies over all vectors in the column space $C(A)$. Hence, to get the best approximate solution we need $A\mathbf{x}$ to be the vector in $C(A)$ that minimizes $\|Ax - \mathbf{b}\|$ so in other words we should let

$$A\mathbf{x} = \operatorname{proj}_{C(A)} \mathbf{b}. \tag{9.8}$$

In theory, we have now solved our problem. Given A and $\mathbf{b}$, we can calculate $\operatorname{proj}_{C(A)} \mathbf{b}$ and then solve (9.8), which will now be consistent since the right hand side is in the column space of A. However, this is not practical. To calculate the projection $\operatorname{proj}_{C(A)} \mathbf{b}$ with the projection formula (9.4) requires us to have an orthogonal basis for $C(A)$ and to get that we must first find a basis for $C(A)$ and then apply the Gram-Schmidt process. Totally there will be many steps to perform. Fortunately however, there is a shortcut.

Using the equation (9.8) and Theorem 9.2.8, we have

$$\mathbf{b} - A\mathbf{x} = \mathbf{b} - \operatorname{proj}_{C(A)} \mathbf{b} = \operatorname{proj}_{C(A)^\perp} \mathbf{b},$$

so $\mathbf{b} - A\mathbf{x}$ lies in the orthogonal complement $C(A)^\perp$ of $C(A)$. However, using the simple fact that the column space equals the row space of the transposed matrix, together with Theorem 9.2.14, we have

$$C(A)^\perp = R(A^T)^\perp = N(A^T),$$

so $\mathbf{b} - A\mathbf{x}$ must belong to the nullspace of A^T, i.e. we must have

$$A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0},$$

which we can rewrite as

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

We summarize our discussion in the following theorem.

Theorem 9.3.2. *For a linear system $A\mathbf{x} = \mathbf{b}$, the associated normal system*

$$A^T A\mathbf{x} = A^T\mathbf{b},$$

is consistent and all solutions to the normal system are least squares solutions of $A\mathbf{x} = \mathbf{b}$. Furthermore, for any least squares solution $\mathbf{x}$, we have

$$\operatorname{proj}_{C(A)} \mathbf{b} = A\mathbf{x}.$$

Example 9.3.3. Find a least squares solution to the system

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

Example 9.3.4. Find the orthogonal projection of $\mathbf{v} = (-3, -3, 8, 9)$ on $\mathcal{W} = \operatorname{span}\{(3, 1, 3, 0), (1, 2, 1, 1), (-

9.4 Change of basis

Example 9.4.1. Consider $\mathbb{R}^2$ and the basis $\mathcal{B} = \{(1,1),(-1,1)\}$. Suppose we want to find $(\mathbf{v})_{\mathcal{B}} = (k_1, k_2)$, where $\mathbf{v} = (2,3)$. For that we need to solve $k_1(1,1) + k_2((-1,1) = (2,3)$, which is the linear system

$$k_1 - k_2 = 2,$$
$$k_1 + k_2 = 3,$$

or in matrix form

$$A[\mathbf{v}]_{\mathcal{B}} = [\mathbf{v}], \tag{9.9}$$

where

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

(Here, we use the matrix notation: $[\mathbf{v}] = \begin{bmatrix} a \\ b \end{bmatrix}$ for $\mathbf{v} = (a,b)$ and $[\mathbf{v}]_{\mathcal{B}} = [(\mathbf{v})_{\mathcal{B}}]$.)

Since the two columns of A are the basis vectors of $\mathcal{B}$, which are linearly independent, we see from Proposition 7.4.16 that A is invertible. Let $P = A^{-1} = \begin{bmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{bmatrix}$. Then

$$[\mathbf{v}]_{\mathcal{B}} = P[\mathbf{v}]. \tag{9.10}$$

In the above example, note that with the standard basis $\mathcal{B}' = \{\mathbf{v}_1' = (1,0), \mathbf{v}_1' = (0,1)\}$ and the basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1 = (1,1)$ and $\mathbf{v}_2 = (-1,1)$, we have

$$[\mathbf{v}_1']_{\mathcal{B}} = \begin{bmatrix} 1/2 \\ -1/2 \end{bmatrix}, \quad [\mathbf{v}_2']_{\mathcal{B}} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix},$$

Hence,

$$P = \begin{bmatrix} \begin{bmatrix} \mathbf{v}_1 \end{bmatrix}_{\mathcal{B}} \begin{bmatrix} \mathbf{v}_2 \end{bmatrix}_{\mathcal{B}} \end{bmatrix},$$

and from equations (9.10) and (9.9) we have

$$[\mathbf{v}]_{\mathcal{B}} = P[\mathbf{v}]_{\mathcal{B}'}, \quad \text{and } [\mathbf{v}]_{\mathcal{B}'} = P^{-1}[\mathbf{v}]_{\mathcal{B}}.$$

Inspired by the observations above, we make the following generalization.

