Contents

Preface

Linear Algebra. These two words, when read and interpreted as two separate words, we are familiar with them. However, putting them together, we get this new realm of Mathematics which might not seem too scary! Yet, most readers might be overwhelmed by the abstract nature of the content and find the theorems and concepts covered in Linear Algebra daunting. However, what makes Linear Algebra stand out from the rest? Let us take a walk down *memory lane* first.

At the start of learning Calculus, students would be exposed to the derivative and the integral. For example, proving that

$$\frac{d}{dx}(\sin x) = \cos x$$

by first principles and evaluating

$$\int_0^6 x + \frac{x^3}{3}\ dx$$

are easy. However, Linear Algebra requires the reader to look at Mathematics from a completely different lens. One needs to have a thorough understanding of the definitions and theorems and work through the examples with pen and paper, and appreciate how each step of the working is related to a definition/theorem. In addition, Linear Algebra is a branch of Mathematics which gives young students a taste of formal and rigorous mathematical proof. Personally, I would advice the reader to write the proofs of the theorems and understand each step of the argument. No formula came out of thin air, analogous to how sleight of hand works. Whenever you are in doubt, talk to your friends and teachers for ideas and seek clarification.

Personally, I started learning Linear Algebra in Secondary Four. However, I learnt it in an unconventional manner as I did not learn the topics in sequence. I started off with the section on eigenvalues and eigenvectors first because I was interested in the method to compute matrices raised to a power. This method is known as diagonalisation. Subsequently, when I took up Further Mathematics in Junior College, I read my school's notes and learnt the content sequentially, starting with matrices, then linear spaces, linear transformations and finally, eigenvalues and eigenvectors. I struggled learning Linear Algebra at the start due to its abstract nature, and it was not like other mathematical content which I learnt before like Calculus, Geometry and Number Theory.

How I studied Linear Algebra was through asking my friends teachers for hints and solutions, as well as reading up online materials for reference. When looking for a textbook (be it physical or e-book), it is good to find one with visuals since it would be very helpful when dealing with the geometrical interpretation of certain concepts. Some textbooks which I have read since my time in Junior College are 'Linear Algebra and Its Applications' by David C Lay, Judith McDonald, and Steven R Lay, 'Linear Algebra Concepts and Techniques on Euclidean Spaces' by Ma Siu Lun, Ng Kah Loon, Victor Tan, 'Matrices and Linear Algebra' by Hans Schneider and 'Elementary Linear Algebra' by Howard Anton.

To better appreciate geometrical idea in Linear Algebra, I chanced upon the videos made by 3Blue1Brown, whose name is no stranger to Mathematics fanatics. Actually, the first few videos which I watched on his channel were not related to Linear Algebra, but were about the hardest problem on the hardest test (Putnam Competition problem) and the infamous windmill problem from the International Mathematical Olympiad (IMO) 2011. Having said that, do check out his compilation on Linear Algebra. It greatly helped me visualise key concepts and appreciate it from a different point-of-view. blackpenredpen and Dr Peyam also produce great videos on Linear Algebra too, and I have been watching them since the start of Junior College. Moreover, Dr. Peyam has a video on 111 true/false questions about Linear Algebra, which is very helpful when you have finished learning a topic and wish to revise by checking whether you have any conceptual errors. To those who have the ability to listen to long lectures, MIT OpenCourseWare, a channel by the Massachusetts Institute of Technology, published a series of lectures filmed back in 2005. The lecturer is Gilbert Strang, who explains concepts in a clear and concise manner.

Appreciate the beauty of Linear Algebra by finding connections with it to the real world. I kid you not, there are loads of applications which hinge on these *seemingly innocent*, yet powerful two words. When I was in Year Two of Junior College, I was given the opportunity to present to students and teachers in my school about the applications of Mathematics in movie graphics - which of course, uses Linear Algebra. It started off with the idea of linear transformations like translation, scaling, shear, and rotation in particular, in $\mathbb{R}^3$, then moving on to the concept of subdivision surfaces and the puppet warp tool. A subdivision surface is a curved surface represented by the specification of a coarser polygonal mesh and produced by a recursive algorithmic method. To create this, we have to use the idea of Riemann Sums. In particular, I shared with the audience members about the Catmull-Clark Algorithm, and how it is applied to movie graphics, such as in the ones made by Pixar. Google to find out why!

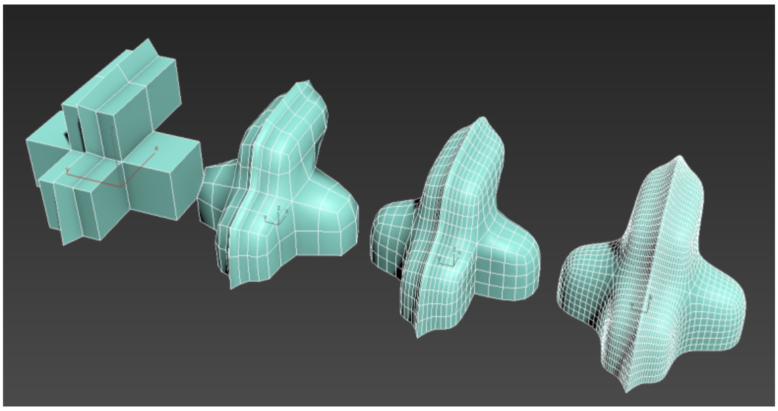

Figure 1: Subdivision surfaces

Now, it is your turn to venture into the world of Linear Algebra. Some of the ideas in each application section might require knowledge from other topics. It would be best if the reader has a good foundational understanding of the topics first before delving into the respective application sections.

For MA2001 (Linear Algebra I), the topics covered are:
1. Gaussian Elimination and Linear Systems
2. Matrices
3. Vector Spaces
4. Vector Spaces Associated with Matrices
5. Orthogonality
6. Diagonalisation
7. Linear Transformations

For MA2101 (Linear Algebra II), the topics covered are:
8. General Vector Spaces
9. General Linear Transformations
10. Multilinear Forms and Determinants
11. Diagonalisation and Jordan Canonical Forms
12. Inner Product Spaces

Lastly, please do not sell the notes. Instead, if you know someone who needs it in his/her studies, feel free to share it. I hope that this would act as a great supplement.

Thang Pang Ern

1 Gaussian Elimination and Linear Systems

1.1 Linear Systems and their Solutions

1.1.1 General Equation of a Line

Briefly recall from O-Level Mathematics (4048) that a line in the xy-plane can be represented algebraically by an equation of the form $ax + by = c$, where $a, b \neq 0$. Making y the subject yields

$$y = -\frac{a}{b}x + \frac{c}{b},$$

so the gradient of the line is $-\frac{a}{b}$ and the y-intercept is $\frac{c}{b}$. An equation of the above form is a linear equation in terms of x and y (or have variables x and y).

1.1.2 Extension to n Variables

A linear equation in n variables $x_1, x_2, \ldots, x_n$ has the form

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

where $a_i, b \in \mathbb{R}$ for $1 \leq i \leq n$. In this case, not all the a_i's need to be zero. If $a_i, b \in \mathbb{R}$ are all zero, then the equation is a zero equation, and is a non-zero equation otherwise.

Suppose $s_1, s_2, \ldots, s_n \in \mathbb{R}$. If $x_i = s_i$ is a solution to the equation

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

for all $1 \leq i \leq n$. The set of all solutions to the equation is the solution set of the equation and an expression that gives all these solutions is the general solution for the equation.

1.1.3 Geometric Interpretation of Solutions in $\mathbb{R}^2$ and $\mathbb{R}^3$

Lines can either be parallel or non-parallel. Consider two lines, l_1 and l_2, with the equations

$$l_1 : y = m_1x + c_1 \text{ and } l_2 : y = m_2x + c_2.$$

Parallel but non-intersecting: We have $m_1 = m_2$, so both lines have the same gradient and do not intersect. There are no solutions.

Parallel but intersecting: We have $m_1 = m_2$ and $c_1 = c_2$, so both lines have the same equation. There are infinitely many solutions.

Non-parallel: The lines intersect at a point. There exists only one solution. The coordinates of the point of intersection are

$$\left(\frac{c_2 - c_1}{m_1 - m_2}, \frac{m_1c_2 - m_2c_1}{m_1 - m_2}\right).$$

Next, we consider the intersection of 2 planes in xyz-space, namely P_1 and P_2. They are given by

$$P_1 : a_1x + b_1y + c_1z = d_1 \text{ and } P_2 : a_2x + b_2y + c_2z = d_2,$$

where $a_1, b_1, c_1 \neq 0$ and $a_2, b_2, c_2 \neq 0$. There are three cases to consider.

Case 1: No solutions

The system of equations has no solutions if and only if P_1 and P_2 are different but parallel planes.

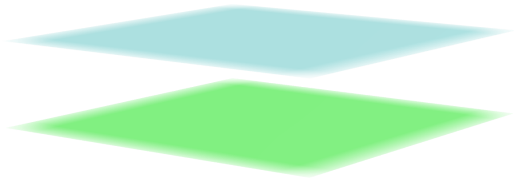

Figure 2: Geometric Interpretation of Case 1

Case 2: Infinitely many solutions (common line)

The system of equations has infinitely many solutions if and only if P_1 and P_2 intersect at a line.

Figure 3: Geometric Interpretation of Case 2

Case 3: Infinitely many solutions (common plane)

The system of equations has infinitely many solutions if and only if P_1 and P_2 are the same plane.

Figure 4: Geometric Interpretation of Case 3

Next, we consider the possible interactions between 3 planes in the xyz-space, where the third plane, P_3, has equation

$$P_3 : a_3x + b_3y + c_3z = d_3,$$

where for each i, a_i, b_i, c_i are not all zero. Here, we consider 8 cases.

For Cases 1, 2 and 3, the three normal lines to their respective planes are collinear. For Cases 4 and 5, only two of the normal lines are collinear. Lastly, for Cases 6, 7 and 8, none of the normal lines are collinear.

Case 1: No solutions
Here, P_1, P_2 and P_3 are parallel and distinct. It is clear that there are no points of intersection, and hence no solutions to the system of linear equations. Just to add on, a system of linear equations with no solutions is *inconsistent*.

Figure 5: Geometric Interpretation of Case 1

Case 2: No solutions
Without a loss of generality, suppose any two of the planes (say P_1 and P_2) are coincident, but the third plane (in this case is P_3) is parallel but not lying on the intersection of P_1 and P_2. Hence, it is clear that there are no solutions as well.

Figure 6: Geometric Interpretation of Case 2

Case 3: Infinitely many solutions
The three planes P_1, P_2 and P_3 are coincident so there are infinitely many solutions to the system of linear equations.

Figure 7: Geometric Interpretation of Case 3

Case 4: No solutions
Suppose two planes, say P_1 and P_2, are parallel. A third plane, P_3, cuts the other two such that P_3 is not parallel to P_1 (and thus P_2 too). Hence, the system of equations is inconsistent, implying that there are no solutions.

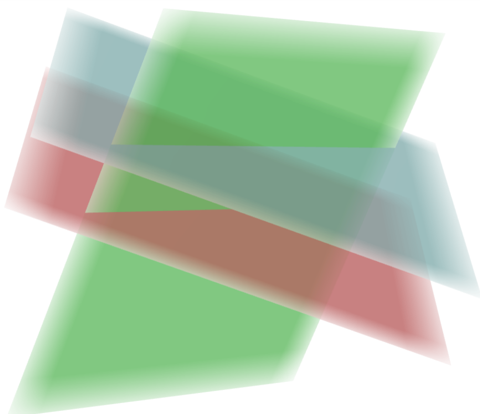

Figure 8: Geometric Interpretation of Case 4

Case 5: Infinitely many solutions (line)
Suppose P_1 and P_2 are coincident. P_3 slices P_1 and P_2 in such a way that P_3 is not parallel to the other two planes. Hence, all three planes intersect at a line, implying that there are infinitely many solutions.

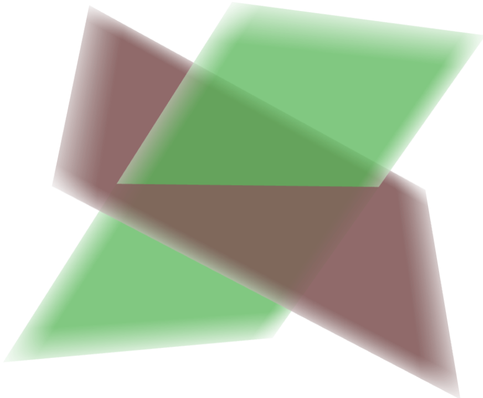

Figure 9: Geometric Interpretation of Case 5

Case 6: No solutions
Now, the normal lines of the planes are coplanar and the planes intersect in pairs (i.e. P_1 and P_2 intersect, but do not intersect with P_3, and vice versa). The system of solutions is clearly inconsistent as there are no points of intersection.

Geometrically, the inner triangular figure formed by the interactions between P_1, P_2 and P_3 is known as a triangular prism and resembles the shape of a very famous chocolate bar! Make a guess!

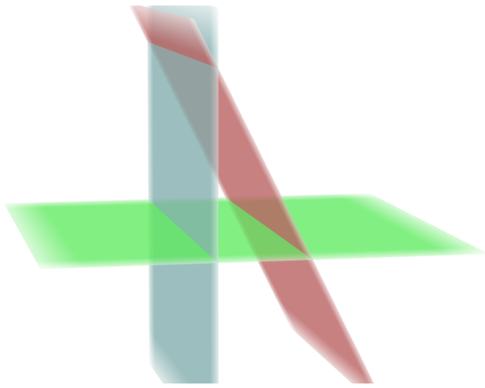

Figure 10: Geometric Interpretation of Case 6

Case 7: Infinitely many solutions (line)
The normal lines are coplanar and the planes intersect each other at a line. Thus, the system of equations has infinitely many solutions.

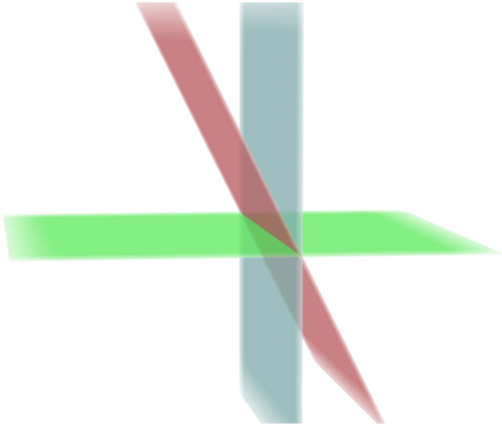

Figure 11: Geometric Interpretation of Case 7

Case 8: One solution
Now, the normal lines are not coplanar. The intersection of the three planes is a point.

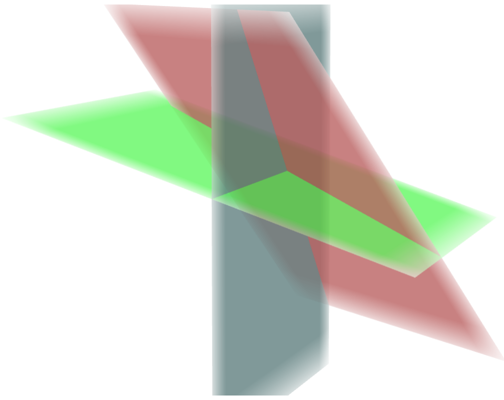

Figure 12: Geometric Interpretation of Case 8

1.2 Elementary Row Operations (EROs)

A system of linear equations with n variables and m equations

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

can be represented by a rectangular array of numbers as shown:

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_n \end{array}\right)$$

The above matrix is called the augmented matrix of the system.

The augmented matrix is analogous to the conventional methods of solving a system of linear equations by substitution and/or elimination.

Example: Consider the following system of equations:

$$\begin{aligned} 2x + 3y &= 8 \\ 4x - 5y &= -6 \end{aligned}$$

Solution: To solve it, observe that two times of the first equation yields $4x + 6y = 16$. By considering this equation and the second equation, $(4x + 6y) - (4x - 5y) = 16 - (-6)$, implying that $y = 2$, and hence $x = 1$.

We can express the above equation as an augmented matrix too. The matrix representation is

$$\left(\begin{array}{cc|c} 2 & 3 & 8 \\ 4 & -5 & -6 \end{array}\right)$$

Let the first row and second row be denoted by R_1 and R_2 respectively. In a similar fashion as before, we perform the operation $-2R_1 + R_2 \to R_2$, which means that the new second row is produced by taking the sum of $-2R_1$ and R_2. We write the operation in the following manner:

$$\left(\begin{array}{cc|c} 2 & 3 & 8 \\ 4 & -5 & -6 \end{array}\right) \xrightarrow{-2R_1+R_2\to R_2} \left(\begin{array}{cc|c} 2 & 3 & 8 \\ 0 & -11 & -22 \end{array}\right)$$

Hence, we have eliminated the term in x in R_2. For the new second row, we have $-11y = -22$, implying that $y = 2$, which is the same as what we obtained previously. □

Now, we state the three elementary row operations, also known as EROs. When solving a system of linear equations, the three techniques are
(i): multiplying an equation by a non-zero constant
(ii): interchanging the two equations
(iii): adding a multiple of one equation to another

In terms of augmented matrix, these correspond to the following EROs respectively:
(i): multiplying a row by a non-zero constant (i.e. $kR_1 \to R_1$, where $k \in \mathbb{R}$)
(ii): interchanging two rows (i.e. $R_1 \leftrightarrow R_2$)
(iii): adding a multiple of one row to another row (i.e. $-4R_1 + R_2 \to R_2$)

In the example above, we made use of the first and third EROs.

1.2.1 Row Equivalence

Two augmented matrices are row equivalent if one can be obtained from the other by a series of EROs. If augmented matrices of two systems of linear equations are row equivalent, then the two systems have the same set of solutions. The proof hinges on the idea of elementary matrices, which will be discussed in due course.

1.2.2 Gaussian Elimination and Row-Echelon Form (REF)

An augmented matrix, or any matrix in general, is in row-echelon form, or REF, if
(1): Any rows consisting entirely of zeros are grouped together at the bottom of the matrix.
(2): In any two successive rows that do not consist entirely of zeros, the first non-zero number in the lower row occurs further to the right than the first non-zero number in the higher row. The first non-zero number in a row is the *leading entry*, or the *pivot* of the row.

For example, the matrices $\mathbf{A}$ and $\mathbf{B}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 & 3 & 4 \\ 0 & 2 & 5 & 6 & 3 \\ 0 & 0 & 0 & 4 & 0 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} -1 & 2 & 4 \\ 0 & 0 & 2 \end{pmatrix},$$

are in REF.

For a matrix in REF, a column is called a pivot column if it contains a pivot point; otherwise, it is a non-pivot column.

The next part deals with Gaussian Elimination, which is named after Carl Friedrich Gauss, also known as the Prince of Mathematics. He proved the Fundamental Theorem of Algebra in his doctoral thesis in 1799. The theorem states that any polynomial of degree n with complex coefficients has n complex roots (considering multiplicity).

Let $\mathbf{A}$ and $\mathbf{R}$ be row-equivalent augmented matrices. If $\mathbf{R}$ is in REF, then $\mathbf{R}$ is an REF of $\mathbf{A}$ and $\mathbf{A}$ has an REF, which is $\mathbf{R}$.

Gaussian Elimination Process

The Gaussian Elimination is the following algorithm:
Step 1: Locate the leftmost column that does not consist entirely of zeros.
Step 2: Interchange the top row with another row, if necessary, to bring a non-zero entry to the top of the column found in Step 1.
Step 3: For each row below the top row, add a suitable multiple of the top row to it so that the entry below the leading entry of the top row becomes zero.
Step 4: Cover the top row of the matrix and repeat with Step 1 applied to the submatrix that remains. Continue until the entire matrix is in REF.