Theorem 9.4.2. *Let $\mathcal{V}$ be a finite dimensional vector space, and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\mathcal{B}' = \{\mathbf{v}_1', \mathbf{v}_2', \ldots, \mathbf{v}_n'\}$, be two bases for $\mathcal{V}$. Then the matrix*

$$P = \begin{bmatrix} \begin{bmatrix} \mathbf{v}_1' \end{bmatrix}_{\mathcal{B}} \begin{bmatrix} \mathbf{v}_2' \end{bmatrix}_{\mathcal{B}} \cdots \begin{bmatrix} \mathbf{v}_n' \end{bmatrix}_{\mathcal{B}} \end{bmatrix},$$

is invertible and has the property that for any $\mathbf{v} \in \mathcal{V}$,

$$[\mathbf{v}]_{\mathcal{B}} = P[\mathbf{v}]_{\mathcal{B}'}, \quad \text{and } [\mathbf{v}]_{\mathcal{B}'} = P^{-1}[\mathbf{v}]_{\mathcal{B}}.$$

Proof. Suppose

$$[\mathbf{v}]_{\mathcal{B}'} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix},$$

which means that

$$\mathbf{v} = k_1\mathbf{v}_1' + k_2\mathbf{v}_2' + \ldots + k_n\mathbf{v}_n'.$$

Then

$$[\mathbf{v}]_{\mathcal{B}} = [k_1\mathbf{v}_1' + k_2\mathbf{v}_2' + \ldots + k_n\mathbf{v}_n']_{\mathcal{B}} = k_1[\mathbf{v}_1']_{\mathcal{B}} + k_2[\mathbf{v}_2']_{\mathcal{B}} + \ldots + k_n[\mathbf{v}_n']_{\mathcal{B}} =$$

$$\begin{bmatrix} \begin{bmatrix} \mathbf{v}_1' \end{bmatrix}_{\mathcal{B}} \begin{bmatrix} \mathbf{v}_2' \end{bmatrix}_{\mathcal{B}} \cdots \begin{bmatrix} \mathbf{v}_n' \end{bmatrix}_{\mathcal{B}} \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = P[\mathbf{v}]_{\mathcal{B}'}.$$

Furthermore, since $\mathcal{B}'$ is linearly independent, Proposition 7.4.16 tells us that P is invertible, so multiplying the above equation with P^{-1} gives us

$$[\mathbf{v}]_{\mathcal{B}'} = P^{-1}[\mathbf{v}]_{\mathcal{B}}.$$

□

Definition 9.4.3. The matrix P above is called the *transition matrix* from $\mathcal{B}'$ to $\mathcal{B}$.

Note: The transition matrix from $\mathcal{B}$ to $\mathcal{B}'$ will be P^{-1}.

Example 9.4.4. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v_2}\}$ and $\mathcal{B}' = \{\mathbf{v\prime}_1, \mathbf{v}_2'\}$, where $\mathbf{v}_1 = (1, 0)$, $\mathbf{v}_2 = (0, 1)$, $\mathbf{v}_1' = (2, 1)$ and $\mathbf{v}_2' = (-3, 4)$, be two bases of $\mathbb{R}^2$.

(a) Find the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$.

(b) Find the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$.

(c) Use the matrix above to evaluate $(\mathbf{w})_{\mathcal{B}'}$, when $(\mathbf{w})_{\mathcal{B}} = (3, -5)$.

(d) Check your work by evaluating $(\mathbf{w})_{\mathcal{B}'}$ directly.

9.5 Orthogonal matrices

When changing basis, we are often interested in just rotating the coordinate system and maybe change the orientation (i.e. between right handed and left handed coordinates). A transition matrix that does that is an orthogonal matrix, although we will state its definition somewhat differently.