The steps seem complicated but let us go through an example to reduce a matrix to its REF. Consider the matrix

$$\begin{pmatrix} 9 & 8 & 1 & 2 & -3 \\ 1 & 2 & 0 & 0 & 2 \\ 2 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}.$$

Solution: The leading entry of the second row is 1, so it would be better to swap rows 1 and 2 in order to make the row operations process more efficient.

$$\begin{pmatrix} 9 & 8 & 1 & 2 & -3 \\ 1 & 2 & 0 & 0 & 2 \\ 2 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 9 & 8 & 1 & 2 & -3 \\ 2 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}$$

$$\xrightarrow[-2R_1+R_3 \to R_3]{-9R_1+R_2 \to R_2} \begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & -10 & 1 & 2 & -21 \\ 0 & -3 & 1 & 0 & -4 \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}$$

$$\xrightarrow{-\frac{10}{3}R_3+R_2 \to R_2} \begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & \frac{10}{3} & 2 & -\frac{23}{3} \\ 0 & -3 & 1 & 0 & -4 \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & -3 & 1 & 0 & -4 \\ 0 & 0 & \frac{10}{3} & 2 & -\frac{23}{3} \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}$$

The matrix

$$\begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & -3 & 1 & 0 & -4 \\ 0 & 0 & \frac{10}{3} & 2 & -\frac{23}{3} \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}$$

is in REF. Observe the numbers in red, which are the leading entries of the respective rows. □

1.2.3 Gauss-Jordan Elimination and Reduced Row-Echelon Form (RREF)

An augmented matrix is in reduced row-echelon form (RREF) and has the following additional properties:
(3): The leading entry of every non-zero row is 1
(4): In each pivot column, except the pivot point, all other entries are zero

For example, the matrices $\mathbf{C}$ and $\mathbf{D}$, where

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 2 & 3 & 4 \\ 0 & 1 & 5 & 6 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \text{ and } \mathbf{D} = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

are in RREF since each leading entry is 1.

Consider the matrix $\mathbf{J}$, where

$$\mathbf{J} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

It is clear that $\mathbf{J}$ is neither in REF or RREF.

We make a remark about the five matrices $\mathbf{A}, \mathbf{B}$ (mentioned under REF), $\mathbf{C}, \mathbf{D}$ and $\mathbf{J}$.

Matrix	REF	RREF
A	Yes	No
B	Yes	No
C	Yes	Yes
D	Yes	Yes
J	No	No

For Gauss-Jordan Elimination (named after Carl Gauss and Camille Jordan), we reduce it to REF first. Then, we adopt the following procedures to reduce it to an RREF.

Gauss-Jordan Elimination Process

The Gaussian-Jordan Elimination is a continuation of the Gaussian Elimination. The only difference is that the former requires these two additional steps:
Step 5: Multiply by a suitable constant to each row so that all the leading entries become 1.
Step 6: Beginning with the last non-zero row and working upwards, add suitable multiples of each row to the rows above to introduce zeros above the leading entries.

Previously, we performed EROs on the matrix

$$\begin{pmatrix} 9 & 8 & 1 & 2 & -3 \\ 1 & 2 & 0 & 0 & 2 \\ 2 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}$$

to reduce it to its row-echelon form. We can express the statement as such:

$$\begin{pmatrix} 9 & 8 & 1 & 2 & -3 \\ 1 & 2 & 0 & 0 & 2 \\ 2 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix} \xrightarrow{\text{REF}} \begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & -3 & 1 & 0 & -4 \\ 0 & 0 & \frac{10}{3} & 2 & -\frac{23}{3} \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix}$$

To further reduce the matrix to its RREF, we have to ensure that the leading entries -3, $\frac{10}{3}$ and 2 must be 1. Hence,

$$\begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & -3 & 1 & 0 & -4 \\ 0 & 0 & \frac{10}{3} & 2 & -\frac{23}{3} \\ 0 & 0 & 0 & 2 & 3 \end{pmatrix} \xrightarrow[R_3 \div \frac{10}{3} \to R_3,\ R_4 \div 2 \to R_4]{R_2 \div (-3) \to R_2} \begin{pmatrix} 1 & 2 & 0 & 0 & 2 \\ 0 & 1 & -\frac{1}{3} & 0 & \frac{4}{3} \\ 0 & 0 & 1 & \frac{3}{5} & -\frac{23}{10} \\ 0 & 0 & 0 & 1 & \frac{3}{2} \end{pmatrix}$$

As all the leading entries are 1, we are almost to getting the matrix to its RREF! All that is left is to ensure that the leading coefficient of the second, third and fourth rows are always strictly to the right of the leading coefficient of the row above it. Hence, the RREF is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \frac{23}{15} \\ 0 & 1 & 0 & 0 & \frac{7}{30} \\ 0 & 0 & 1 & 0 & -\frac{32}{10} \\ 0 & 0 & 0 & 1 & \frac{3}{2} \end{pmatrix}.$$

We consider the following augmented matrices representing systems of linear equations:

$$\mathbf{A} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 7 & 7 \end{array}\right)$$

Solution: $\mathbf{A}$ is the augmented matrix of a system of linear equations in 3 variables, say x_1, x_2 and x_3. It is clear that $x_1 = 1$, $x_2 = 5$ and $7x_3 = 7 \implies x_3 = 1$. $\square$

$$\mathbf{B} = \left(\begin{array}{cccc|c} 5 & 3 & -2 & 0 & 4 \\ 0 & 1 & 2 & 1 & 8 \\ 0 & 0 & 0 & 3 & 6 \end{array}\right)$$

B is the augmented matrix of a system of linear equations in 4 variables, say x_1, x_2, x_3 and x_4. As $3x_4 = 6 \implies x_4 = 2$, we only need to solve the following system of equations:

$$\begin{aligned} 5x_1 + 3x_2 - 2x_3 &= 4 \\ x_2 + 2x_3 &= 6 \end{aligned}$$

There are 2 equations but 3 unknowns. We make a remark before solving the system of linear equations.

REMARK

Suppose a system of linear equations has m equations and n variables. If $m < n$, then the system is underdetermined. If $m > n$, then the system is overdetermined.

For the case where $m < n$, we introduce a term known as the degree of freedom (df). This term also appears in Statistics when one is conducting a χ^2-test. In Linear Algebra, when solving a system of linear equations with more variables than unknowns, the degree of freedom is defined by the number of independent variables which must be specified to uniquely determine a solution. It is the following formula: $df = n - m$.

Solution: As there are more variables than equations, we set $x_3 = \lambda$, where $\lambda \in \mathbb{R}$ is a *free variable*. From here, $x_2 = 6 - 2\lambda$ and thus, $x_1 = -\frac{14}{5} + \frac{8}{5}\lambda$. Hence,

$$\begin{aligned} x_1 &= -\frac{14}{5} + \frac{8}{5}\lambda \\ x_2 &= 6 - 2\lambda \\ x_3 &= \lambda \\ x_4 &= 2 \end{aligned}$$

This implies that the system of equations has infinitely many solutions due to the presence of the parameter λ. Note that $n - m = 4 - 3 = 1$, implying that the system has 1 degree of freedom. Moreover, the method of finding the values of x_1, x_2, x_3 and x_4 is known as back substitution. $\square$

We consider the last matrix, **C**, where

$$\mathbf{C} = \left(\begin{array}{ccc|c} 2 & 1 & 6 & 4 \\ 0 & 0 & 0 & -4 \\ -2 & -2 & 0 & 6 \end{array}\right)$$

Solution: From the second row, it is clear that $0x_1 + 0x_2 + 0x_3 = -4 \implies 0 = -4$, which is a contradiction! Thus, we say that the system is inconsistent. $\square$

1.3 Applications

1.3.1 Macroeconomics: An Introduction to Input-Output (IO) Analysis

Wassily Leontief was a Soviet-American economist known for his research on input-output (IO) analysis and how changes in one economic sector may affect other sectors. In 1973, he was awarded the Nobel Memorial Prize in Economic Sciences for the development of this theory.

The IO model depicts inter-industry relationships within an economy, showing how output from one industrial sector may become an input to another industrial sector. In the inter-industry matrix, column entries represent inputs to an industrial sector, while row entries represent outputs from a given sector. This shows how dependent each sector is on every other sector, both as a customer of outputs from other sectors and as a supplier of inputs. Sectors may also depend internally on a portion of their own production as delineated by the entries of the matrix diagonal. Each column of the IO matrix shows the monetary value of inputs to each sector and each row represents the value of each sector's outputs.

Suppose an economy is divided into n sectors. We introduce a matrix, $\mathbf{C}$, also known as the consumption matrix. The consumption matrix shows the quantity of inputs needed to produce one unit of a good. Its columns represent the value of goods demanded from each sector per unit output. Let $\mathbf{x}$ be the product produced, where each entry of $\mathbf{x}$ denotes he monetary value of the output of sector i. Let $\mathbf{d}$ be the vector denoting the final demand, representing the value of goods demanded from the non-productive part of the economy.

Putting everything together, for the total demand to balance production,

$$\mathbf{x} = \mathbf{C}\mathbf{x} + \mathbf{d} \text{ or } (\mathbf{I} - \mathbf{C})\mathbf{x} = \mathbf{d}.$$

If $\mathbf{I} - \mathbf{C}$ is invertible, then there is a unique equilibrium output level (i.e. system of linear equations has a unique solution). In most scenarios, the column sums of $\mathbf{C}$ are less than 1, in which case $\mathbf{I} - \mathbf{C}$ is invertible.

Moreover, if the principal minors of $\mathbf{I} - \mathbf{A}$ are all positive, then the entries in $\mathbf{x}$ will be non-negative, making the output economically feasible. This can be verified using the identity

$$(\mathbf{I} - \mathbf{C}) \sum_{i=0}^{m} \mathbf{C}^i = \mathbf{I} - \mathbf{C}^{m+1},$$

where $\mathbf{C}^0 = \mathbf{I}$ and $\mathbf{I} - \mathbf{C}^{m+1} \to \mathbf{I}$ as $m \to \infty$. This provides us a way of approximating $(\mathbf{I} - \mathbf{C})^{-1}$.

1.3.2 Electrical Circuits: Kirchoff's Laws

Kirchhoff's Circuit Laws are two equations that deal with the current and potential difference in the lumped element model of electrical circuits.

Kirchoff's Current Law states that the algebraic sum of currents in a network of conductors meeting at a point is zero. In other words, for any node in an electrical circuit, the sum of currents flowing into that node is equal to the sum of currents flowing out of that node.

Kirchoff's Voltage Law states that the directed sum of the potential differences (or voltages) around any closed loop is zero.

Consider the complicated electrical circuit shown below. Note the currents flowing in respective sections of the circuit, namely $I_1, I_2, I_3, \ldots, I_7$.

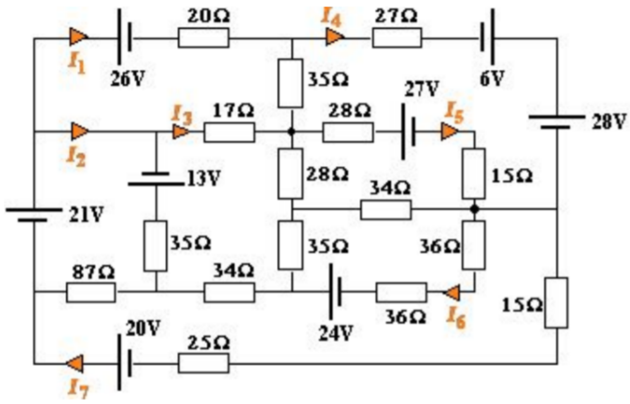

Figure 13: A complicated electrical circuit

We can form the following system of linear equations (those proficient in Physics should be able to obtain it):

$$\begin{aligned}
72I_1 - 17I_3 - 35I_4 &= -26 \\
122I_2 - 35I_3 - 87I_7 &= 34 \\
-87I_2 - 34I_3 - 72I_6 + 233I_7 &= -4 \\
-17I_1 - 35I_2 + 149I_3 - 28I_5 - 35I_6 - 34I_7 &= -13 \\
-28I_3 - 43I_4 + 105I_5 - 34I_6 &= -27 \\
-35I_3 - 34I_5 + 141I_6 - 72I_7 &= 24 \\
-35I_1 + 105I_4 - 43I_5 &= 5
\end{aligned}$$

We can form a matrix equation of the form $\mathbf{Ax} = \mathbf{b}$ representing the above data, where

$$\mathbf{A} = \begin{pmatrix} 72 & 0 & -17 & -35 & 0 & 0 & 0 \\ 0 & 122 & -35 & 0 & 0 & 0 & -87 \\ 0 & -87 & -34 & 0 & 0 & -72 & 233 \\ -171 & -35 & 149 & 0 & -28 & -35 & -34 \\ 0 & 0 & -28 & -43 & 105 & -36 & 0 \\ 0 & 0 & -35 & 0 & -34 & 141 & -72 \\ -35 & 0 & 0 & 105 & -43 & 0 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} -26 \\ 34 \\ -4 \\ -13 \\ -27 \\ 24 \\ 5 \end{pmatrix}.$$

We can solve for the values of $I_1, I_2, I_3, \ldots, I_7$ by noting that $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ or Gaussian Elimination.

Using matrices to solve for currents is more efficient than solely using physics formulas. A current can be found by an inverse matrix or EROs. The complexity of a circuit might be intimidating at first glance but through Linear Algebra, it can be understood by everyone.

1.3.3 Recreational Mathematics: Magic Square

A square array of numbers, usually positive integers, is called a magic square if the sums of the numbers in each row, each column, and both main diagonals are the same.

Suanfa tongzong is a mathematical text written by 16th century Chinese Mathematician Cheng Dawei published in 1592. One of the pages in the text shows a 9×9 magic square.

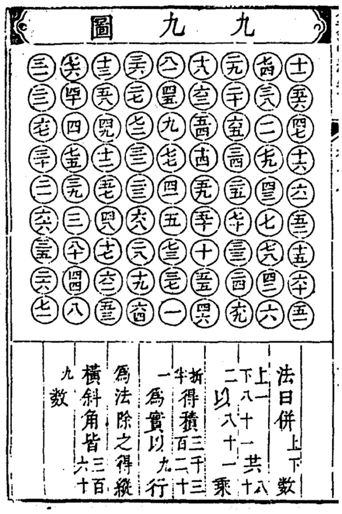

Figure 14: A page displaying a 9×9 magic square from Cheng Dawei's *Suanfa tongzong*

Suppose we have a 2×2 magic square with matrix representation

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

where the sum of rows, columns and diagonals add up to a constant, say s. That is, $a + b = s$, $c + d = s$, $a + c = s$, $b + d = s$, $a + d = s$ and $b + c = s$. This system has the unique solution $a = b = c = d = \frac{s}{2}$. Hence, the set of 2×2 magic squares is a one-dimensional subspace of $\mathcal{M}_{2\times 2}$.

For a 3×3 magic square, we consider the following matrix:

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

Hence, we have the following system of equations:

$$\begin{aligned} a + b + c &= s \\ d + e + f &= s \\ g + h + i &= s \\ a + d + g &= s \\ b + e + h &= s \\ c + f + i &= s \\ a + e + i &= s \\ c + e + g &= s \end{aligned}$$

Suppose $s = 20$. We can form the following matrix equation to represent the above data:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \\ i \end{pmatrix} = \begin{pmatrix} 20 \\ 20 \\ 20 \\ 20 \\ 20 \\ 20 \\ 20 \\ 20 \end{pmatrix}$$

The degree of freedom is 2, so we can generate infinitely many magic squares by setting h and i as our free variables. This is left as an exercise to the reader.

There is a magic square which is very special to me. It is Ramanujan's Magic Square, named after the brilliant Indian Mathematician Srinivasa Ramanujan. I came across it when I was 10 years old.

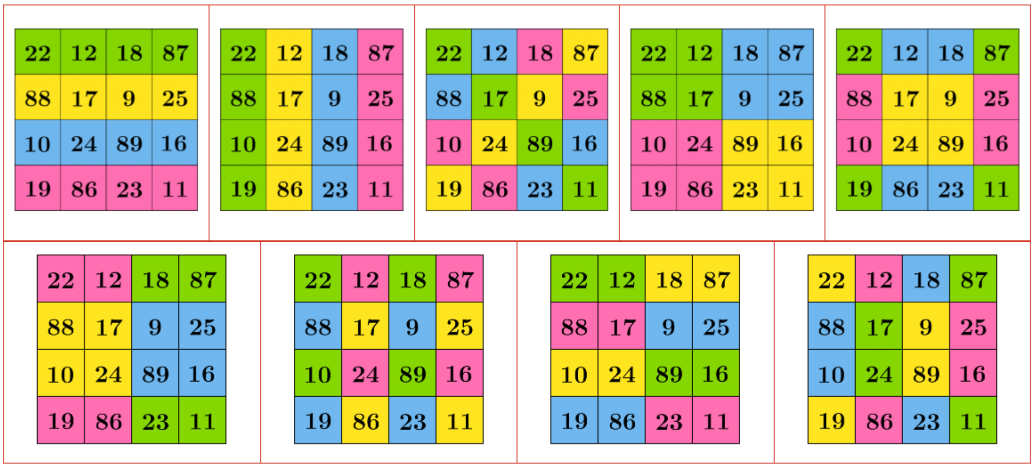

Figure 15: Ramanujan's Magic Square

Unlike any conventional magic square which satisfies the properties that the row sum, column sum and diagonal sum are all the same (reference to first three magic squares), Ramanujan's Magic Square has other surprising features. These are evident from the top right picture onward. The sum of numbers in the same-coloured square is always equal to one another! Furthermore, Ramanujan's birthday is on 22^{th} December 1887. Try to spot these feature in the magic square!

2 Matrices

A matrix is a rectangular array of numbers. The numbers in the array are known as the entries in the matrix. The size of a matrix is $m \times n$, where m is the number of rows and n is the number of columns. The (i, j)-entry of a matrix is the number in the i^{th} row and j^{th} column of a matrix.

A column matrix, or a column vector, is a matrix with one column. A row matrix, or a row vector, is a matrix with one row.

In general, an $m \times n$ matrix, $\mathbf{A}$, can be written as

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

or $\mathbf{A} = (a_{ij})_{m\times n}$, where a_{ij} is the (i, j)-entry of $\mathbf{A}$. If the size of the matrix is already known, then we can write it as $\mathbf{A} = (a_{ij})$.

2.1 Matrix Operations

Two matrices are equal if they have the same size and their corresponding entries are equal. That is, suppose $\mathbf{A} = (a_{ij})_{m\times n}$ and $\mathbf{B} = (b_{ij})_{p\times q}$. If $\mathbf{A}$ and $\mathbf{B}$ are equal, it implies that $m = p$, $n = q$ and $a_{ij} = b_{ij}$ for all i, j.

Suppose $\mathbf{A} = (a_{ij})_{m\times n}$ and $\mathbf{B} = (a_{ij})_{m\times n}$. We define matrix addition and matrix subtraction as follows:

$$\mathbf{A} \pm \mathbf{B} = (a_{ij} \pm b_{ij})_{m\times n}$$

That is,

$$\begin{aligned}\mathbf{A} \pm \mathbf{B} &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \pm \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \dots & a_{1n} \pm b_{1n} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \dots & a_{2n} \pm b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \pm b_{m1} & a_{m2} \pm b_{m2} & \dots & a_{mn} \pm b_{mn} \end{pmatrix}\end{aligned}$$

We can also multiply a matrix by a real scalar. Each entry of the matrix has to be multiplied by the scalar. Let $\lambda \in \mathbb{R}$. Then, λ is called the scalar and

$$\begin{aligned}c\mathbf{A} &= c\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \\ &= \begin{pmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{pmatrix}\end{aligned}$$

2.1.1 Matrix Multiplication

Given $\mathbf{A} = (a_{ij})_{m\times p}$ and $\mathbf{B} = (b_{ij})_{p\times n}$, the product $\mathbf{AB}$ is defined to be an $m \times n$ matrix whose (i,j)-entry is

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$$

for $1 \le i \le m$ and $1 \le j \le n$.