Definition 9.5.1. An invertible matrix A is said to be an *orthogonal matrix* if $A^{-1} = A^T$.

This definition seems to have very little to do with orthogonality, but indeed there is a connection as explained by the following theorem.

Theorem 9.5.2. *For an $n \times n$ matrix A, the following are equivalent.*

1. *A is an orthogonal matrix.*
2. *The columns of A form an ON-basis for $\mathbb{R}^n$ (with the Euclidean inner product).*
3. *The rows of A form an ON-basis for $\mathbb{R}^n$*

For a proof of this theorem you can look up any textbook in linear algebra.

We can easily deduce the following facts.

Theorem 9.5.3.
- *A product of orthogonal matrices is orthogonal.*
- *If A is an orthogonal matrix, then* $\det A = 1$ *or* $\det A = -1$.

Proof. Tutorial problem. □

And we also have the following equivalences.

Theorem 9.5.4. *If A is an $n \times n$ matrix, then the following are equivalent.*

(a) *A is orthogonal.*

(b) $\|A\mathbf{x}\| = \|\mathbf{x}\|$ *for all* $\mathbf{x} \in \mathbb{R}^n$.

(c) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proof. The proofs that (a)⇒(b) and (b)⇒(c) are tutorial problems. The proof that (c)⇒(a) you can look up in for example [1, Theorem 6.6.3]. □

This has a geometric interpretation. If A is an orthogonal matrix, then part (b) says that multiplication with A preserves length and part (c) says that multiplication with A preserves the Euclidean inner product, so in particular it preserves orthogonality. Another result of interest is this.

Theorem 9.5.5. *If P is the transition matrix from one orthonormal basis to another orthonormal basis for an inner product space. Then P is an orthogonal matrix.*

A proof can be found in [1, Theorem 6.6.4].

9.6 Diagonalization

9.6.1 Digagonalization

Example 9.6.1. Consider $\mathbb{R}^2$ and let ℓ be a line through the origin, making a 30° angle with the positive x-axis. Let T be the linear transformation that projects a vector orthogonally on ℓ. We can calculate the standard matrix A for T and we'll get

$$A = \begin{bmatrix} 3/4 & \sqrt{3}/4 \\ \sqrt{3}/4 & 1/4 \end{bmatrix}.$$

This means that if $\mathcal{B}$ is the standard basis, we have

$$[T(\mathbf{v})]_{\mathcal{B}} = A[\mathbf{v}]_{\mathcal{B}}$$

On the other hand, let $\mathcal{B}' = \{(\sqrt{3}/2, 1/2), (-1/2, \sqrt{3}/2)\}$. Note that the basis vectors of $\mathcal{B}'$ are obtained by rotating the standard basis vectors 30° anticlockwise. Hence they define a coordinate system $x'y'$ where the x'-axis is parallel to ℓ and the y'-axis is perpendicular to it. Hence, in this coordinate system, T is just projection on the x'-axis and we realize that

$$[T(\mathbf{v})]_{\mathcal{B}'} = D[\mathbf{v}]_{\mathcal{B}'}, \tag{9.11}$$

where

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Furthermore, if P is the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$, we see that

$$[T(\mathbf{v})]_{\mathcal{B}'} = P^{-1}[T(\mathbf{v})]_{\mathcal{B}} = P^{-1}A[\mathbf{v}]_{\mathcal{B}} = P^{-1}AP[\mathbf{v}]_{\mathcal{B}'}.$$

Comparing the above equation with equation (9.11), suggests that

$$D = P^{-1}AP,$$

which we can indeed verify by calculating P and P^{-1} (do that!).

The above example shows that with a good choice of coordinates, the matrix defining T becomes significantly simpler. A natural question to ask is then, given a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$, can we find a basis for $\mathbb{R}^n$ such that the matrix representation of T in this basis is as simple as possible? We will not solve this problem in general, but as the above example shows, there are situations where we can find a basis such that the matrix for T is a diagonal matrix D and we are going to find out exactly when this is possible.

If we look at the above example again, the basis vectors of $\mathcal{B}'$, expressed in the standard basis $\mathcal{B}$, form the columns of the transition vector P and we have the relation

$$D = P^{-1}AP,$$

where A is the standard matrix for T and D, a diagonal matrix which is the matrix for T relative to the basis $\mathcal{B}'$. Our problem about finding a basis giving a diagonal matrix for T can then be formulated in the following manner.