The product $\mathbf{AB}$ can be computed when the number of columns of $\mathbf{A}$ is equal to the number of rows of $\mathbf{B}$. We refer to $\mathbf{AB}$ as the pre-multiplication of $\mathbf{A}$ to $\mathbf{B}$ and $\mathbf{BA}$ as the post-multiplication of $\mathbf{A}$ to $\mathbf{B}$.

Matrix multiplication is not commutative. That is, in general, for matrices $\mathbf{A}$ and $\mathbf{B}$, the matrix products $\mathbf{AB}$ and $\mathbf{BA}$ are different even if the products exist. This idea will be talked about in greater detail in our discussion regarding the laws of matrix operations.

Consider the equation

$$(3x - 2y)(x + y) = 0,$$

which is clear that $3x - 2y = 0$ or $x + y = 0$. However, this does not apply to matrices! That is, if $\mathbf{AB} = \mathbf{0}$, where $\mathbf{0}$ is the zero matrix (which has zeros as all the entries), it does not necessarily imply that $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.

We now state some laws for matrix multiplication.

(1) Associative law for matrix multiplication:

If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are $m \times p$, $p \times q$ and $q \times n$ matrices respectively, then $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.

(2) Distributive laws for matrix addition and multiplication:

If $\mathbf{A}$, $\mathbf{B}_1$ and $\mathbf{B}_2$ are $m \times p$, $p \times n$ and $p \times n$ matrices respectively, then $\mathbf{A}(\mathbf{B}_1 + \mathbf{B}_2) = \mathbf{AB}_1 + \mathbf{AB}_2$.

If $\mathbf{A}$, $\mathbf{C}_1$ and $\mathbf{C}_2$ are $p \times n$, $m \times p$ and $m \times p$ matrices respectively, then $(\mathbf{C}_1 + \mathbf{C}_2)\mathbf{A} = \mathbf{C}_1\mathbf{A} + \mathbf{C}_2\mathbf{A}$.

2.1.2 Laws of Matrix Operations

Given a matrix $\mathbf{A}$, we normally use $-\mathbf{A}$ to denote the matrix $(-1)\mathbf{A}$.

Also, matrix subtraction can be defined using matrix addition. Given two matrices, $\mathbf{A}$ and $\mathbf{B}$, of the same size, then $\mathbf{A} - \mathbf{B}$ is defined to be $\mathbf{A} + (-\mathbf{B})$.

Now, we state some laws of matrix operations, which are believed to be trivial. Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be matrices of the same size and $c, d \in \mathbb{R}$.

(1) Commutative law for matrix addition: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

(2) Associative law for matrix addition: $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$

(3): $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$

(4): $(c + d)\mathbf{A} = c\mathbf{A} + d\mathbf{A}$

(5): $(cd)\mathbf{A} = c(d\mathbf{A}) = d(c\mathbf{A})$

(6): $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$

(7): $\mathbf{A} - \mathbf{A} = \mathbf{0}$

(8): $0\mathbf{A} = \mathbf{0}$ (note that 0 is the number zero and $\mathbf{0}$ is the zero matrix)

2.1.3 Power of Matrices

Let $\mathbf{A}$ be a square matrix and n be a non-negative integer. Define $\mathbf{A}^n$ as

$$\mathbf{A}^n = \begin{cases} \mathbf{I} & \text{if } n = 0 \\ \underbrace{\mathbf{AA}\ldots\mathbf{A}}_{n \text{ times}} & \text{if } n \ge 1. \end{cases}$$

A common misconception is some might raise each entry of the matrix to the power n, which is definitely false by the definition! That is, for instance,

$$\begin{pmatrix} 6 & 7 \\ 2 & -1 \end{pmatrix}^5 \neq \begin{pmatrix} 6^5 & 7^5 \\ 2^5 & (-1)^5 \end{pmatrix}.$$

For a square matrix $\mathbf{A}$ and non-negative integers m and n, we have $\mathbf{A}^m\mathbf{A}^n = \mathbf{A}^{m+n}$.

As matrix multiplication is not commutative, in general, $(\mathbf{AB})^2$ and $\mathbf{A}^2\mathbf{B}^2$ may be different.

You might wonder if there is an efficient method to compute a matrix when it is raised to a large power. For example, how do we find $\mathbf{A}^{100}$ quickly without repeatedly punching digits into a calculator? We will discuss this under the topic of diagonalisation where we have to make use of the eigenvalues and eigenvectors of a matrix.

2.2 Special Matrices

2.2.1 Square Matrix

A square matrix is a matrix with the same number of rows and columns. A square matrix of size $n \times n$ is of n.

2.2.2 Diagonal Matrix

For a square matrix $\mathbf{A} = (a_{ij})$ of order n, the diagonal of $\mathbf{A}$ is the sequence of entries $a_{11}, a_{22}, \ldots, a_{nn}$. Each entry a_{ii}, where $1 \leq i \leq n$ is a diagonal entry while a_{ij}, where $i \neq j$, is a non-diagonal entry.

Consider the following matrix, where the red entries are the diagonal entries and the black entries are the non-diagonal entries:

$$\begin{pmatrix} \color{red}{a_{11}} & a_{12} & \ldots & a_{1n} \\ a_{21} & \color{red}{a_{22}} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & \color{red}{a_{nn}} \end{pmatrix}$$

A square matrix is called a diagonal matrix if all its non-diagonal entries are zero. This means that $\mathbf{A}$ is a diagonal matrix if and only if $a_{ij} = 0$ whenever $i \neq j$.

2.2.3 Tridiagonal Matrix

A tridiagonal matrix has non-zero elements only on the main diagonal (i.e. diagonal entries are non-zero), the lower diagonal (the first diagonal below this; also known as sub diagonal) and the upper diagonal (the first diagonal above the main diagonal; also known as super diagonal).

Just to jump the gun, in relation to determinants, the determinant of a tridiagonal matrix can be computed quickly by establishing a suitable recurrence relation.

2.2.4 Scalar Matrix

A diagonal matrix is a scalar matrix if all of its diagonal entries are the same. That is, $\mathbf{A}$ is a scalar matrix if and only if

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

where c is a constant. For example, the following matrix is a scalar matrix:

$$\begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

2.2.5 Identity Matrix

A diagonal matrix is called an identity matrix if all of its diagonal entries are 1. Even though the identity matrix of order n is denoted by $\mathbf{I}_n$, we usually write it simply as $\mathbf{I}$ when there is no confusion involved. For example, the 2×2 and 3×3 identity matrices are denoted by

$$\mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

2.2.6 Zero Matrix

A matrix with all entries equal to zero is called the zero mtrix. $\mathbf{0}_{m \times n}$ is used to denote the zero matrix of size $m \times n$. However, similar to what was mentioned about identity matrices, we commonly write the zero matrix as $\mathbf{0}$ when there is no confusion involved.

2.2.7 Symmetric and Skew-Symmetric Matrices

A square matrix (a_{ij}) is symmetric if $a_{ij} = a_{ji}$ for all i, j. This is precisely the definition of the transpose of a matrix. Let $\mathbf{A}$ be a matrix (need not be square). Then, the transpose of $\mathbf{A}$ is denoted by $\mathbf{A}^{\mathrm{T}}$. The entries of $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}$ are a_{ij} and a_{ji} respectively.

Example: $\mathbf{M}$ and $\mathbf{N}$ are symmetric matrices since $\mathbf{M} = \mathbf{M}^{\mathrm{T}}$ and $\mathbf{N} = \mathbf{N}^{\mathrm{T}}$.

$$\mathbf{M} = \begin{pmatrix} 2 & 1 & 6 & -2 \\ 1 & 3 & 0 & -1 \\ 6 & 0 & 0 & 0 \\ -2 & -1 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{N} = \begin{pmatrix} 0 & 4 \\ 4 & 2 \end{pmatrix}.$$

A skew-symmetric matrix (a_{ij}) is skew-symmetric if $a_{ij} = -a_{ji}$ for all i, j.

Let $\mathbf{A}$ be a square matrix. Then, $\mathbf{A}$ is symmetric if and only if $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$, whereas $\mathbf{A}$ is skew-symmetric if and only if $\mathbf{A}^{\mathrm{T}} = -\mathbf{A}$.

Just like how every function can be expressed as a sum of an odd function and an even function, we have a similar result for matrices. That is, every square matrix $\mathbf{A}$ can be expressed as a sum of a symmetric matrix and a skew-symmetric matrix. In equation form, we have

$$\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}}) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}}).$$

Note that $\frac{1}{2}(\mathbf{A} + \mathbf{A}^{\mathrm{T}})$ is symmetric and $\frac{1}{2}(\mathbf{A} - \mathbf{A}^{\mathrm{T}})$ is skew-symmetric.

2.2.8 Upper and Lower Triangular Matrices

A triangular matrix is *upper* if all the entries in the lower diagonal are zero, and similarly, a triangular matrix is *lower* if all the entries in the upper diagonal are zero.

Hence, a square matrix (a_{ij}) is upper triangular if $a_{ij} = 0$ whenever $i > j$, and lower triangular if $a_{ij} = 0$ whenever $i < j$.

Example: Consider the following matrices $\mathbf{W}$ and $\mathbf{X}$:

$$\mathbf{W} = \begin{pmatrix} 3 & 2 & 6 & 5 \\ 0 & 7 & 7 & 6 \\ 0 & 0 & -4 & 3 \\ 0 & 0 & 0 & -9 \end{pmatrix} \text{ and } \mathbf{X} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & -7 & 0 \\ 2 & 10 & -0 \end{pmatrix}$$

$\mathbf{W}$ is an upper triangular matrix whereas $\mathbf{X}$ is a lower triangular matrix.

2.3 Block Matrices and Linear Systems

A block matrix is one that is broken into sections called blocks or submatrices. A matrix interpreted as a block matrix can be visualised as the original matrix with a collection of horizontal and vertical lines, which break it up, or partition it, into a collection of smaller matrices.

Given an $m \times p$ matrix $\mathbf{A}$ with q row partitions and s column partitions, we can express $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1s} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{q1} & \mathbf{A}_{q2} & \cdots & \mathbf{A}_{qs} \end{pmatrix}$$

and a $p \times n$ matrix $\mathbf{B}$ with s row partitions and r column partitions as

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1r} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{s1} & \mathbf{B}_{s2} & \cdots & \mathbf{B}_{sr} \end{pmatrix}$$

that are compatible with the partitions of $\mathbf{A}$. The product $\mathbf{C} = \mathbf{AB}$ can be performed blockwise, where $\mathbf{C}$ is an $m \times n$ matrix with q row partitions and r column partitions. The matrices in the resulting matrix $\mathbf{C}$ can be obtained by the following equation:

$$\mathbf{C}_{qr} = \sum_{i=1}^{s} \mathbf{A}_{qi}\mathbf{B}_{ir}$$

Example: the matrix $\mathbf{P}$, where

$$\mathbf{P} = \begin{pmatrix} 1 & 2 & 2 & 7 \\ 1 & 5 & 6 & 2 \\ 3 & 3 & 4 & 5 \\ 3 & 3 & 6 & 7 \end{pmatrix}$$

can be partitioned into four 2×2 blocks, namely

$$\mathbf{P}_{11} = \begin{pmatrix} 1 & 2 \\ 1 & 5 \end{pmatrix}, \ \mathbf{P}_{12} = \begin{pmatrix} 2 & 7 \\ 6 & 2 \end{pmatrix}, \ \mathbf{P}_{21} = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix} \text{ and } \mathbf{P}_{22} = \begin{pmatrix} 4 & 5 \\ 6 & 7 \end{pmatrix}.$$

Hence,

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix}.$$

Recall that matrices can be used to solve a system of linear equations. Given a matrix equation $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is the coefficient matrix, $\mathbf{x}$ is the variable matrix and $\mathbf{b}$ is the constant matrix. The solution to the unknown $\mathbf{x}$ is simply $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Consider the system of linear equations:

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

Note that $\mathbf{A}$, $\mathbf{x}$ and $\mathbf{b}$ denote the following matrices:

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

We can write $\mathbf{A}$ as $\begin{pmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n\end{pmatrix}$, where $\mathbf{c}_j$ is the j^{th} column of $\mathbf{A}$. Thus, the linear system can be also written as

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

or these forms too:

$$x_1\mathbf{c}_1+x_2\mathbf{c}_2+\ldots+x_n\mathbf{c}_n=\sum_{j=1}^{n}x_j\mathbf{c}_j=\mathbf{b}$$

Example: The system of linear equations

$$\begin{aligned}-x+y+2z&=3\\5x-y+6z&=16\\x+3y-5z&=10\end{aligned}$$

can be written as

$$x\begin{pmatrix}-1\\5\\1\end{pmatrix}+y\begin{pmatrix}1\\-1\\3\end{pmatrix}+z\begin{pmatrix}2\\6\\-5\end{pmatrix}=\begin{pmatrix}3\\16\\10\end{pmatrix}.$$

2.4 Matrix Transposition

Let $A=(a_{ij})$ be an $m\times n$ matrix. The transpose of $\mathbf{A}$, denoted by $\mathbf{A}^{\mathrm{T}}$, is the matrix whose (i,j)-entry is a_{ji}. That is, the rows of $\mathbf{A}$ become the columns of $\mathbf{A}^{\mathrm{T}}$ and vice versa. For example, a matrix, $\mathbf{A}$, and its transpose, $\mathbf{A}^{\mathrm{T}}$ are

$$\mathbf{A}=\begin{pmatrix}2&3&-3\\4&6&1\end{pmatrix}\text{ and }\mathbf{A}^{\mathrm{T}}=\begin{pmatrix}4&2\\6&3\\1&-3\end{pmatrix}.$$

Some basic properties of transpose are stated below. Let $\mathbf{A}$ be an $m\times n$ matrix. Then,
(1): $\left(\mathbf{A}^{\mathrm{T}}\right)^{\mathrm{T}}=\mathbf{A}$
(2): If $\mathbf{B}$ is an $m\times n$ matrix, then $(\mathbf{A}+\mathbf{B})^{\mathrm{T}}=\mathbf{A}^{\mathrm{T}}+\mathbf{B}^{\mathrm{T}}$
(3): If $c\in\mathbb{R}$, then $(c\mathbf{A})^{\mathrm{T}}=c\mathbf{A}^{\mathrm{T}}$
(4): If $\mathbf{B}$ is an $n\times p$ matrix, then $(\mathbf{AB})^{\mathrm{T}}=\mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$

The fourth property seems interesting so we shall provide a proof for it.

Proof: Recall that the (i,j)-entry of $\mathbf{AB}$ is the following sum:

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

When we take the transpose of $\mathbf{AB}$, the (i,j)-entry of $(\mathbf{AB})^{\mathrm{T}}$ becomes the (j,i)-entry of $\mathbf{AB}$, which is

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

Consider the entries b_{ki} and a_{jk}. Taking their respective transposes yield b_{ik} and a_{kj}, so for $1\le k\le n$, b_{ik} gives the (i,k)-entry of $\mathbf{B}^{\mathrm{T}}$ and a_{kj} gives the (k,j)-entry of $\mathbf{A}^{\mathrm{T}}$. $\square$

2.5 Inverse of Square Matrices

Let $a,b\in\mathbb{R}$, where $a\neq 0$. Then, the solution to the equation $ax=b$ is $x=\frac{b}{a}=a^{-1}b$. As there is no such thing as the division of matrices, to solve the matrix equation $\mathbf{Ax}=\mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are square matrices, we have a similar property to a^{-1} in the computation of the solution of $ax=b$ as mentioned earlier.

Let $\mathbf{A}$ be a square matrix of order n. Then, $\mathbf{A}$ is invertible if there exists a square matrix of order n such that

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

$\mathbf{B}$ is called the inverse of $\mathbf{A}$. If a square matrix does not have an inverse, it is singular. In relation to determinants, this also implies that $\det(\mathbf{A}) = 0$.

Hence, for a square matrix $\mathbf{A}$ which does not have an inverse, then
(1): $\mathbf{A}$ is not invertible
(2): $\mathbf{A}$ is singular
(3): $\det(\mathbf{A}) = 0$

To prove that a matrix is singular, one way is by establishing a contradiction.

Example: To show that

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

is singular, suppose on the contrary that $\mathbf{A}$ has an inverse, $\mathbf{B}$. That is,

$$\mathbf{B} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

By definition,

$$\mathbf{BA} = \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

but

$$\mathbf{BA} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a+b & 0 \\ c+d & 0 \end{pmatrix}.$$

Comparing the bottom right entry of $\mathbf{BA}$, both equations imply that $1 = 0$, which is a contradiction! Hence, $\mathbf{A}$ has to be singular. □

The cancellation laws for multiplication laws are as follows: for an invertible $m \times m$ matrix $\mathbf{A}$,
(1): if $\mathbf{B}_1$ and $\mathbf{B}_2$ are $m \times n$ matrices such that $\mathbf{AB}_1 = \mathbf{AB}_2$, then $\mathbf{B}_1 = \mathbf{B}_2$.
(2): if $\mathbf{C}_1$ and $\mathbf{C}_2$ are $n \times m$ matrices such that $\mathbf{C}_1\mathbf{A} = \mathbf{C}_2\mathbf{A}$, then $\mathbf{C}_1 = \mathbf{C}_2$.

Note that if $\mathbf{A}$ is not invertible, then the cancellation laws may not hold.

THEOREM

The inverse of a matrix is unique. That is, if $\mathbf{B}$ and $\mathbf{C}$ are inverses of a square matrix $\mathbf{A}$, then $\mathbf{B} = \mathbf{C}$.

Proof: Since $\mathbf{B}$ and $\mathbf{C}$ are inverses of $\mathbf{A}$, then

$$\mathbf{AB} = \mathbf{I},\ \mathbf{BA} = \mathbf{I},\ \mathbf{AC} = \mathbf{I} \text{ and } \mathbf{CA} = \mathbf{I}.$$

By considering $\mathbf{AB} = \mathbf{I}$, we have $\mathbf{CAB} = \mathbf{CI}$. Since $\mathbf{CAB}$=$\mathbf{IB}$, then $\mathbf{IB} = \mathbf{IC}$. We conclude that $\mathbf{B} = \mathbf{C}$. □

Some other properties regarding the inverse of a matrix are as follows: for invertible matrices $\mathbf{A}$ and $\mathbf{B}$ and a non-zero scalar c,
(1): $c\mathbf{A}$ is invertible and $(c\mathbf{A})^{-1} = \frac{1}{c}\mathbf{A}^{-1}$
(2):$\mathbf{A}^{\mathrm{T}}$ is invertible and $\left(\mathbf{A}^{\mathrm{T}}\right)^{-1} = \left(\mathbf{A}^{-1}\right)^{\mathrm{T}}$
(3): $\mathbf{A}^{-1}$ is invertible and $\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$
(4): $\mathbf{AB}$ is invertible and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
(5): $\mathbf{A}^r\mathbf{A}^s = \mathbf{A}^{r+s}$ for any $r, z \in \mathbb{Z}$

(6): $\mathbf{A}^n$ is invertible and $\left(\mathbf{A}^n\right)^{-1} = \left(\mathbf{A}^{-1}\right)^n$

We state a generalisation of the fourth property, which can be proven by induction. If $\mathbf{A}_1, \mathbf{A}_2, \ldots \mathbf{A}_k$ are invertible matrices, then the product $\mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_k$ is invertible and

$$\left(\mathbf{A}_1\mathbf{A}_2 \ldots \mathbf{A}_k\right)^{-1} = \mathbf{A}_k^{-1} \ldots \mathbf{A}_2^{-1}\mathbf{A}_1^{-1}.$$

Moreover, the fourth property reminds us of one of the properties of transpose due to its strong semblance. We state a proof for it.

Proof:

$$\begin{aligned}
(\mathbf{AB})^{-1} &= (\mathbf{AB})^{-1}\,\mathbf{I} \\
&= (\mathbf{AB})^{-1}\left(\mathbf{AA}^{-1}\right) \\
&= (\mathbf{AB})^{-1}\left(\mathbf{AIA}^{-1}\right) \\
&= (\mathbf{AB})^{-1}\left(\mathbf{ABB}^{-1}\mathbf{A}^{-1}\right) \\
&= (\mathbf{AB})^{-1}(\mathbf{AB})\left(\mathbf{B}^{-1}\mathbf{A}^{-1}\right) \\
&= \mathbf{I}\left(\mathbf{B}^{-1}\mathbf{A}^{-1}\right) \\
&= \mathbf{B}^{-1}\mathbf{A}^{-1}
\end{aligned}$$

and we are done. □

Example: If $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices of the same size, $\mathbf{A}+\mathbf{B}$ is invertible, we wish to prove that $\mathbf{A}^{-1}+\mathbf{B}^{-1}$ is invertible and

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{B}.$$

Solution:

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

This implies that $\mathbf{A}^{-1}+\mathbf{B}^{-1}$ is a product of invertible matrices, and hence, invertible. The second result can be proven by taking inverse of

$$\mathbf{A}^{-1}+\mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{A}+\mathbf{B})\mathbf{A}^{-1}$$

on both sides. □

2.5.1 Elementary Matrices

Consider the matrices $\mathbf{A}$ and $\mathbf{B}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 4 & -2 & 6 & 12 \\ 1 & 4 & 4 & 0 \end{pmatrix}.$$

What is the relationship between $\mathbf{A}$ and $\mathbf{B}$? From $\mathbf{A}$ to $\mathbf{B}$, the second row is multiplied by 2. That is, $R_2 \times 2 \to R_2$. By setting

$$\mathbf{E}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we have $\mathbf{E}_1\mathbf{A} = \mathbf{B}$. Observe that $\mathbf{E}_1$ is similar to the identity matrix, but the center entry is 2, which is a caused by the row operation 'second row of $\mathbf{A}$ multiplied by 2'.

An elementary matrix is a matrix which differs from the identity matrix by a single elementary row operation.

Pre-multiplication by an elementary matrix represents elementary row operations while post-multiplication represents elementary column operations. All elementary matrices are invertible and their inverses are also elementary matrices.

There are three types of elementary matrices, which correspond to three types of row operations, which are row switching, row multiplication and row addition.

Row Switching
A row within the matrix can be switched with another row. For a matrix $\mathbf{A}$ with m rows, if the i^{th} and j^{th} rows switch, then

$$R_i \leftrightarrow R_j \text{ for } i \neq j.$$

Let us state an example of row switching. Consider the matrices $\mathbf{C}$ and $\mathbf{D}$, where

$$\mathbf{C} = \begin{pmatrix} 2 & 1 & 3 & 4 \\ 4 & 2 & 4 & 5 \\ -1 & -1 & 3 & 0 \end{pmatrix} \text{ and } \mathbf{D} = \begin{pmatrix} -1 & -1 & 3 & 0 \\ 2 & 1 & 3 & 4 \\ 4 & 2 & 4 & 5 \end{pmatrix}.$$

This is slightly complicated. The first row of $\mathbf{C}$ becomes the second row of $\mathbf{D}$, the second row of $\mathbf{C}$ becomes the third row of $\mathbf{D}$ and the third row of $\mathbf{C}$ becomes the first row of $\mathbf{D}$. Hence, $\mathbf{EC} = \mathbf{D}$, where

$$\mathbf{E} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Row Multiplication
Each element in a row can be multiplied by a non-zero constant. It is also known as scaling a row. For a matrix $\mathbf{A}$ with m rows and $1 \leq i \leq m$,

$$R_i \times k \to R_i, \text{ where } k \neq 0.$$

Row Addition
A row can be replaced by the sum of that row and a multiple of another row. That is,

$$R_i + kR_j \to R_i \text{ where } i \neq j.$$

Let us state an example of row addition. Consider the matrices $\mathbf{P}$ and $\mathbf{Q}$, where

$$\mathbf{P} = \begin{pmatrix} 2 & 4 & 6 & 0 \\ 0 & 0 & 1 & 2 \\ 3 & 1 & 2 & 2 \end{pmatrix} \text{ and } \mathbf{Q} = \begin{pmatrix} 5 & 5 & 8 & 2 \\ 0 & 0 & 1 & 2 \\ 3 & 1 & 2 & 2 \end{pmatrix}.$$

Note that from $\mathbf{P}$ to $\mathbf{Q}$, we take the third row of $\mathbf{P}$ and add it to its first row. Hence, $\mathbf{EP} = \mathbf{Q}$, where

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The red entry is a resultant of adding the third row to the first.

Now that I have briefly explained the ideas of elementary matrices, recall that if two linear systems of equations $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{Cx} = \mathbf{d}$ have the same solutions, then the augmented matrices $(\mathbf{A}|\mathbf{b})$ and $(\mathbf{C}|\mathbf{d})$ are row equivalent. This can be proven by considering a product of elementary matrices since they are closely related to EROs and can reduce an augmented matrix to its REF or RREF.

THEOREM

We will state some equivalent statements. Let $\mathbf{A}$ be a square matrix. Then,
(1): $\mathbf{A}$ is invertible
(2): The linear system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution (i.e. only solution is $\mathbf{x} = \mathbf{0}$)
(3): The RREF of $\mathbf{A}$ is an identity matrix
(4): $\mathbf{A}$ can be expressed as a product of elementary matrices

Proof:
(1) $\implies$ **(2)** : Since $\mathbf{A}$ is invertible, then

$$\begin{aligned}\mathbf{Ax}=\mathbf{0} &\implies \mathbf{A}^{-1}\mathbf{Ax}=\mathbf{A}^{-1}\mathbf{0}\\ &\implies \mathbf{Ix}=\mathbf{0}\\ &\implies \mathbf{x}=\mathbf{0}\end{aligned}$$

This indicates that the system $\mathbf{Ax}=\mathbf{0}$ has only the trivial solution. □

(2) $\implies$ **(3)** : Suppose the system $\mathbf{Ax}=\mathbf{0}$ has only the trivial solution. The augmented matrix of the system is $(\mathbf{A}|\mathbf{0})$. Since the number of columns of $\mathbf{A}$ is equal to the number of rows in $\mathbf{A}$, then the RREF of the augmented matrix cannot have any zero rows, implying that the RREF is $(\mathbf{I}|\mathbf{0})$. □

(3) $\implies$ **(4)** : Since the RREF of $\mathbf{A}$ is $\mathbf{I}$, there exist elementary matrices $\mathbf{E}_1,\mathbf{E}_2,\ldots,\mathbf{E}_k$ such that

$$\mathbf{E}_k\ldots\mathbf{E}_2\mathbf{E}_1\mathbf{A}=\mathbf{I}.$$

Thus,

$$\mathbf{A}=(\mathbf{E}_k\ldots\mathbf{E}_2\mathbf{E}_1)^{-1}\mathbf{I}=(\mathbf{E}_k\ldots\mathbf{E}_2\mathbf{E}_1)^{-1}=\mathbf{E}_1^{-1}\mathbf{E}_2^{-1}\ldots\mathbf{E}_k^{-1}$$

and we are done. □

(4) $\implies$ **(1)** : As $\mathbf{A}$ is a product of elementary matrices and recall that elementary matrices are invertible. Thus, the result follows. □

This asserts that a square matrix is invertible if and only if its RREF is an identity matrix. This can be used to check if a square marix is invertible.

Example: The matrix

$$\begin{pmatrix}1&-1&3\\1&3&-3\\5&3&3\end{pmatrix}$$

is singular because its RREF is

$$\begin{pmatrix}1&0&\frac{3}{2}\\0&1&-\frac{3}{2}\\0&0&0\end{pmatrix}.$$

Observe the row of zeros in the third row, which implies that the matrix is singular.

THEOREM

Suppose $\mathbf{A}$ and $\mathbf{B}$ are square matrices of the same size. If $\mathbf{AB}=\mathbf{I}$, then $\mathbf{A}$ and $\mathbf{B}$ are both invertible, and

$$\mathbf{A}^{-1}=\mathbf{B},\ \mathbf{B}^{-1}=\mathbf{A}\text{ and }\mathbf{BA}=\mathbf{I}.$$

THEOREM

If $\mathbf{A}$ is singular. then $\mathbf{AB}$ and $\mathbf{BA}$ are singular.

We provide a proof for the second theorem.
Proof: Suppose $\mathbf{A}$ is singular. That is, for some $\mathbf{x}\neq\mathbf{0}$, we have $\mathbf{Ax}=\mathbf{0}$. Hence, $(\mathbf{BA})\,\mathbf{x}=\mathbf{B}\,(\mathbf{0})=\mathbf{0}$. Then, the coefficient matrix $\mathbf{BA}$ of the homogeneous linear system $(\mathbf{BA})\,\mathbf{x}=\mathbf{0}$ has a non-trivial solution, which concludes that $\mathbf{BA}$ is singular.

To prove that $\mathbf{AB}$ is singular, we consider two cases, namely if $\mathbf{B}$ is singular and $\mathbf{B}$ is non-singular. If $\mathbf{B}$ is singular, then for some $\mathbf{y}\neq\mathbf{0}$, we have $\mathbf{By}=\mathbf{0}$. Pre-multiplying both sides by $\mathbf{A}$, we have $(\mathbf{AB})\,\mathbf{y}=\mathbf{0}$ and the result follows. If $\mathbf{B}$ is non-singular, then it has an inverse (meaning $\mathbf{B}$ is invertible). Thus, $\mathbf{By}=\mathbf{x}$ and so $(\mathbf{AB})\,\mathbf{y}=\mathbf{Ax}=\mathbf{0}$. □

2.6 Determinants

2.6.1 Geometrical Interpretation

Before we formally introduce the determinant (whose definition is slightly complicated), we first explain what it means geometrically.

For two coplanar vectors in $\mathbb{R}^3$, namely $\begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$, recall from H2 Mathematics that the absolute value of the cross product of two vectors which have a common initial point gives the area of a parallelogram. That is, by considering a parallelogram $OABC$, its area is $|\mathbf{a} \times \mathbf{b}|$, where $\mathbf{a}$ and $\mathbf{b}$ are the vectors representing the line segments OA and OB in their respective directions. By setting $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$, we have

$$\text{area of parallelogram} = a_1 b_2 - a_2 b_1.$$

The cross product is only defined in three-dimensional space.

We extend this to constructing a three-dimensional figure, called a parallelepiped, using three vectors. A parallelepiped is a shape whose faces are all parallelograms.

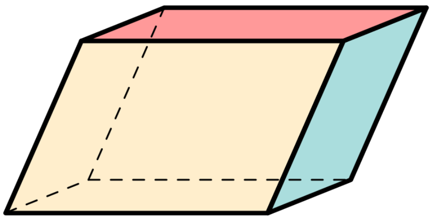

Figure 16: A parallelepiped

Intuitively, the volume (since it is a three-dimensional figure now) of the parallelepiped is the absolute value of the cross product of the three vectors with a common initial point. Suppose the three vectors are

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \text{ and } \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}.$$

Then,

$$\text{volume of parallelepiped} = |(\mathbf{a} \times \mathbf{b}) \bullet \mathbf{c}| = a_1 (b_2 c_3 - b_3 c_2) - b_1 (a_2 c_3 - a_3 c_2) + c_1 (a_2 b_3 - a_3 b_2).$$

We provide a proof for the fact that the volume of the parallelepiped is $|(\mathbf{a} \times \mathbf{b}) \bullet \mathbf{c}|$.
Proof: Note that $\mathbf{a} \times \mathbf{b}$ gives a vector normal to both $\mathbf{a}$ and $\mathbf{b}$. Let the angle between $\mathbf{c}$ and $\mathbf{a} \times \mathbf{b}$ be θ. Then, $|\mathbf{c}| \cos\theta = |\mathbf{a} \times \mathbf{b}|$. Since $|\mathbf{c}| \cos\theta$ is regarded as the *height* of the parallelepiped, then its volume is $|\mathbf{a} \times \mathbf{b}||\mathbf{c}| \cos\theta$, yielding the same conclusion as before. □

Now, what do the quantities $a_1 b_2 - a_2 b_1$ and $a_1 (b_2 c_3 - b_3 c_2) - b_1 (a_2 c_3 - a_3 c_2) + c_1 (a_2 b_3 - a_3 b_2)$ mean in relation to determinants?

2.6.2 Formal Introduction

For a 2×2 matrix $\mathbf{A}$ and 3×3 matrix $\mathbf{B}$, where

$$\mathbf{A} = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix},$$

their respective determinants, denoted by $\det(\mathbf{A})$ and $\det(\mathbf{B})$ respectively, are

$$\det(\mathbf{A}) = a_1b_2 - a_2b_1 \text{ and } \det(\mathbf{B}) = a_1\left(b_2c_3 - b_3c_2\right) - b_1\left(a_2c_3 - a_3c_2\right) + c_1\left(a_2b_3 - a_3b_2\right).$$

Definition of Determinant

Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix. Let $\mathbf{M}_{ij}$ be an $(n-1)\times(n-1)$ matrix obtained from $\mathbf{A}$ by deleting the i^{th} row and j^{th} column. Then, the determinant of $\mathbf{A}$ is defined as

$$\det(\mathbf{A}) = \begin{cases} a_{11} & \text{if } n = 1 \\ a_{11}A_{11} + a_{12}A_{12} + \ldots + a_{1n}A_{1n} & \text{if } n > 1, \end{cases}$$

where

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

is the (i,j)-cofactor of $\mathbf{A}$. The way the determinant is defined known as the method of cofactor expansion (also known as Laplace Expansion).

For an $n \times n$ matrix $\mathbf{A} = (a_{ij})$, $\det(\mathbf{A})$ is usually written as

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}.$$

The interested reader can read up on a more complicated definition of the determinant called Leibniz's Formula for Determinants. It will be studied under MA2101 though.

Example: We start off simple with a 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. To find $\mathbf{M}_{11}$, delete the first row and first column of $\mathbf{A}$, so we have the entry d. That is, $\mathbf{M}_{11} = (d)$ so $\det(\mathbf{M}_{11}) = d$. Next, we delete the first row and second column of $\mathbf{A}$ to get the entry c. Substituting everything ino the formula, the determinant of $\mathbf{A}$ is

$$\det(\mathbf{A}) = aA_{11} + bA_{12} = ad - bc.$$

Example: For a 3×3 matrix $\mathbf{B} = \begin{pmatrix} -3 & -2 & 4 \\ 4 & 3 & 1 \\ 0 & 2 & 4 \end{pmatrix}$, its determinant is

$$\begin{aligned} \det(\mathbf{B}) &= (-3)\det\begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix} - (-2)\det\begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix} + 4\det\begin{pmatrix} 4 & 3 \\ 0 & 2 \end{pmatrix} \\ &= 34 \end{aligned}$$

The same technique can be applied to 4×4 matrices and matrices of higher orders.

In general, for a 3×3 matrix $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, its determinant is

$$aei + bfg + cdh - ceg - afh - bdi.$$

We will later state a way to memorise this formula (only applies to 3×3 matrices) known as the Rule of Sarrus.

We will now state some properties of determinants.
(1): If $\mathbf{A}$ is a square matrix, then $\det(\mathbf{A}) = \det\left(\mathbf{A}^{\mathrm{T}}\right)$
Proof: Use mathematical induction and cofactor expansion. Left as an exercise to the reader. □

(2): The determinant of a square matrix with two identical rows is zero

(3): The determinant of a square matrix with two identical columns is zero

Proof: Use mathematical induction and cofactor expansion. The general case where the order of $\mathbf{A}$ is n is left as an exercise to the reader. We will provide a proof for 3×3 matrix with two identical rows. Let $\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ d & e & f \end{pmatrix}$. Expanding along the first column, we have

$$\begin{aligned} \det(\mathbf{A}) &= a \det \begin{pmatrix} e & f \\ e & f \end{pmatrix} - d \det \begin{pmatrix} b & c \\ e & f \end{pmatrix} + d \det \begin{pmatrix} b & c \\ e & f \end{pmatrix} \\ &= a \det \begin{pmatrix} e & f \\ e & f \end{pmatrix} \end{aligned}$$

which, of course, implies that the determinant is 0. □

(4): The determinant of a square matrix with either a row of zeros or a column of zeros is zero
Proof: This can be proven using cofactor expansion on the row/column containing the zeros. □

(5): A square matrix $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$
Proof: First, suppose $\mathbf{A}$ is invertible. Then, we wish to prove that $\det(\mathbf{A}) \neq 0$. Since $\mathbf{A}$ is invertible, then it can be expressed as a product of elementary matrices. Let $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_k$ be elementary matrices such that

$$\mathbf{B} = \mathbf{E}_k \ldots \mathbf{E}_2 \mathbf{E}_1 \mathbf{A}$$

is the RREF of $\mathbf{A}$. Then,

$$\det(\mathbf{B}) = \det(\mathbf{E}_k) \ldots \det(\mathbf{E}_2) \det(\mathbf{E}_1) \det(\mathbf{A})$$

and so $\mathbf{B} = \mathbf{I}$ and the result follows.

To prove $\det(\mathbf{A}) \neq 0 \implies \mathbf{A}$ is invertible, it would be better to prove the contrapositive statement. That is, if $\mathbf{A}$ is singular, then $\det(\mathbf{A}) = 0$. This is clearly true because $\mathbf{B}$ contains a row consisting entirely of zeros. Hence, $\det(\mathbf{B}) = 0$. Since $\det(\mathbf{E})_i \neq 0$ for all $1 \leq i \leq k$, then $\det(\mathbf{A}) = 0$. □

For two square matrices $\mathbf{A}$ and $\mathbf{B}$ of order n and a scalar c, then we establish properties 6 to 8.
(6): $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$
Proof: The matrix $c\mathbf{A}$ is obtained from $\mathbf{A}$ by multiplying c to every row of $\mathbf{A}$. □

(7): $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$
Proof: We consider 2 cases, namely if $\mathbf{A}$ is singular and if $\mathbf{A}$ is invertible. If $\mathbf{A}$ is singular, then $\mathbf{AB}$ is singular so $\det(\mathbf{AB}) = 0$, and the result follows. If $\mathbf{A}$ is invertible, it can be written as a product of elementary matrices. That is, $\mathbf{A} = \mathbf{E}_1 \mathbf{E}_2 \ldots \mathbf{E}_k$.

$$\begin{aligned} \det(\mathbf{AB}) &= \det(\mathbf{E}_1 \mathbf{E}_2 \ldots \mathbf{E}_k \mathbf{B}) \\ &= \det(\mathbf{E}_1) \det(\mathbf{E}_2) \ldots \det(\mathbf{E}_k) \det(\mathbf{B}) \\ &= \det(\mathbf{E}_1 \mathbf{E}_2 \ldots \mathbf{E}_k) \det(\mathbf{B}) \\ &= \det(\mathbf{A}) \det(\mathbf{B}) \end{aligned}$$

□

The transition from the first line to the second line uses a property regarding determinants involving elementary matrices which we will mention in the next section.

(8): If $\mathbf{A}$ is invertible, then

$$\det\left(\mathbf{A}^{-1}\right) = \frac{1}{\det(\mathbf{A})}.$$

Proof: Since $\mathbf{A}$ is invertible, then $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ and taking the determinants on both sides and doing a rearrangement of the equation (note that $\det(\mathbf{I}) = 1$), the result follows. $\square$

2.6.3 Influence of EROs on the Determinant of a Matrix

How EROs affect the Determinant of a Matrix

Let $\mathbf{A}$ and $\mathbf{B}$ be square matrices. Then, we have the following properties:
(1): If $\mathbf{B}$ is obtained from $\mathbf{A}$ by multiplying one row of $\mathbf{A}$ by a constant k, then $\det(\mathbf{B}) = k\det(\mathbf{A})$
(2): If $\mathbf{B}$ is obtained from $\mathbf{A}$ by interchanging two rows of $\mathbf{A}$, then $\det(\mathbf{B}) = -\det(\mathbf{A})$
(3): If $\mathbf{B}$ is obtained from $\mathbf{A}$ by adding a multiple of one row of $\mathbf{A}$ to another row, then $\det(\mathbf{B}) = \det(\mathbf{A})$
(4): Let $\mathbf{E}$ be an elementary matrix of the same size as $\mathbf{A}$. Then, $\det(\mathbf{E}\mathbf{A}) = \det(\mathbf{E})\det(\mathbf{A})$.