Question:
Given an $n \times n$ matrix A, is there an invertible matrix P such that $D = P^{-1}AP$ is a diagonal matrix?

If the answer to the above question is yes, then we say that A is *diagonalizable*, and we say that P *diagonalizes* A.

In our example, the basis vectors for $\mathcal{B}'$ which form the columns of P are actually eigenvectors of A (this is geometrically evident), and this is no coincidence. As a matter of fact we have the following theorem.

Theorem 9.6.2. *Let A be an $n \times n$ matrix. Then the following are equivalent:*

(a) The matrix A is diagonalizable.

(b) There is a basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

Proof. First we prove that $(a) \Rightarrow (b)$:
If A is diagonalizable, this means that there is an invertible matrix P such that

$$D = P^{-1}AP \tag{9.12}$$

is a diagonal matrix. We will prove that the columns of P are all eigenvectors of A. Then, since P is invertible, the Propositions 7.4.11 and 7.4.16 tell us that these vectors span $\mathbb{R}^n$ and are linearly independent. Hence they form a basis of $\mathbb{R}^n$.

It remains for us to prove that the columns of P are eigenvectors of A. To do that we rewrite equation (9.12) as

$$PD = AP.$$

Evaluating the left hand side first, we let

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}, \quad \text{and } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Hence

$$PD = \begin{bmatrix} \lambda_1 p_{11} & \lambda_2 p_{12} & \cdots & \lambda_n p_{1n} \\ \lambda_1 p_{21} & \lambda_2 p_{22} & \cdots & \lambda_n p_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda_1 p_{n1} & \lambda_2 p_{n2} & \cdots & \lambda_n p_{nn} \end{bmatrix} = \left[\lambda_1 \mathbf{p}_1 \,\middle|\, \lambda_2 \mathbf{p}_2 \,\middle|\, \cdots \,\middle|\, \lambda_n \mathbf{p}_n \right],$$

where $\mathbf{p}_1, \ldots, \mathbf{p}_n$ are the column vectors of P.

Evaluating the right hand side, using that [jth column of AP] = A[jth column of P], we have

$$AP = \left[A\mathbf{p}_1 \,\middle|\, A\mathbf{p}_2 \,\middle|\, \cdots \,\middle|\, A\mathbf{p}_n \right].$$

Since $PD = AP$, we get $A\mathbf{p}_j = \lambda_j \mathbf{p}_j$ for $1 \leq j \leq n$. Furthermore, the invertibility of P, guarantees that all $\mathbf{p}_j \neq \mathbf{0}$, so indeed each $\mathbf{p}_j$ is an eigenvector of A, with eigenvalue λ_j.

We proceed by proving that $(b) \Rightarrow (a)$:
Let $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ be a basis of $\mathbb{R}^n$, where each $\mathbf{p}_j$ is an eigenvector of A with corresponding eigenvalue λ_j, i.e. $A\mathbf{p}_j = \lambda_j \mathbf{p}_j$, $1 \leq j \leq n$. We let

$$P = \left[\mathbf{p}_1 \,\middle|\, \mathbf{p}_2 \,\middle|\, \cdots \,\middle|\, \mathbf{p}_n \right], \quad \text{and } D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

and we will show that with this choice of P and D, we have $PD = AP$. Then, since $\{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is linearly independent, Proposition 7.4.16 tells us that P is invertible, so we can rewrite $PD = AP$ as $D = P^{-1}AP$.

It remains for us to show that $PD = AP$. By the same calculation as before, we have

$$PD = \left[\lambda_1\mathbf{p}_1 \,\middle|\, \lambda_2\mathbf{p}_2 \,\middle|\, \cdots \,\middle|\, \lambda_n\mathbf{p}_n \right].$$

On the other hand, using that $A\mathbf{p}_j = \lambda_j \mathbf{p}_j$, we have

$$AP = \left[A\mathbf{p}_1 \,\middle|\, A\mathbf{p}_2 \,\middle|\, \cdots \,\middle|\, A\mathbf{p}_n \right] = \left[\lambda_1\mathbf{p}_1 \,\middle|\, \lambda_2\mathbf{p}_2 \,\middle|\, \cdots \,\middle|\, \lambda_n\mathbf{p}_n \right],$$

which concludes our proof. □

The above theorem not only tells us when A is diagonalizable. The proof also tells us that when it is, then the matrix P that diagonalizes A is formed by taking the n linearly independent eigenvectors of A and put them as columns in P. The diagonal matrix D will then have the corresponding eigenvalues in the diagonal.