2.6.4 Rule of Sarrus

Consider a 3×3 matrix $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$.

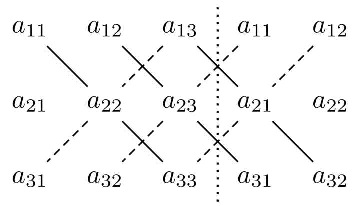

Figure 17: Rule of Sarrus

From the above figure, by taking the sum of the products of entries along the solid diagonals, we have

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}.$$

By taking the sum of the products of entries along the dotted diagonal, we have

$$a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}.$$

Then, the determinant of the matrix is

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

2.6.5 Sylvester's Determinant Theorem

Sylvester's Determinant Theorem

Sylvester's Determinant Theorem, also known as the Weinstein-Aronszajn Identity, states that if $\mathbf{A}$ and $\mathbf{B}$ matrices of the same order, say order m, and $\mathbf{A}$ is invertible, then

$$\det(\mathbf{I}_m + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_m + \mathbf{B}\mathbf{A}).$$

Proof: Starting from the left side,

$$\begin{aligned}\det(\mathbf{I}_m + \mathbf{AB}) &= \det(\mathbf{I_m})\det(\mathbf{I}_m + \mathbf{AB}) \\ &= \det(\mathbf{AA}^{-1})\det(\mathbf{I}_m + \mathbf{AB}) \\ &= \det(\mathbf{A})\det(\mathbf{A}^{-1})\det(\mathbf{I}_m + \mathbf{AB}) \\ &= \det(\mathbf{A}^{-1})\det(\mathbf{I}_m + \mathbf{AB})\det(\mathbf{A}) \\ &= \det(\mathbf{A}^{-1}\mathbf{I}_m + \mathbf{A}^{-1}\mathbf{AB})\det(\mathbf{A}) \\ &= \det(\mathbf{A}^{-1} + \mathbf{B})\det(\mathbf{A}) \\ &= \det(\mathbf{A}^{-1}\mathbf{A} + \mathbf{BA}) \\ &= \det(\mathbf{I}_m + \mathbf{BA})\end{aligned}$$

□

2.7 Methods to find Inverse of Square Matrices

2.7.1 Method of Cofactor Expansion

For an $n \times n$ matrix $\mathbf{A} = (a_{ij})$, $\det(\mathbf{A})$ can be expressed as a cofactor expansion using any row or column of $\mathbf{A}$. That is,

$$\begin{aligned}\det(\mathbf{A}) &= a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in} \text{ cofactor expansion along } i^{\text{th}} \text{ row} \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \ldots + a_{nj}A_{nj} \text{ cofactor expansion along } j^{\text{th}} \text{ column}\end{aligned}$$

for any $1 \le i \le n$ and $1 \le j \le n$. For those interested in the proof, the idea will be discussed in one of the chapters under MA2101, which hinges on multilinear forms and partity of a permutation.

Example: We consider a 3×3 matrix

$$\begin{pmatrix} 2 & 2 & 1 \\ -3 & 6 & 1 \\ 4 & 3 & 0 \end{pmatrix}.$$

Expanding along the second row yields

$$\det\begin{pmatrix} 2 & 2 & 1 \\ -3 & 6 & 1 \\ 4 & 3 & 0 \end{pmatrix} = 3\det\begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix} +6\det\begin{pmatrix} 2 & 1 \\ 4 & 0 \end{pmatrix} + -1\det\begin{pmatrix} 2 & 2 \\ 4 & 3 \end{pmatrix} = -31$$

whereas expanding along the third column yields

$$\det\begin{pmatrix} 2 & 2 & 1 \\ -3 & 6 & 1 \\ 4 & 3 & 0 \end{pmatrix} = 1\det\begin{pmatrix} -3 & 6 \\ 4 & 3 \end{pmatrix} -1\det\begin{pmatrix} 2 & 2 \\ 4 & 3 \end{pmatrix} + 0\det\begin{pmatrix} 2 & 2 \\ -3 & 6 \end{pmatrix} = -31$$

too. Hence, we claim that if $\mathbf{A}$ is a triangular matrix, then the determinant of $\mathbf{A}$ is equal to the product of the diagonal entries of $\mathbf{A}$. Without a loss of generality, we consider the case where $\mathbf{A}$ is upper triangular (of course, the theorem also applies to lower triangular matrices).

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nn} \end{pmatrix},$$

then

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

This can be proven using mathematical induction.

2.7.2 Adjoint

Define the adjoint (or adjugate) of an $n \times n$ matrix $\mathbf{A}$ by

$$\text{adj}(\mathbf{A}) = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1n} \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} A_{11} & A_{21} & \ldots & A_{n1} \\ A_{12} & A_{22} & \ldots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \ldots & A_{nn} \end{pmatrix},$$

where A_{ij} is the (i, j)-cofactor of $\mathbf{A}$.

THEOREM

If $\mathbf{A}$ is invertible, then

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

2.7.3 Augmented Matrix

We can perform EROs to transform an augmented matrix from $(\mathbf{A}|\mathbf{I})$ to $(\mathbf{I}|\mathbf{A}^{-1})$ since pre-multiplications of a matrix by elementary matrices correspond to performing EROs on the matrix.

Example: Suppose we wish to find the inverse of the matrix $\mathbf{J}$, where

$$\mathbf{J} = \begin{pmatrix} 1 & 9 & 0 \\ -5 & 5 & 6 \\ -3 & 4 & 2 \end{pmatrix}.$$

Solution: As $\det(\mathbf{J}) = -86$ which is non-zero, then $\mathbf{J}$ is invertible and hence, its inverse exists. We write the augmented matrix $(\mathbf{J}|\mathbf{I})$ and perform EROs until we obtain $\mathbf{I}$ on the left of the vertical bar.

$$\left(\begin{array}{ccc|ccc} 1 & 9 & 0 & 1 & 0 & 0 \\ -5 & 5 & 6 & 0 & 1 & 0 \\ -3 & 4 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow[3R_1+R_3 \to R_3]{5R_1+R_2 \to R_2} \left(\begin{array}{ccc|ccc} 1 & 9 & 0 & 1 & 0 & 0 \\ 0 & 50 & 6 & 5 & 1 & 0 \\ 0 & 31 & 2 & 3 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{-\frac{31}{50}R_2+R_3 \to R_3} \left(\begin{array}{ccc|ccc} 1 & 9 & 0 & 1 & 0 & 0 \\ 0 & 50 & 6 & 5 & 1 & 0 \\ 0 & 0 & -\frac{43}{25} & -\frac{1}{10} & -\frac{31}{50} & 1 \end{array}\right)$$

$$\xrightarrow{R_3 \div (-\frac{43}{25}) \to R_3} \left(\begin{array}{ccc|ccc} 1 & 9 & 0 & 1 & 0 & 0 \\ 0 & 50 & 6 & 5 & 1 & 0 \\ 0 & 0 & 1 & \frac{5}{86} & \frac{31}{86} & -\frac{25}{43} \end{array}\right)$$

$$\xrightarrow{-6R_3+R_2 \to R_2} \left(\begin{array}{ccc|ccc} 1 & 9 & 0 & 1 & 0 & 0 \\ 0 & 50 & 0 & \frac{200}{43} & -\frac{50}{43} & \frac{150}{43} \\ 0 & 0 & 1 & \frac{5}{86} & \frac{31}{86} & -\frac{25}{43} \end{array}\right)$$

$$\xrightarrow{R_2 \div 50 \to R_2} \left(\begin{array}{ccc|ccc} 1 & 9 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{4}{43} & -\frac{1}{43} & \frac{3}{43} \\ 0 & 0 & 1 & \frac{5}{86} & \frac{31}{86} & -\frac{25}{43} \end{array}\right)$$

$$\xrightarrow{-9R_2+R_1 \to R_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{7}{43} & \frac{9}{43} & -\frac{27}{43} \\ 0 & 1 & 0 & \frac{4}{43} & -\frac{1}{43} & \frac{3}{43} \\ 0 & 0 & 1 & \frac{5}{86} & \frac{31}{86} & -\frac{25}{43} \end{array}\right)$$

Hence, $\mathbf{J}^{-1} = \begin{pmatrix} \frac{7}{43} & \frac{9}{43} & -\frac{27}{43} \\ \frac{4}{43} & -\frac{1}{43} & \frac{3}{43} \\ \frac{5}{86} & \frac{31}{86} & -\frac{25}{43} \end{pmatrix}$. □

2.8 Applications

2.8.1 Efficient Method to solve a System of Linear Equations: Cramer's Rule

Cramer's Rule can be used to solve a system of linear equations.

Suppose $\mathbf{Ax} = \mathbf{b}$ is a linear system where $\mathbf{A}$ is an $n \times n$ matrix. Let $\mathbf{A}_i$ be the matrix obtained from $\mathbf{A}$ by replacing the i^{th} column of $\mathbf{A}$ by $\mathbf{b}$. If $\mathbf{A}$ is invertible, then the system has only one solution. That is,

$$\mathbf{x} = \frac{1}{\det(\mathbf{A})}\begin{pmatrix} \det(\mathbf{A}_1) \\ \det(\mathbf{A}_2) \\ \vdots \\ \det(\mathbf{A}_n) \end{pmatrix}.$$

Proof: Set $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$. Then, as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, by replacing $\mathbf{A}^{-1}$ with $\frac{1}{\det(\mathbf{A})}\operatorname{adj}(\mathbf{A})$, then

$$\mathbf{x} = \frac{1}{\det(\mathbf{A})}(\operatorname{adj}(\mathbf{A}))\mathbf{b},$$

where

$$x_i = \frac{b_1A_{1i} + b_2A_{2i} + \ldots + b_nA_{ni}}{\det(\mathbf{A})} = \frac{\det(\mathbf{A}_i)}{\det(\mathbf{A})}$$

for $1 \le i \le n$. $\square$

For the case where $n = 3$, we let $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, where $\mathbf{A}$ is the matrix representation to a system of linear equations in x, y and z. That is, $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$. Then, Cramer's Rule states that

$$x = \frac{\det\begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix}}{\det(\mathbf{A})},\ y = \frac{\det\begin{pmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{pmatrix}}{\det(\mathbf{A})} \text{ and } z = \frac{\det\begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{pmatrix}}{\det(\mathbf{A})}$$

2.8.2 Art of Polynomials: The Lagrange Interpolation and Vandermonde Matrix

We start off simple by asking the following question:

Given three points with distinct x-coordinates, how do we find a quadratic curve which passes through them?

Suppose the coordinates are $(-1, 3)$, $(2, 5)$ and $(3, 4)$ and the quadratic equation is of the form $f(x) = ax^2+bx+c$. Substituting $f(-1) = 3$, $f(2) = 5$ and $f(3) = 4$, we obtain the following matrix equation:

$$\begin{pmatrix} 1 & -1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 4 \end{pmatrix}.$$

Solve the equation however you want to get $a = -\frac{5}{12}$, $b = \frac{13}{12}$ and $c = \frac{9}{2}$. As the inverse of the matrix is unique, we only have one solution to the matrix equation (this piece of information is very important for the general case with $n + 1$ data points). That is, the quadratic equation passing through the three points is

$$y = -\frac{5}{12}x^2 + \frac{13}{12}x + \frac{9}{2}.$$

Lagrange polynomials are studied in Numerical Analysis and they are used for polynomial interpolation. For a given set of points x_j, y_j with no two x_j values equal, the Lagrange polynomial is the polynomial of lowest degree that assumes at each value x_j the corresponding value y_j.

Solving an interpolation problem leads to a problem in Linear Algebra amounting to inversion of a matrix. We use a special matrix called a square Vandermonde matrix to help us solve such a problem.

For an $n \times n$ Vandermonde Matrix, $\mathbf{V}$, we can write it as

$$\mathbf{V} = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ 1 & x_3 & x_3^2 & \ldots & x_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{pmatrix}.$$

The matrix equation depicting a scenario where a curve passes through $n+1$ points is

$$\begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n} \\ 1 & x_3 & x_3^2 & \ldots & x_3^{n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n+1} & x_{n+1}^2 & \ldots & x_{n+1}^{n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n+1} \end{pmatrix}.$$

The determinant of an $n \times n$ Vandermonde Matrix is

$$\det(\mathbf{V}) = \prod_{1 \le i < j \le n} (x_j - x_i).$$

Proof: We will use the method of cofactor expansion. By subtracting to each column, the preceding column multiplied by x_1, the determinant is unchanged. We will obtain the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & x_2 - x_1 & x_2(x_2 - x_1) & x_2^2(x_2 - x_1) & \cdots & x_2^{n-2}(x_2 - x_1) \\ 1 & x_3 - x_1 & x_3(x_3 - x_1) & x_3^2(x_3 - x_1) & \cdots & x_3^{n-2}(x_3 - x_1) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n - x_1 & x_n(x_n - x_1) & x_n^2(x_n - x_1) & \cdots & x_n^{n-2}(x_n - x_1) \end{pmatrix}.$$

Performing cofactor expansion on the first row yields the following result:

$$\begin{aligned} \det(\mathbf{V}) &= \det \begin{pmatrix} x_2 - x_1 & x_2(x_2 - x_1) & x_2^2(x_2 - x_1) & \ldots & x_2^{n-2}(x_2 - x_1) \\ x_3 - x_1 & x_3(x_3 - x_1) & x_3^2(x_3 - x_1) & \ldots & x_3^{n-2}(x_3 - x_1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n - x_1 & x_n(x_n - x_1) & x_n^2(x_n - x_1) & \ldots & x_n^{n-2}(x_n - x_1) \end{pmatrix} \\ &= (x_2 - x_1)(x_3 - x_1)\ldots(x_n - x_1) \det \begin{pmatrix} 1 & x_2 & x_2^2 & \ldots & x_2^{n-2} \\ 1 & x_3 & x_3^2 & \ldots & x_3^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-2} \end{pmatrix} \\ &= \prod_{1 < j \le n} (x_j - x_1) \det(\mathbf{V}'), \end{aligned}$$

where $\mathbf{V}'$ is also a Vandermonde Matrix but of a smaller order. Repeatedly applying this process will yield the determinant formula as the product of all $x_j - x_i$ such that $i < j$. $\square$

Recall that $x_i \neq x_j$ as we regard them as distinct x-coordinates. Thus, $x_j - x_i \neq 0$, and so $\det(\mathbf{V}) \neq 0$, implying that $\mathbf{V}^{-1}$ exists.

3 Vector Spaces

Before the introduction to Euclidean n-spaces, I would not talk about the geometric definitions of vector addition and subtraction, as well as multiplication by a scalar as these should be covered in O- and A-Level Mathematics. The same can be said for coordinate systems where we deal with vectors on an xy-plane or vectors in xyz-space.

3.1 Euclidean n-Space

An n-vector or ordered n-tuple of real numbers has the form

$$(u_1, u_2, \ldots, u_i, \ldots, u_n)$$

where $u_1, u_2, \ldots, u_n$ are real numbers. The number u_i in the i^{th} position of an n-vector is called the i^{th} component or the i^{th} coordinate of the n-vector.

We state some properties.
(1): $\mathbf{u}$ and $\mathbf{v}$ are equal if and only if $u_i = v_i$ for all $1 \leq i \leq n$
(2): The addition $\mathbf{u} + \mathbf{v}$ of $\mathbf{u}$ and $\mathbf{v}$ is defined by

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

(3): Let $c \in \mathbb{R}$. The scalar multiple $c\mathbf{u}$ of $\mathbf{u}$ is defined by

$$c\mathbf{u} = (cu_1, cu_2, \ldots, cu_n).$$

(4): The n-vector $(0, 0, \ldots, 0)$ is the *zero vector* and it is denoted by 0
(5): Define the negative of $\mathbf{u}$ to be $(-1)\mathbf{u}$ and denote it by $-\mathbf{u}$. That is,

$$-\mathbf{u} = (-u_1, -u_2, \ldots, -u_n).$$

(6): Similar to property **(2)**, the subtraction $\mathbf{u} - \mathbf{v}$ of $\mathbf{v}$ from $\mathbf{u}$ is defined by $\mathbf{u} + (-\mathbf{v})$. That is,

$$\mathbf{u} - \mathbf{v} = (u_1 - v_1, u_2 - v_2, \ldots, u_n - v_n).$$

We can identify n-vectors $(u_1, u_2, \ldots, u_n)$ with a $1 \times n$ matrix called a row vector or an $n \times 1$ matrix called a column vector.

Other properties are as follows: Let $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ be n-vectors and $c, d \in \mathbb{R}$. Then,
(1): $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutative law for addition)
(2): $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (associative law for addition)
(3): $\mathbf{u} + \mathbf{0} = \mathbf{u} = \mathbf{0} + \mathbf{u}$ (existence of an additive identity)
(4): $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (existence of additive inverse)
(5): $c(d\mathbf{u}) = (cd)\mathbf{u}$ (associative law for multiplication)
(6): $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ (distributive law)
(7): $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$ (distributive law)
(8): $1\mathbf{u} = \mathbf{u}$ (existence of a multiplicative identity)

The set of all n-vectors of real numbers is called the Euclidean n-space or simply, the n-space. $\mathbb{R}$ is used to denote the set of all real nummbers and $\mathbb{R}^n$ is used to denote the Euclidean n-space.

Hence, $\mathbf{u} \in \mathbb{R}^n$ if and only if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ for some $u_1, u_2, \ldots, u_n \in \mathbb{R}$.

In set notation, the above can be represented by

$$\mathbb{R}^n = \{(u_1, u_2, \ldots, u_n) | u_1, u_2, \ldots, u_n \in \mathbb{R}\}.$$

3.2 Subsets of $\mathbb{R}^n$

Example: Let

$$B = \{(u_1, u_2, u_3, u_4) | u_1 = 0 \text{ and } u_2 = u_4\}.$$

It means that B is a subset of $\mathbb{R}^4$ such that $(u_1, u_2, u_3, u_4) \in B$ if and only if $u_1 = 0$ and $u_2 = u_4$. For example, $(0, \pi, \pi, \pi) \in B$ but $(0, 1, 3, 2) \notin B$. In general, we can write

$$B = \{(0, a, b, a) | a, b \in \mathbb{R}\}.$$

If a system of linear equations has n variables, then its solution set is a subset of $\mathbb{R}^n$.

Example: The solution set of the linear equation

$$\begin{aligned} x + y + z &= 0 \\ x - y + 2z &= 1 \end{aligned}$$

can be expressed implicitly as

$$\{(x, y, z) | x + y + z = 0 \text{ and } x - y + 2z = 1\}$$

or explicitly in terms of a free variable t, where $t \in \mathbb{R}$. That is,

$$\left\{ \frac{1}{2} - \frac{3}{2}t, -\frac{1}{2} + \frac{1}{2}t, t \,\middle|\, t \in \mathbb{R} \right\}.$$

3.2.1 Solution Sets for Lines and Planes

Now, we are going to discuss how to express lines in $\mathbb{R}^2$ and $\mathbb{R}^3$, as well as planes in $\mathbb{R}^3$, which is merely revisiting H2 Mathematics content.

Lines in $\mathbb{R}^2$

A line in $\mathbb{R}^2$ can be expressed implicitly in set notation by

$$\{(x, y) | ax + by = c\},$$

where a, b and c are real constants and a and b are both non-zero. If $a \neq 0$, then the line can be expressed as

$$\left\{ \left(\frac{c - bt}{a}, t \right) \middle| \, t \in \mathbb{R} \right\},$$

whereas if $b \neq 0$, then the line can be expressed as

$$\left\{ \left(t, \frac{c - at}{b} \right) \middle| \, t \in \mathbb{R} \right\}.$$

Planes in $\mathbb{R}^3$

A plane in $\mathbb{R}^3$ can be expressed implicitly in set notation by

$$\{(x, y, z) | ax + by + cz = d\},$$

where a, b, c and d are real constants and a, b and c are not all zero. Explictly, the plane can be expressed in terms of two free variables, say s and t, where $s, t \in \mathbb{R}$. They are namely the following ways:

$$\begin{aligned} &\left\{ \left(\frac{d - bs - ct}{a}, s, t \right) \middle| \, s, t \in \mathbb{R} \right\} \text{ if } a \neq 0, \\ &\left\{ \left(s, \frac{d - as - ct}{b}, t \right) \middle| \, s, t \in \mathbb{R} \right\} \text{ if } b \neq 0, \\ &\left\{ \left(s, t, \frac{d - as - bt}{c} \right) \middle| \, s, t \in \mathbb{R} \right\} \text{ if } c \neq 0. \end{aligned}$$

Lines in $\mathbb{R}^3$

A line in $\mathbb{R}^3$ cannot be regarded by a single equation as in the case of $\mathbb{R}^2$. Instead, it can be regarded as the intersection of two non-parallel planes and hence, written implictly as

$$\{(x,y,z)|a_1x+b_1y+c_1z=d_1 \text{ and } a_2x+b_2y+c_2z=d_2\}$$

for some suitable choice of constants a_1,b_1,c_1,d_1 and a_2,b_2,c_2,d_2.

A line in $\mathbb{R}^3$ is usually represented explicitly in set notation by

$$\{(a_0,b_0,c_0)+t(a,b,c)|t\in\mathbb{R}\},$$

where (a_0,b_0,c_0) is a point on the line and (a,b,c) is the direction of the line.

3.2.2 Finite Sets

Let S be a finite set. Then, $|S|$ is used to denote the number of elements contained in S. We call $|S|$ the cardinality of S.

Example: If

$$S_1=\{1,2,3,\ldots,n+1\},$$

then the cardinality of S, or $|S|$, is just $n+1$.

3.3 Linear Combination and Span

3.3.1 Linear Combination

Linear Combination

Let $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k$ be vectors in $\mathbb{R}^n$. For any real numbers $c_1,c_2,\ldots,c_k$, the vector

$$\sum_{i=1}^{k}c_i\mathbf{u}_i=c_1\mathbf{u}_1+c_2\mathbf{u}_2+\ldots+c_k\mathbf{u}_k$$

is a linear combination of $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k$.

We can test whether a vector $\mathbf{v}$ is a linear combination of $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k$ by forming a system of linear equations. If the system has a solution, then $\mathbf{v}$ is indeed a linear combination of all the $\mathbf{u}_i$'s.

Example: We set $\mathbf{u}_1=\begin{pmatrix}1\\-4\\3\end{pmatrix}$, $\mathbf{u}_2=\begin{pmatrix}1\\2\\6\end{pmatrix}$, $\mathbf{u}_3=\begin{pmatrix}2\\1\\-3\end{pmatrix}$ and $\mathbf{v}=\begin{pmatrix}4\\7\\5\end{pmatrix}$ and ask if $\mathbf{v}$ is a linear combintion of $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$. That is, does there exist c_1,c_2 and c_3 such that

$$c_1\begin{pmatrix}1\\-4\\3\end{pmatrix}+c_2\begin{pmatrix}1\\2\\6\end{pmatrix}+c_3\begin{pmatrix}2\\1\\-3\end{pmatrix}=\begin{pmatrix}4\\7\\5\end{pmatrix}?$$

Solution: We can rewrite this as a matrix equation with the column vector $\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix}$ as our variable. Then,

$$\begin{pmatrix}1&1&2\\-4&2&1\\3&6&-3\end{pmatrix}\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix}=\begin{pmatrix}4\\7\\5\end{pmatrix}.$$

Note that $\det\begin{pmatrix}1&1&2\\-4&2&1\\3&6&-3\end{pmatrix}\neq 0$, which implies that $\begin{pmatrix}1&1&2\\-4&2&1\\3&6&-3\end{pmatrix}$ has an inverse, and there exist c_1,c_2 and c_3 satisfying the matrix equation. Thus, we conclude that $\mathbf{v}$ can be written as a linear combination of $\mathbf{u}_1$, $\mathbf{u}_2$

and $\mathbf{u}_3$. To find the values of $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$, it will be left as an exercise. $\square$

Example: We state a more obvious example. Now, we set $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 1 \\ 3 \\ 10 \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 2 \\ 1 \\ 20 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} 4 \\ 6 \\ 33 \end{pmatrix}$ and ask if $\mathbf{w}$ is a linear combintion of $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$.

Solution: One can observe that $\mathbf{w} = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3$, which implies that $\mathbf{w}$ is a linear combination of $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_3$. Alternatively, one could manually compute the inverse of the coefficient matrix to solve for the scalars c_1, c_2 and c_3. $\square$

Example: Now, we state a case where we only have two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ and ask if $\mathbf{x}$ can be written as a linear combination of them. Suppose $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 1 \\ 3 \\ 10 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix}$.

Solution: We can create the following system of linear equations:

$$\begin{aligned} c_1 + c_2 &= 7 \\ 2c_1 + 3c_2 &= 2 \\ 3c_1 + 10c_2 &= 1 \end{aligned}$$

Solving the first two equations yields $c_1 = 19$ and $c_2 = -12$. However, substituting into the third equation, we have $3(19) + 10(-12) = 1 \implies -63 = 1$, which is a contradiction. Thus, $\mathbf{x}$ cannot be written as a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$. $\square$

3.3.2 Standard Basis

Not to be confused with standard base at Stuff'd, the standard basis of a coordinate vector space is the set of vectors whose components are all zero, except one that equals 1. The standard basis for the three-dimensional space $\mathbb{R}^3$ is formed by the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \text{ and } \mathbf{e}_3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

For any $(x, y, z) \in \mathbb{R}^3$, note that

$$(x, y, z) = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3.$$

Hence, every vector in $\mathbb{R}^3$ is a linear combination of $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{e}_3$. $\mathbf{e}_1, \mathbf{e}_2$ and $\mathbf{e}_3$ are known as the directional vectors of the x-axis, y-axis and z-axis respectively.

3.3.3 Span

Span

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a set of vectors in $\mathbb{R}^n$. The set of all linear combinations of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$

$$\{c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k | c_1, c_2, \ldots, c_k \in \mathbb{R}\}$$

is called a linear span of S and is denoted by $\text{span}(S)$ or $\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$.

Example: From the previous examples,

$$\mathbf{v} = \begin{pmatrix} 4 \\ 7 \\ 5 \end{pmatrix} \in \text{span}\left\{ \begin{pmatrix} 1 \\ -4 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 6 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \right\}$$

but

$$\mathbf{x} = \begin{pmatrix} 7 \\ 2 \\ 1 \end{pmatrix} \notin \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 10 \end{pmatrix} \right\}$$

Now, we will state several examples related to span.

Example: Let $S = \{(1,0,0,-1),(0,1,1,0)\} \subseteq \mathbb{R}^4$. Then, every element in S can be expressed as

$$\lambda(1,0,0,-1)+\mu(0,1,1,0)=(\lambda,\mu,\mu,-\lambda),$$

where $\lambda,\mu \in \mathbb{R}$. Thus, span$(S)$ comprises vectors of the form $(\lambda,\mu,\mu,-\lambda)$, where $\lambda,\mu \in \mathbb{R}$.

One question which comes to one's mind is when is $\operatorname{span} S = \mathbb{R}^n$.

THEOREM

Let $S = \{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k\} \subseteq \mathbb{R}^n$, where $\mathbf{u}_i = (a_{i1},a_{i2},\ldots,a_{in})$ for $1 \le i \le k$. For any vector $\mathbf{v} = (v_1,v_2,\ldots,v_n) \in \mathbb{R}^n$, $\mathbf{v} \in \operatorname{span}(S)$ if and only if the vector equation

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

has a solution for $c_1,c_2,\ldots,c_k$, meaning that the following system of linear equations is consistent:

$$\begin{aligned}
a_{11}c_1+a_{21}c_2+\ldots+a_{k1}c_k&=v_1\\
a_{12}c_1+a_{22}c_2+\ldots+a_{k2}c_k&=v_2\\
&\vdots\\
a_{1n}c_1+a_{2n}c_2+\ldots+a_{kn}c_k&=v_n
\end{aligned}$$

THEOREM

Let $\mathbf{A} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_k\end{pmatrix}$.
(i): If an REF of $\mathbf{A}$ has no zero row, the system is always consistent regardless the values of $v_1,v_2,\ldots,v_n$ and span$(S)=\mathbb{R}^n$.
(ii): If an REF of $\mathbf{A}$ has at least one zero row, the system is not always consistent and span$(S) \neq \mathbb{R}^n$.

Note that if $k<n$, then span$(S) \neq \mathbb{R}^n$. In particular, we have the following results:

REMARK

(i): one vector cannot span $\mathbb{R}^2$
(ii): one or two vectors cannot span $\mathbb{R}^3$
(iii): $\mathbf{0} \in \operatorname{span}(S)$
(iv): For any $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_r \in \operatorname{span}(S)$ and $c_1,c_2,\ldots,c_r \in \mathbb{R}$,

$$c_1\mathbf{v}_1+c_2\mathbf{v}_2+\ldots+c_r\mathbf{v}_r \in \operatorname{span}(S).$$

The proof for **(iii)** is trivial (I'm not kidding) so we will only provide a proof for **(iv)**.
Proof: Our objective is to show that

$$c_1\mathbf{v}_1+c_2\mathbf{v}_2+\ldots+c_r\mathbf{v}_r$$

is a linear combination of $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k$. Rewriting $\mathbf{v}_i$ as

$$a_{i1}\mathbf{u}_1+a_{i2}\mathbf{u}_2+\ldots+a_{ik}\mathbf{u}_k,$$

where $1 \leq i \leq r$, we have

$$\begin{aligned} c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_r\mathbf{v}_r &= c_1(a_{11}\mathbf{u}_1 + a_{12}\mathbf{u}_2 + \ldots + a_{1k}\mathbf{u}_k) \\ &+ c_2(a_{21}\mathbf{u}_1 + a_{22}\mathbf{u}_2 + a_{2k}\mathbf{u}_k) \\ &+ \ldots \\ &+ c_r(a_{r1}\mathbf{u}_1 + a_{r2}\mathbf{u}_2 + \ldots + a_{rk})\mathbf{u}_k \\ &= (c_1a_{11} + c_2a_{21} + \ldots + c_ra_{r1})\mathbf{u}_1 \\ &+ (c_1a_{12} + c_2a_{22} + \ldots + c_ra_{r2})\mathbf{u}_2 \\ &+ \ldots \\ &+ (c_1a_{1k} + c_2a_{2k} + \ldots + c_ra_{rk})\mathbf{u}_k \end{aligned}$$

and the result follows. $\square$

Example: To show that $\text{span}\{(1,0,1),(1,1,0),(0,1,1)\} = \mathbb{R}^3$, we have to show that for any $(x,y,z) \in \mathbb{R}^3$ there exist $a,b,c \in \mathbb{R}$ such that

$$a(1,0,1) + b(1,1,0) + c(0,1,1) = (x,y,z).$$

Solution: This is slightly different from the usual questions we discussed on a system of linear equations in the past but the idea is to prove that this system is consistent for all $x,y,z \in \mathbb{R}$. Thus, we consider reducing the coefficient matrix to its REF. Note that REF would be sufficient since RREF has an extra condition that the leading entries must be 1.

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & 1 & 1 & y \\ 1 & 0 & 1 & z \end{array}\right) \xrightarrow{\text{REF}} \left(\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & 1 & 1 & y \\ 0 & 0 & 2 & z-x+y \end{array}\right)$$

As the system is consistent regardless the values of x,y and z, then the result follows. $\square$

THEOREM

Given two subsets S_1 and S_2 of $\mathbb{R}^n$, where $S_1 = \{\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k\}$ and $S_2 = \{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m\}$, $\text{span}(S_1) \subseteq \text{span}(S_2)$ if and only if each $\mathbf{u}_i$ is a linear combination of $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$.

Proof: The case where $\text{span}(S_1) \subseteq \text{span}(S_2) \implies$ each $\mathbf{u}_i$ is a linear combination of $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$ should be clear because $S_1 \subseteq \text{span}(S_1) \subseteq \text{span}(S_2)$.

Now, we prove that if $\mathbf{u}_i$ is a linear combination of $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_m$, then $\text{span}(S_1) \subseteq \text{span}(S_2)$. It is clear that $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_k \in \text{span}(S_2)$. Let $\mathbf{w}$ be any vector in $\text{span}(S_1)$. Then,

$$\mathbf{w} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k$$

for some $c_1,c_2,\ldots,c_k \in \mathbb{R}$. Hence, $\mathbf{w} \in \text{span}(S_2)$ and the result follows. $\square$

3.3.4 Equal Sets

To show that two sets A and B are equal (i.e. $A = B$), then we need to show that $A \subseteq B$ and $B \subseteq A$. Thus, we have to show that the elements in A are linear combinatons of elements in B and vice versa. One efficient method to prove this is by forming an augmented matrix.

Example: Suppose

$$A = \left\{\begin{pmatrix}1\\0\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\-1\\2\end{pmatrix}, \begin{pmatrix}2\\1\\-1\\4\end{pmatrix}\right\} \text{ and } B = \left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\-1\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\1\\-1\end{pmatrix}\right\}.$$

We wish to prove if $\text{span}(A) = \text{span}(B)$.

Solution: First, we check whether $\text{span}(A) \subseteq \text{span}(B)$. We form the following augmented matrix:

$$\left(\begin{array}{ccc|c|c|c} 1 & -1 & -1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & -1 & 1 & 0 & -1 & -1 \\ 1 & 1 & -1 & 1 & 2 & 4 \end{array}\right) \xrightarrow{\text{REF}} \left(\begin{array}{ccc|c|c|c} 1 & -1 & -1 & 1 & 0 & 2 \\ 0 & 2 & 2 & -1 & 1 & -1 \\ 0 & 0 & 2 & -1 & -1 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

and as all three systems are consistent, then our claim that $\text{span}(A) \subseteq \text{span}(B)$ is true.

Now, we check whether $\text{span}(B) \subseteq \text{span}(A)$. We can form an augmented matrix representing the information above and note that not all the systems are consistent and thus, our claim is false. This will be left as an exercise. Hence, $\text{span}(A) \neq \text{span}(B)$. □

3.3.5 Redundant Vectors

Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \in \mathbb{R}^n$. If $\mathbf{u}_k$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}$, then

$$\text{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}) = \text{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}, \mathbf{u}_k).$$

3.4 Subspaces

Subspaces

Let V be a subset of $\mathbb{R}^n$. Then, V is called a subspace of $\mathbb{R}^n$ if $V = \text{span}(S)$, where $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ for some $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \in \mathbb{R}^n$. We say that V is a subspace spanned by S or V is a subspace spanned by the vectors in S. Hence, S spans V.

Example: Let $V_1 = \{(a + 4b, a) | a, b \in \mathbb{R}\} \subseteq \mathbb{R}^2$. For any $a, b \in \mathbb{R}$, $(a + 4b, a) = a(1, 1) + b(4, 0)$. Thus, $V_1 = \text{span}\{(1, 1), (4, 0)\}$ is a subspace of $\mathbb{R}^2$.

Example: Let $V_2 = \{(x, y, z) | x + y - z = 0\} \subseteq \mathbb{R}^3$. Note that the system has two degrees of freedom as there is one equation and three unknowns. The equation $x + y - z = 0$ has a general solution

$$(x, y, z) = (-s + t, s, t) = s(-1, 1, 0) + t(1, 0, 1)$$

where $s, t \in \mathbb{R}$. Hence, $V_2 = \text{span}\{(-1, 1, 0), (1, 0, 1)\}$ is a subspace of $\mathbb{R}^3$.

Since the linear span of a set of vectors is the smallest linear subspace that contains the set, we state an alternative definition of subspace which was previously used for span.

Let V be a subspace of $\mathbb{R}^n$. We have the following results:
(i): $\mathbf{0} \in V$
(ii): For any $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ and $c_1, c_2, \ldots, c_r \in \mathbb{R}$,

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_r\mathbf{v}_r \in V.$$

It would actually be easier to prove the following three statements in order to show that a subset is a subspace of a vector space.

Proving that a Subset is a Subspace of a Vector Space

(i): $\mathbf{0} \in V$ (V is non-empty)
(ii): For any two vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$, their sum is also in V. That is, $\mathbf{v}_1 + \mathbf{v}_2 \in V$. (closure under addition)
(iii): Let $\alpha \in \mathbb{R}$ and $\mathbf{v} \in V$. Then, $\alpha\mathbf{v} \in V$. (closure under scalar multiplication)

As such, statements **(ii)** and **(iii)** will imply that a linear combination of vectors in V is also in V. Note that to disprove that a subset is a subspace of a vector space, a contradiction would suffice.

Example: We shall prove that $V_3 = \left\{(x, y, z)|x^2 \le y^2 \le z^2)\right\} \subseteq \mathbb{R}^3$ is not a subspace of $\mathbb{R}^3$.

Solution: As the square of any real number is non-negative, to provide the contradiction, we need to consider the polarity of numbers. Note that $(1, 1, 1), (1, 1, -1) \in V_4$. However, when we take the sum of these two vectors, we obtain $(2, 2, 0)$. However, this is impossible as $(x, y, z) = (2, 2, 0)$ does not satisfy the inequality $x^2 \le y^2 \le z^2$. Hence, the result follows. $\square$

Sum of Subspaces

Let V and W be subspaces of $\mathbb{R}^n$. Define

$$V + W = \{\mathbf{v} + \mathbf{w}|\mathbf{v} \in V \text{ and } \mathbf{w} \in W\}.$$

We shall prove that $V + W$ is also a subspace of $\mathbb{R}^n$.

Proof: Here, we are dealing with a sum of two subspaces. Let $V = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ and $W = \text{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$. Then,

$$\begin{aligned} V + W &= \{\mathbf{v} + \mathbf{w}|\mathbf{v} \in V \text{ and } \mathbf{w} \in W\} \\ &= \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_m\mathbf{v}_m + \beta_1\mathbf{w}_1 + \beta_2\mathbf{w}_2 + \ldots + \beta_n\mathbf{w}_n|\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \ldots, \beta_n \in \mathbb{R}\} \end{aligned}$$

which is the definition of $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$. $\square$

Intersection of Subspaces

Let V and W be subspaces of $\mathbb{R}^n$. Then, $V \cap W$ is a subspace of $\mathbb{R}^n$.

Proof: Since both V and W contain the zero vector, then the zero vector is also contained in $V \cap W$ and hence, $V \cap W$ is non-empty.

Let $\mathbf{u}$ and $\mathbf{v}$ be any two vectors in $V \cap W$, and let α and β be any real numbers. Since $\mathbf{u}$ and $\mathbf{v}$ are contained in V, then $\alpha\mathbf{u} + \beta\mathbf{v}$ is contained in V. The same argument can be used to prove that $\alpha\mathbf{u} + \beta\mathbf{v}$ is contained in W. $\square$

Union of Subspaces

Let V and W be subspaces of $\mathbb{R}^n$. Then, $V \cup W$ is a subspace of $\mathbb{R}^n$ if and only if $V \subseteq W$ or $W \subseteq V$.

Proof: If $V \subseteq W$ or $W \subseteq V$, then $V \cup W$ is a subspace of $\mathbb{R}^n$.

Next, suppose $V \nsubseteq W$ and $V \cup W$. We wish to prove that $W \subseteq V$. Take any vector $\mathbf{x} \in W$. Since $V \nsubseteq W$, there exists a vector $\mathbf{y} \in V$ but $\mathbf{y} \nsubseteq W$. As $\mathbf{x}, \mathbf{y} \in V \cup W$, then $\mathbf{x} + \mathbf{y} \in V \cup W$. This means that $\mathbf{x} + \mathbf{y} \in V$ or $\mathbf{x} + \mathbf{y} \in W$.