Note: If we have n linearly independent vectors in $\mathbb{R}^n$, then Theorem 7.4.16 guarantees that the matrix with these columns is invertible and then Theorem 7.4.11 tells us that they span $\mathbb{R}^n$. In other words, n linearly independent vectors in $\mathbb{R}^n$ will form a basis for $\mathbb{R}^n$, so in order to find a basis consisting of eigenvectors, it's enough to find n linearly independent eigenvectors.

Example 9.6.3. Is the matrix

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

diagonalizable? If it is, then find a matrix P that diagonalizes A and evaluate D.

When we are looking for n linearly independent eigenvectors, there is some terminology and a result which are useful.

Definition 9.6.4. If λ is an eigenvalue of A, then the solution space of the homogeneous system

$$(\lambda I - A)\mathbf{x} = \mathbf{0},$$

is called the *eigenspace* corresponding to λ.

In other words, the eigenspace corresponding to an eigenvalue λ consists of all the eigenvectors corresponding to this eigenvalue together with the zero vector.

The following theorem helps us when we're looking for linearly independent eigenvectors.

Theorem 9.6.5. *Suppose $\mathcal{S}$ is a set of eigenvectors of a matrix. If any subset of $\mathcal{S}$ with only eigenvectors from the same eigenspace is linearly independent, then $\mathcal{S}$ itself is linearly independent.*

For a proof of this statement you can look up a more advanced textbook in linear algebra.

Example 9.6.6. Is the matrix

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

diagonalizable? If it is, then find a matrix P that diagonalizes A and evaluate D.

9.6.2 Orthogonal diagonalization

We are often more interested in orthonormal bases than other bases. In the problem we just considered, we were looking for a basis relative to which the matrix for a linear transformation is diagonal. We might strengthen our demands further and ask not just for a basis but for an orthonormal basis. Since the matrix P that diagonalizes A is the transition matrix from the new basis to the standard basis for $\mathbb{R}^n$, both of which are now orthonormal bases, then Theorem 9.5.5 tells us that we should be looking for an *orthogonal matrix* P. Our problem is the following.

Question:
Given an $n \times n$ matrix A, is there an orthogonal matrix P such that $D = P^{-1}AP = P^TAP$ is a diagonal matrix?

If the answer to the above question is yes, then we say that A is *orthogonally diagonalizable*, and we say that P *orthogonally diagonalizes* A.

Analougously to Theorem 9.6.2

Theorem 9.6.7. *Let A be an $n \times n$ matrix. Then the following are equivalent:*

(a) The matrix A is orthogonally diagonalizable.

(b) There is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

(c) A is symmetric.

We omit the proof of this theorem, but note that it looks very much like Theorem 9.6.2 except that "diagonalizable" has been replaced by "orthogonally diagonalizable", "basis" has been replaced with "orthogonal basis" and *there is an extra equivalent statement that is very easy to check*. We can immediately, by inpection, see if a matrix is symmetric or not and hence deduce whether A is orthogonally diagonalizable or not.

Like before, we construct P by choosing eigenvectors of A as columns, but now we don't just require the columns to be linearly independent, but also orthonormal (since P is supposed to be an orthogonal matrix). Like Theorem 9.6.5 helped us find linearly independent eigenvectors, the following theorem helps us find orthonormal ones.

Theorem 9.6.8. *If A is a symmetric matrix, then eigenvector from different eigenspaces are orthogonal.*

As a consequence, we have the following procedure for orthogonally diagonalizing a symmetric matrix.

Step 1. Find a basis for each eigenspace of A.

Step 2. Apply the Gram-Schmidt process to each of these mases to obtain an ON-basis for each eigenspace.

Step 3. Form a matrix P by using the basis vectors constructed as columns. Step 2 together with the preceding theorem guarantees that the columns will form an orthonormal set and Theorem 9.6.7 that these are a basis for $\mathbb{R}^n$. The matrix P orthogonally diagonalizes A.

Example 9.6.9. Find an orthogonal matrix that diagonalizes

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

To be done during lecture.

Bibliography

[1] H. Anton *Elementary Linear Algebra*, 9th ed. Wiley, 2005.