Assume that $\mathbf{x} + \mathbf{y} \in W$. As W is a subspace of $\mathbb{R}^n$ and $-\mathbf{x} \in W$, then $\mathbf{y} = (\mathbf{x} + \mathbf{y}) + (-\mathbf{x}) \in W$, which contradicts that $\mathbf{y} \nsubseteq W$. Hence, $\mathbf{x} + \mathbf{y} \in V$. As V is a subspace of $\mathbb{R}^n$ and $-\mathbf{y} \in V$, then $\mathbf{x} = (\mathbf{x} + \mathbf{y}) + (-\mathbf{y}) \in V$. Since every vector in W is contained in V, the result follows. $\square$

REMARK

In general, the union of two subspaces V and W is not a subspace of $\mathbb{R}^n$. Let $V = \{(x, 0)|x \in \mathbb{R}\}$ and $W = \{(0, y)|y \in \mathbb{R}\}$. Then, V and W are lines through the origin and hence, are subspaces of $\mathbb{R}^n$. However, $V \cup W$ is a union of two lines, which is not a subspace of $\mathbb{R}^n$.

3.4.1 Trivial Subspaces

Let $\mathbf{0}$ be the zero vector of $\mathbb{R}^n$. The set $\{\mathbf{0}\} = \text{span}\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$ and is known as the zero space.

In relation to the standard basis, let $\mathbf{e}_1 = (1, 0, \ldots, 0), \mathbf{e}_2 = (0, 1, 0, \ldots, 0)$ and $\mathbf{e}_n = (0, 0, \ldots, 0, 1)$ be vectors in $\mathbb{R}^n$. Then, any vector $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$ can be written as

$$\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + \ldots + u_n\mathbf{e}_n.$$

Hence, $\mathbb{R}^n = \text{span}\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a subspace of $\mathbb{R}^n$.

3.4.2 Geometrical Interpretation of Subspaces

The following are all the subspaces of $\mathbb{R}^2$.
(i): the zero space $\{(0, 0)\}$
(ii): lines through the origin
(iii): $\mathbb{R}^2$

We provide a proof for the second statement.
Proof: First, we show that the set of lines through the origin is a non-empty subspace. Note that any line through the origin has the equation $y = mx$, where $m \in \mathbb{R}$. As $y = 0$ passes through the origin, then the subspace is non-empty.

Next, we prove that the subspace is closed under addition and scalar multiplication. Suppose $m_1, m_2, \alpha \in \mathbb{R}$. Note that $y = m_1x$ and $y = m_2x$ pass through the origin. Thus,

$$\begin{aligned} y &= (\alpha m_1 + m_2)x \\ &= \alpha m_1 x + m_2 x \end{aligned}$$

which is a line that passes through the origin. $\square$

REMARK

The above proof shows that we can prove that the subspace is closed under addition and closed multiplication simultaneously. Such a technique is useful, especially when we move on to the section on linear transformations later.

The following are all the subspaces of $\mathbb{R}^3$.
(i): the zero space $\{(0, 0, 0)\}$
(ii): lines through the origin
(iii): planes containing the origin
(iv): $\mathbb{R}^3$

The solution set of a homogeneous system of linear equations in n variables is a subspace of $\mathbb{R}^n$.

3.5 Linear Independence

Given a subspace $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$, how do we know whether there are redundant vectors among $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$? This encompasses the idea of linear independence, or rather linear dependence.

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \subseteq \mathbb{R}^n$, where $k \geq 2$. Consider the vector equation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k = \mathbf{0}$$

where $c_1, c_2, \ldots, c_k$ are variables. Note that $c_1 = c_2 = \ldots = c_k = 0$ satisfies the above equation and hence, is a solution to it. Recall that this solution is called the trivial solution.

Linear Independence

(i): S is a linearly independent set and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are linearly independent if

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k = \mathbf{0}$$

has only the trivial solution.

Alternatively, S is linearly independent if and only if no vector in S can be written as a linear combination of other vectors in S. Hence, there is no redundant vector in the set.

Linear Dependence

S is a linearly dependent set and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are linearly dependent if

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k = \mathbf{0}$$

has non-trivial solutions. This means that there exists $a_1, a_2, \ldots, a_k$, which are not all zero, such that

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \ldots + a_k\mathbf{u}_k = \mathbf{0}.$$

Alternatively, S is linearly dependent if and only if at least one vector $\mathbf{u}_i \in S$ can be written as a linear combination of other vectors in S, meaning that

$$\mathbf{u}_i = a_1\mathbf{u}_1 + \ldots + a_{i-1}\mathbf{u}_{i-1} + a_{i+1}\mathbf{u}_{i+1} + \ldots + a_k\mathbf{u}_k$$

for some $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k \in \mathbb{R}$. Hence, there exists at least one redundant vector in the set.

THEOREM

Let $S = \{\mathbf{u}\} \subseteq \mathbb{R}^n$. If S is linearly dependent, then there exists a real number $a \neq 0$ such that $a\mathbf{u} = \mathbf{0}$.

Proof: For any $a \neq 0$, $a\mathbf{u} = \mathbf{0} \iff \mathbf{u} = a^{-1}\mathbf{0} = \mathbf{0}$. Hence, S is linearly dependent if and only if $a\mathbf{u} = \mathbf{0}$. $\square$

Hence, in relation to what was covered in linear span, there are no redundant vectors among $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ if and only if $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are linearly independent.

Example: Are the vectors $(1, 0, 0, 1), (0, 2, 1, 0)$ and $(1, -1, 1, 1)$ linearly independent?

Solution: Consider the equation

$$c_1(1, 0, 0, 1) + c_2(0, 2, 1, 0) + c_3(1, -1, 1, 1) = (0, 0, 0, 0)$$

and by converting it into a matrix equation, we have

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Using EROs, we can reduce the coefficient matrix to

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & \frac{3}{2} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Hence, we have $\frac{3}{2}c_3 = 0 \implies c_3 = 0$ and so by backward substitution, we have $c_1 = c_2 = 0$ as well. Thus, the only solution is the trivial solution and so, the vectors are linearly independent. $\square$

THEOREM

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \subseteq \mathbb{R}^n$. If $k > n$, S is linearly dependent.

We can prove it by forming a system of n equations with k unknowns, implying that it has non-trivial solutions. In particular, in $\mathbb{R}^2$, a set of three or more vectors must be linearly dependent and in $\mathbb{R}^3$, a set of four or more vectors must be linearly dependent.

THEOREM

Let $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^3$ such that $V = \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$ and $W = \operatorname{span}\{\mathbf{u}, \mathbf{w}\}$ are planes in $\mathbb{R}^3$. If $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ are linearly independent, then $V \cap W = \operatorname{span}\{\mathbf{u}\}$. If $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ are not linearly independent, then $V \cap W = V = W$.

Proof: If $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ are linearly independent, then the two planes V and W intersect at the line spanned by $\mathbf{u}$. Hence the first result follows. $\square$

For the second result, V and W are planes in $\mathbb{R}^3$. Hence, $\mathbf{u}$ and $\mathbf{v}$ are linearly independent and $\mathbf{u}$ and $\mathbf{w}$ are linearly independent. Since all three vectors are linearly dependent, then they must lie on the same plane. $\square$

THEOREM

Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ be vectors in $\mathbb{R}^n$ and $\mathbf{P}$ a square matrix of order n. If $\mathbf{P}\mathbf{u}_1, \mathbf{P}\mathbf{u}_2, \ldots, \mathbf{P}\mathbf{u}_k$ are linearly independent, then $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are linearly independent.

Proof: Suppose

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k = \mathbf{0}.$$

Pre-multiplying both sides of the equation by $\mathbf{P}$,

$$c_1\mathbf{P}\mathbf{u}_1 + c_2\mathbf{P}\mathbf{u}_2 + \ldots + c_k\mathbf{P}\mathbf{u}_k = \mathbf{0}.$$

Hence, $c_1 = c_2 = \ldots = c_k = 0$ and the result follows. $\square$

3.5.1 Geometric Interpretation of Linear Independence

In $\mathbb{R}^2$ or $\mathbb{R}^3$, two vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent if and only if they lie on the same line but linearly independent if neither vector is a multiple of the other.

Consider the following figure where $\operatorname{span}(\mathbf{u}, \mathbf{v})$ is the x_1x_2-plane. On the left diagram, we see that if $\mathbf{w}$ can be written as a linear combination of $\mathbf{u}$ and $\mathbf{v}$, then the set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent. On the other hand, the right diagram presents us a scenario where $\mathbf{w}$ cannot be written as a linear combination of $\mathbf{u}$ and $\mathbf{v}$ because $\mathbf{w}$ is vertically above the x_1x_2 plane. Thus, for the right diagram, the set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent.

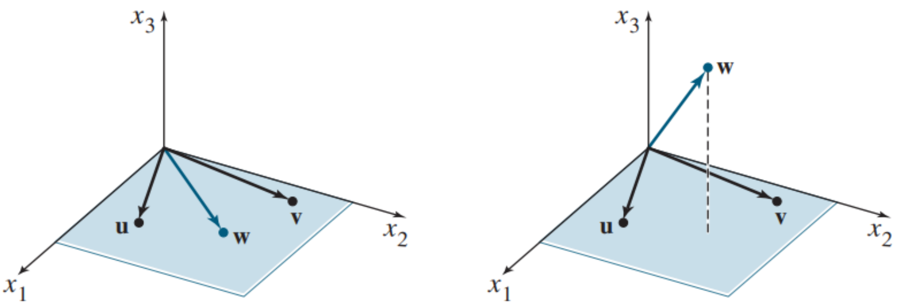

Figure 18: Geometric Interpretation of Linear Independence

3.6 Basis and Dimension

3.6.1 Basis

Basis

Let V be a vector space and $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a subset of V. Then, for S to be a basis (plural: bases) for V, it must satisfy the following two conditions:
(1): S is linearly independent
(2): S spans V

A basis for V can be used to build a coordinate system for V. Moreover, the basis is the set of the smallest size which can span V. For convenience, the empty set $\varnothing$ is defined to be the basis for the zero space. Except the zero space, any vector space has infinitely many different bases. We will now revisit the section on the standard basis for $\mathbb{R}^n$, but of course, placing more importance on the idea of basis.

Let $E = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$, where $\mathbf{e}_1 = (1, 0, \ldots, 0), \mathbf{e}_2 = (0, 1, 0, \ldots, 0)$ and $\mathbf{e}_n = (0, 0, \ldots, 0, 1)$ are vectors in $\mathbb{R}^n$. For any $\mathbf{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$, $\mathbf{u}$ can be written as a linear combination of the $\mathbf{e}_i$'s. That is,

$$\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + \ldots + u_n\mathbf{e}_n.$$

Thus, $\mathbb{R}^n = \operatorname{span}(E)$ and hence, E spans $\mathbb{R}^n$. Next, we set

$$c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \ldots + c_n\mathbf{e}_n = \mathbf{0}.$$

It is clear that $c_1 = c_2 = \ldots = c_n = 0$, implying that the vector equation has only the trivial solution and thus, E is linearly independent. This is why E is a basis for $\mathbb{R}^n$ and in particular, known as the standard basis.

3.6.2 Basis Theorem

Basis Theorem

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a basis for V. Then any set in V containing more than k vectors must be linearly dependent. Moreover, if V has a basis of k vectors, then every basis must also have k vectors.

3.6.3 Uniqueness of Basis Vectors

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a basis for a vector space V. Then, every vector $\mathbf{v} \in V$ can be expressed in the form

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

in exactly one way, where $c_1, c_2, \ldots, c_k \in \mathbb{R}$. However, why is the expression unique? We will provide a proof.

Proof: Suppose $\mathbf{v}$ can be expressed in two ways. That is,

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k \text{ and } \mathbf{v} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \ldots + d_k\mathbf{u}_k,$$

where $c_1, c_2, \ldots, c_k, d_1, d_2, \ldots, d_k \in \mathbb{R}$. Then

$$\begin{aligned} \mathbf{v} - \mathbf{v} &= c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k - d_1\mathbf{u}_1 - d_2\mathbf{u}_2 - \ldots - d_k\mathbf{u}_k \\ \mathbf{0} &= (c_1 - d_1)\mathbf{u}_1 + (c_2 - d_2)\mathbf{u}_2 + \ldots + (c_k - d_k)\mathbf{u}_k \end{aligned}$$

By definition, all the u_i's are linearly independent for $1 \leq i \leq k$, and hence $c_1 - d_1 = 0, c_2 - d_2 = 0, \ldots, c_k - d_k = 0 \implies c_i = d_i$ for all $1 \leq i \leq k$. Hence, the expression is unique. $\square$

3.6.4 Coordinate Systems

As mentioned, any vector $\mathbf{v} \in V$ can be expressed uniquely as such:

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k.$$

The coefficients $c_1, c_2, \ldots, c_k$ are called the coordinates of $\mathbf{v}$ relative to the basis S.

The vector

$$(\mathbf{v})_S = (c_1, c_2, \ldots, c_k) \in \mathbb{R}^k$$

is called the coordinate vector of $\mathbf{v}$ relative to S.

Let S be a basis for a vector space V.
(i): For any $\mathbf{u}, \mathbf{v} \in V$, $\mathbf{u} = \mathbf{v}$ if and only if $(\mathbf{u})_S = (\mathbf{v})_S$
(ii): For any $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ and $c_1, c_2, \ldots, c_r \in \mathbb{R}$,

$$\left(\sum_{i=1}^{r} c_i \mathbf{v}_i\right)_S = \sum_{i=1}^{r} c_i \left(\mathbf{v}_i\right)_S$$

THEOREM

If $|S| = k$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$, then
(1): $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ are linearly dependent vectors in V if and only if $(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, \ldots, (\mathbf{v}_r)_S$ are linearly dependent vectors in $\mathbb{R}^k$ and equivalently, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ are linearly independent vectors in V if and only if $(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, \ldots, (\mathbf{v}_r)_S$ are linearly independent vectors in $\mathbb{R}^k$

(2): $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} = V$ if and only if $\operatorname{span}\{(\mathbf{v}_1)_S, (\mathbf{v}_2)_S, \ldots, (\mathbf{v}_r)_S\} = \mathbb{R}^k$

3.6.5 Dimension

Let V be a vector space which has a basis with k vectors. Then, any subset of V with more than k vectors is always linearly dependent and any subset of V with less than k vectors cannot span V. This implies that every basis of V has the same size k.

Dimension

The dimension of a vector space, V, is denoted by $\dim(V)$ and it is the number of vectors in a basis for V. The dimension of the zero space is defined to be 0.

Note that $\dim(\mathbb{R}^n) = n$. Except $\{0\}$ and $\mathbb{R}^2$, subspaces of $\mathbb{R}^2$ are lines through the origin which are of dimension 1. Similarly, except $\{0\}$ and $\mathbb{R}^3$, subspaces of $\mathbb{R}^3$ are either lines through the origin, which are of dimension 1, or planes containing the origin, which are of dimension 2.

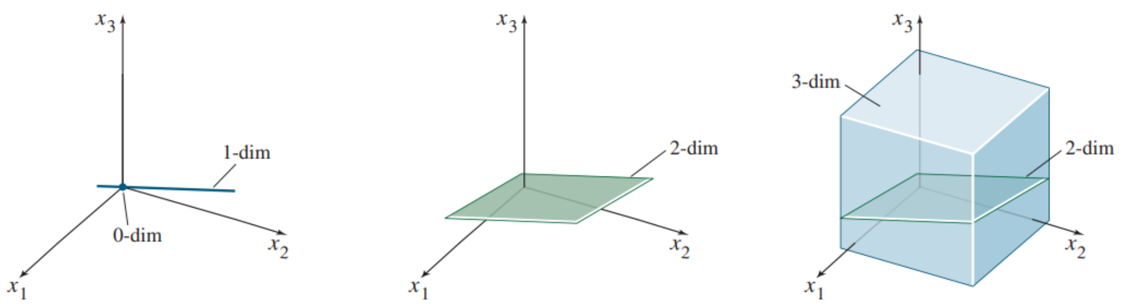

Figure 19: Subspaces of $\mathbb{R}^3$

Example: Consider the subspace $W = \{(x, y, z) | y = z\}$ of $\mathbb{R}^3$. Note that every vector of W is of the form

$$(x, y, y) = x(1, 0, 0) + y(0, 1, 1).$$

It is clear that $W = \operatorname{span}\{(1, 0, 0), (0, 1, 1)\}$ and the vectors $(1, 0, 0)$ and $(0, 1, 1)$ are linearly independent. Thus, a basis for W is $\{(1, 0, 0), (0, 1, 1)\}$. Since there are 2 vectors in the basis of W, then we conclude that $\dim(W) = 2$.

THEOREM

For a vector space V of dimension k and S being a subset of V, we have the following equivalent statements:
(1): S is a basis for V
(2): S is linearly independent and $|S| = k$
(3): S spans V and $|S| = k$

To prove this equivalence, we can use the method of contradiction.

REMARK

Any linearly independent set of exactly k elements in V is automatically a basis for V. Similarly, any set of exactly k elements that spans V is automatically a basis for V. Hence, if we want to check if S is a basis for V, we only need to check for two out of the following three conditions:
(i): S is linearly independent
(ii): S spans V
(iii): $|S| = k$

Now, we state an inequality related to dimension.

THEOREM

Let U be a subspace of a vector space V. Then, $\dim(U) \leq \dim(V)$. Furthermore, if $U \neq V$, then the inequality is strict. That is, $\dim(U) < \dim(V)$.

Proof: Let S be a basis for U. Since $U \subseteq V$, then S is a linearly independent subset of V. Hence,

$$\dim(U) = |S| \leq \dim(V).$$

Now, assume that $\dim(U) = \dim(V)$. As S is linearly independent and $|S| = \dim(V)$, then S is a basis for V. However, $U = \operatorname{span}(S) = V$. Hence, if $U \neq V$, it implies that $\dim(U) < \dim(V)$. □

THEOREM

Analogous to the Principle of Inclusion and Exclusion, there is a similar result for the dimension of subspaces. Let V and W be subspaces of $\mathbb{R}^n$. Then,

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

3.6.6 Invertible Matrix Theorem

The Invertible Matrix Theorem brings together whatever we have learnt from the past few topics. However, it also encompasses properties of row and column spaces, linear transformations and eigenvalues, which are out of the scope of our discussion as of now. At the end of this set of notes on MA2001, there will be a total of 13 equivalent statements which will wrap up our discussion on the Invertible Matrix Theorem.

Invertible Matrix Theorem

Let $\mathbf{A}$ be an $n \times n$ matrix. Then, the following statements are equivalent:
(1): $\mathbf{A}$ is invertible
(2): The linear system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution
(3): The RREF of $\mathbf{A}$ is an identity matrix
(4): $\mathbf{A}$ can be expressed as a product of elementary matrices
(5): $\det(\mathbf{A}) \neq 0$
(6): The rows of $\mathbf{A}$ form a basis for $\mathbb{R}^n$
(7): The columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$

3.7 Polynomial Vector Spaces

For $n \geq 0$, the set $\mathbb{P}_n$ of polynomials of degree at most n consists of all polynomials of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n,$$

where $a_0, a_1, \ldots, a_n$, the coefficients of $x^0, x^1, \ldots, x^n$ respectively, and x, are real numbers. If $p(x) = a_0 \neq 0$, then p is said to have a degree of 0. On the other hand, if all the coefficients are zero (that is $a_1 = a_2 = \ldots = a_n = 0$), then we obtain $p(x) = 0$, and this is called the zero polynomial.

The set of polynomials of degree at most n is a subspace of the set of all real-valued functions. To prove it, it suffices by checking the three axioms aforementioned. In Mathematics, an axiom is a statement or principle that is generally accepted to be true. For example the commutative law for addition $a+b = b+a$ is an axiom.

Proof: The zero polynomial is in $\mathbb{P}_n$ as mentioned and hence $\mathbb{P}_n$ is non-empty.

Next, we consider the polynomial $(\alpha p + q)(x)$ and prove that it must also be a polynomial of degree at most n. Upon proving this, we can conclude that $\mathbb{P}_n$ is closed under addition and scalar multiplication.

Suppose the polynomials $p(x)$ and $q(x)$ are defined as such:

$$\begin{aligned} p(x) &= a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \\ q(x) &= b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n \end{aligned}$$

Then,

$$\begin{aligned} (\alpha p + q)(x) &= \alpha(a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n) + b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n \\ &= \alpha a_0 + b_0 + \alpha a_1 x + b_1 x + \alpha a_2 x^2 + b_2 x^2 + \ldots + \alpha a_n x^n + b_n x^n \\ &= \alpha a_0 + b_0 + x(\alpha a_1 + b_1) + x^2(\alpha a_2 + b_2) + \ldots + x^n(\alpha a_n + b_n) \end{aligned}$$

which concludes the proof. $\square$

We regard derivatives and integrals as *linear operators*, which have the same properties of closure under addition and scalar multiplication as compared to polynomials. These will be discussed under the section of linear transformations.

3.8 Matrices and Vector Spaces

Of course, since matrices and vectors are related, the former would also satisfy the three axioms of a subspace.

The set of all 2×2 matrices is denoted by $\mathcal{M}_{2\times 2}$ and each matrix can be written as $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. It can be verified that $\mathcal{M}_{2\times 2}$ forms a vector space.

However, matrices of the form $\begin{pmatrix} 0 & -3 \\ a & b \end{pmatrix}$ will not form a subspace of $\mathcal{M}_{2\times 2}$ because the zero matrix cannot be obtained regardless of the values of a and b!

3.9 Transition Matrices and Change of Basis

For $\mathbf{v} \in V$, recall that if

$$\mathbf{v} = c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \ldots + c_k \mathbf{u}_k,$$

then the row vector $(\mathbf{v})_S = (c_1, c_2, \ldots, c_k)$ is the coordinate vector of $\mathbf{v}$ relative to S.

We define the column vector

$$[\mathbf{v}]_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

to be the coordinate vector of $\mathbf{v}$ relative to S.

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ are two bases for a vector space V. Take any vector $\mathbf{w} \in V$. Since S is a basis for V, then

$$\mathbf{w} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k$$

for some $c_1, c_2, \ldots, c_k \in \mathbb{R}$. We can express the coordinate vector of $\mathbf{w}$ relative to S (which is $[\mathbf{w}]_S$) as

$$[\mathbf{w}]_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

Since T is a basis for V, we can form the following system of equations:

$$\begin{aligned} \mathbf{u}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \ldots + a_{k1}\mathbf{v}_k \\ \mathbf{u}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \ldots + a_{k2}\mathbf{v}_k \\ &\vdots \\ \mathbf{u}_k &= a_{1k}\mathbf{v}_1 + a_{2k}\mathbf{v}_2 + \ldots + a_{kk}\mathbf{v}_k \end{aligned}$$

for some $a_{11}, a_{12}, \ldots, a_{kk} \in \mathbb{R}$. It can be verified that

$$[\mathbf{w}]_T = \begin{pmatrix} c_1a_{11} + c_2a_{12} + \ldots + c_ka_{1k} \\ c_1a_{21} + c_2a_{22} + \ldots + c_ka_{2k} \\ \vdots \\ c_1a_{k1} + c_2a_{k2} + \ldots + c_ka_{kk} \end{pmatrix}.$$

Note that $[\mathbf{w}]_S$ and $[\mathbf{w}]_T$ are related in the following manner:

$$\begin{aligned} [\mathbf{w}]_T &= \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \ldots & a_{kk} \end{pmatrix} [\mathbf{w}]_S \\ &= \begin{pmatrix} [\mathbf{u}_1]_T & [\mathbf{u}_2]_T & \ldots & [\mathbf{u}_k]_T \end{pmatrix} [\mathbf{w}]_S \end{aligned}$$

We let $\mathbf{P}$ be the matrix $\begin{pmatrix} [\mathbf{u}_1]_T & [\mathbf{u}_2]_T & \ldots & [\mathbf{u}_k]_T \end{pmatrix}$. Then $\mathbf{P}$ is the transition matrix from S to T.

Transition matrices are also known as stochastic matrices. These play a profound role in the study of Markov Chains, which is studied under the section of eigenvalues and eigenvectors. This is where each entry in the transition matrix denotes a probability.

Example: Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be bases for $\mathbb{R}^3$, where $\mathbf{u}_1 = (1, 0, -1), \mathbf{u}_2 = (0, -1, 0), \mathbf{u}_3 = (1, 0, 2), \mathbf{v}_1 = (1, 1, 1), \mathbf{v}_2 = (1, 1, 0)$ and $\mathbf{v}_3 = (-1, 0, 0)$. We wish to obtain the transition matrix from S to T.

Solution: First, we need to find $a_{11}, a_{12}, \ldots, a_{33}$ such that

$$\begin{aligned} \mathbf{u}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + a_{31}\mathbf{v}_3 \\ \mathbf{u}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + a_{32}\mathbf{v}_3 \\ \mathbf{u}_3 &= a_{13}\mathbf{v}_1 + a_{23}\mathbf{v}_2 + a_{33}\mathbf{v}_3 \end{aligned}$$

To solve the first row, we consider the equation

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = a_{11} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + a_{21} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + a_{31} \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix},$$

which yields $a_{11} = -1, a_{21} = 1$ and $a_{31} = -1$. We can repeat this method to solve for the other unknowns but this is slightly tedious. Recall that we can convert the original system of equations into an augmented matrix of the following manner, which would be much more efficient:

$$\left(\begin{array}{ccc|c|c|c} 1 & 1 & -1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 2 \end{array}\right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c|c|c} 1 & 0 & 0 & -1 & 0 & 2 \\ 0 & 1 & 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & -1 & -1 & -1 \end{array}\right)$$

Note that the first three columns denote the $\mathbf{v}_i$'s, whereas the last three columns denote the $\mathbf{u}_i$'s. Solving yields

$$\begin{aligned} \mathbf{u}_1 &= -\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3 \\ \mathbf{u}_2 &= -\mathbf{v}_2 + \mathbf{v}_3 \\ \mathbf{u}_3 &= 2\mathbf{v}_1 - 2\mathbf{v}_2 - \mathbf{v}_3 \end{aligned}$$

Thus, the transition matrix from S to T is

$$\begin{pmatrix} -1 & 0 & 2 \\ 1 & -1 & -2 \\ -1 & -1 & -1 \end{pmatrix}.$$

□

This raises a question. Is the transition matrix from T to S simply the inverse of the transition matrix from S to T ?

Solution: Forming the augmented matrix will yield

$$\begin{aligned} \mathbf{v}_1 &= \frac{1}{3}\mathbf{u}_1 + \frac{2}{3}\mathbf{u}_2 - \frac{2}{3}\mathbf{u}_3 \\ \mathbf{v}_2 &= -\mathbf{u}_1 - \mathbf{u}_2 \\ \mathbf{v}_3 &= \frac{2}{3}\mathbf{u}_1 + \frac{1}{3}\mathbf{u}_2 - \frac{1}{3}\mathbf{u}_3 \end{aligned}$$

and hence, the transition matrix from T to S is

$$\begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -1 & -1 & 0 \\ \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{pmatrix}.$$

Note that

$$\begin{pmatrix} -1 & 0 & 2 \\ 1 & -1 & -2 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -1 & -1 & 0 \\ \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{pmatrix} = \mathbf{I},$$

so the claim is true! We will prove it later. □

THEOREM

Let S and T be two bases for a vector space V and let $\mathbf{P}$ be the transition matrix from S to T. We have the following two results:

(1): $\mathbf{P}$ is invertible

(2): $\mathbf{P}^{-1}$ is the transition matrix from T to S

As mentioned, we will provide a proof for the second result.

Proof: Let $\mathbf{Q}$ be the transition matrix from T to S. Our goal is to show that $\mathbf{QP} = \mathbf{I}$, and hence it will follow that $\mathbf{P}$ is invertible and $\mathbf{P}^{-1} = \mathbf{Q}$.

Suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$. Recall that the coordinate of $\mathbf{u}_1$ relative to S is denoted by $[\mathbf{u}_1]_S$ and can be represented by the matrix

$$[\mathbf{u}_1]_S = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

In general, the coordinate of $\mathbf{u}_i$ relative to S, where $1 \le i \le k$, is a column vector with the entry 1 in the i^{th} row and 0 for the rest. In relation to the standard basis vectors, it is clear that $[\mathbf{u}_i]_S = \mathbf{e}_i$.

Hence, the i^{th} column of $\mathbf{QP}$ can be represented by $\mathbf{QPe}_i$. Thus,

$$\mathbf{QPe}_i = \mathbf{QP}[\mathbf{u}_i]_S = \mathbf{Q}[\mathbf{u}_i]_T = [\mathbf{u}_i]_S = \mathbf{e}_i$$

and the result follows. $\square$

4 Vector Spaces Associated with Matrices

4.1 Row Spaces and Column Spaces

Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix. That is,

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

The row space of $\mathbf{A}$ is the subspace of $\mathbb{R}^n$ spanned by the rows of $\mathbf{A}$ and conversely, the column space of $\mathbf{A}$ is the subspace of $\mathbb{R}^m$ spanned by the columns of $\mathbf{A}$.

Let $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m$ be the m rows of $\mathbf{A}$. That is,

$$\begin{aligned} \mathbf{r}_1 &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{pmatrix} \\ \mathbf{r}_2 &= \begin{pmatrix} a_{21} & a_{22} & \dots & a_{2n} \end{pmatrix} \\ &\vdots \\ \mathbf{r}_m &= \begin{pmatrix} a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \end{aligned}$$

Let $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ be the n columns of $\mathbf{A}$. That is,

$$\mathbf{c}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{c}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \mathbf{c}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Hence, we have the following results:

$$\begin{aligned} \text{row space of } \mathbf{A} &= \operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_m\} \subseteq \mathbb{R}^n \\ \text{column space of } \mathbf{A} &= \operatorname{span}\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\} \subseteq \mathbb{R}^m \end{aligned}$$

By considering transpose, the row space of $\mathbf{A}$ is the same as the column space of $\mathbf{A}^{\mathrm{T}}$ while the column space of $\mathbf{A}$ is the same as the row space of $\mathbf{A}^{\mathrm{T}}$.

Example: Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 0 \\ 1 & -1 & 3 \\ -5 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

where the rows of $\mathbf{A}$ are $(2,-1,0), (1,-1,3), (-5,1,0)$ and $(1,0,1)$ and the columns of $\mathbf{A}$ are $\begin{pmatrix} 2 \\ 1 \\ -5 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 1 \\ -5 \\ 1 \end{pmatrix}$. Let us find the row space of $\mathbf{A}$ and column space of $\mathbf{A}$, and the bases for and dimensions of them.

Solution: The row space of $\mathbf{A}$ is the span of the row vectors. That is,

$$\begin{aligned} &\{a(2,-1,0) + b(1,-1,3) + c(-5,1,0) + d(1,0,1) | a,b,c,d \in \mathbb{R}\} \\ &= \{(2a + b - 5c + d, -a - b + c, 3b + d) | a,b,c,d \in \mathbb{R}\} \end{aligned}$$

which is a subspace of $\mathbb{R}^3$. As such, any basis of the row space of $\mathbf{A}$ contains at most three vectors. We can verify that the row vectors $(2,-1,0),(1,-1,3)$ and $(-5,1,0)$ are linearly independent. Thus, a basis for the row space of $\mathbf{A}$ is $\{(2,-1,0),(1,-1,3),(-5,1,0)\}$ and the dimension of the row space is 3.

The column space of $\mathbf{A}$ is the span of the three column vectors. It is

$$\left\{a\begin{pmatrix}2\\1\\-5\\1\end{pmatrix}+b\begin{pmatrix}-1\\-1\\1\\0\end{pmatrix}+c\begin{pmatrix}0\\3\\0\\1\end{pmatrix}\middle|\, a,b,c\in\mathbb{R}\right\}=\left\{\begin{pmatrix}2a-b\\a-b+3c\\-5a+b\\a+c\end{pmatrix}\middle|\, a,b,c\in\mathbb{R}\right\}$$

which is a subspace of $\mathbb{R}^4$. Similarly, as all three columns of $\mathbf{A}$ are linearly independent, then they form a basis for the column space of $\mathbf{A}$. Note that the dimension of the column space of $\mathbf{A}$ is 3. □

In general, the dimension of the row space is equal to the dimension of the column space. We have an alternative term for it, which is denoted by rank.

In the example above, we can write the columns of $\mathbf{A}$ as $\mathbf{c}_1=(2,1,-5,1)^{\mathrm{T}},\mathbf{c}_2=(-1,-1,1,0)^{\mathrm{T}}$ and $\mathbf{c}_3=(0,3,0,1)^{\mathrm{T}}$. Hence, we can write the column space of $\mathbf{A}$ as

$$\left\{(2a-b,a-b+3c,-5a+b,a+c)^{\mathrm{T}}|a,b,c\in\mathbb{R}\right\}.$$

THEOREM

Recall from the first two topics that two matrices $\mathbf{A}$ and $\mathbf{B}$ are row equivalent if one can be obtained from the other by a series of EROs. Thus, if they are row equivalent, then their row spaces are identical as EROs preserve the row space (but not the column space) of a matrix.

THEOREM

Any matrix is row equivalent to itself. Moreover, if a matrix $\mathbf{C}$ is row equivalent to a matrix $\mathbf{B}$ and $\mathbf{B}$ is row equivalent to another matrix $\mathbf{A}$, then $\mathbf{C}$ is also row equivalent to $\mathbf{A}$.

Any matrix is row equivalent to its REF. In particular, if two matrices have the same REF, then they are row equivalent. Since every matrix has a unique RREF, then two matrices are row equivalent if and only if they have the same RREF.

THEOREM

Let $\mathbf{A}$ be a matrix and $\mathbf{R}$ an REF of $\mathbf{A}$. Then, the set of non-zero rows of $\mathbf{R}$ is a basis for the row space of $\mathbf{A}$ (since they are linearly independent). Since the column space of $\mathbf{A}$ is the row space of $\mathbf{A}^{\mathrm{T}}$, a basis for the column space of $\mathbf{A}$ can be obtained from an REF of $\mathbf{A}^{\mathrm{T}}$.

In general, the column space of $\mathbf{A}$ is not equal to the column space of $\mathbf{B}$.
Example: $\mathbf{A}$ and $\mathbf{B}$, where

$$\mathbf{A}=\begin{pmatrix}0&0\\1&0\end{pmatrix}\text{ and }\mathbf{B}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$$

are row equivalent as we can swap the two rows. However, the column space of $\mathbf{A}$ is not equal to the column space of $\mathbf{B}$.

THEOREM

Let $\mathbf{A}$ be a matrix and $\mathbf{R}$ an REF of $\mathbf{A}$. A basis for the column space of $\mathbf{A}$ can be obtained by taking the columns of $\mathbf{A}$ that correspond to the pivot columns in $\mathbf{R}$.

Example: Consider a matrix $\mathbf{A}$ and its REF $\mathbf{R}$, where

$$\mathbf{A} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -1 & -1 & 2 & -3 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 & -1 \end{pmatrix} \xrightarrow{\text{REF}} \mathbf{R} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ 0 & 0 & \frac{3}{2} & -3 & \frac{3}{2} \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Observe the leading entries of $\mathbf{R}$. The three pivot columns containing these leading entries form a basis for the column space of $\mathbf{R}$, which in turn implies that the three corresponding columns form a basis for the column space of $\mathbf{A}$. That is, a basis for the column space of $\mathbf{A}$ is

$$\left\{ \begin{pmatrix} 2 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

4.1.1 Extending a Linearly Independent Subset to a Basis

Example: Consider the subset $S = \{(1,4,-2,5,1),(2,9,-1,8,2),(2,9,-1,9,3)\}$. We wish to extend S to a basis for $\mathbb{R}^5$. First, we state an algorithm that extends a linearly independent subset S of $\mathbb{R}^n$ to a basis for $\mathbb{R}^n$.

Basis Extension

Step 1: Form a matrix $\mathbf{A}$ using the vectors in S as rows
Step 2: Reduce $\mathbf{A}$ to its REF $\mathbf{R}$
Step 3: Identify the non-pivot columns in $\mathbf{R}$
Step 4: For each non-pivot column, get a vector so as the leading entry of the vector is at that column
Step 5: The union of S and the set of vectors obtained in Step 4 is a basis for $\mathbb{R}^n$

Solution: We write the vectors in S as rows. Hence,

$$\mathbf{A} = \begin{pmatrix} 1 & 4 & -2 & 5 & 1 \\ 2 & 9 & -1 & 8 & 2 \\ 2 & 9 & -1 & 9 & 3 \end{pmatrix}.$$

The REF of $\mathbf{A}$, $\mathbf{R}$ can be written as

$$\mathbf{R} = \begin{pmatrix} 1 & 4 & -2 & 5 & 1 \\ 0 & 1 & 3 & -2 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

The non-pivot columns are the third and fifth columns. In relation to Step 4, the vectors required are $(0,0,1,0,0)$ and $(0,0,0,0,1)$. Hence, taking the union of the two sets yields the conclusion that

$$\{(1,4,-2,5,1),(2,9,-1,8,2),(2,9,-1,9,3),(0,0,1,0,0),(0,0,0,0,1)\}$$

is a basis for $\mathbb{R}^5$. $\square$

THEOREM

Let $\mathbf{A}$ be an $m \times n$ matrix. Then, the

$$\text{column space of } \mathbf{A} = \{\mathbf{Au} | \mathbf{u} \in \mathbb{R}^n\}.$$

That is, a system of linear equations $\mathbf{Ax} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ lies in the column space of $\mathbf{A}$.

Proof: Write $\mathbf{A} = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n \end{pmatrix}$, where $\mathbf{c}_j$ is the j^{th} column of $\mathbf{A}$. For any $\mathbf{u} = (u_1, u_2, \ldots, u_n)^{\mathrm{T}} \in \mathbb{R}^n$,

$$\begin{aligned} \mathbf{Au} &= \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \ldots \\ u_n \end{pmatrix} \\ &= u_1\mathbf{c}_1 + u_2\mathbf{c}_2 + \ldots + u_n\mathbf{c}_n \\ &\in \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\} \\ &= \text{column space of } \mathbf{A} \end{aligned}$$

Next, suppose $\mathbf{b}$ is in the column space of $\mathbf{A}$. That is, $\mathbf{b} \in \text{span}\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$. Hence, there exists $u_1, u_2, \ldots, u_n \in \mathbb{R}$ such that

$$\mathbf{b} = u_1\mathbf{c}_1 + u_2\mathbf{c}_2 + \ldots + u_n\mathbf{c}_n = \mathbf{Au},$$

where $\mathbf{u} = \begin{pmatrix} u_1, u_2, \ldots, u_n \end{pmatrix}^{\mathrm{T}}$. Thus, the result follows. □

4.2 Ranks

THEOREM

The rank of a matrix, $\mathbf{A}$, is the dimension of its row space, or column space. We refer to it as rank($\mathbf{A}$).

rank($\mathbf{A}$) is the number of non-zero rows and the number of pivot columns in an REF of $\mathbf{A}$. The zero matrix, 0 has a rank of 0 and an identity matrix of order n, $\mathbf{I}_n$, has a rank of n.

We shall prove that rank($\mathbf{A}$) is the dimension of the row space of $\mathbf{A}$.
Proof: Let $\mathbf{A}$ be a matrix and $\mathbf{R}$ an REF of $\mathbf{A}$. Since the row space of $\mathbf{A}$ coincides with that of $\mathbf{R}$, then the dimension of the row space of $\mathbf{A}$ is the number of non-zero rows in $\mathbf{R}$, which is also equal to the number of pivot columns in $\mathbf{R}$. On the other hand, the columns of $\mathbf{A}$ that correspond to the pivot columns in $\mathbf{R}$ form a basis for the column space of $\mathbf{A}$. Hence, the dimension of the column space of $\mathbf{A}$ is equal to the number of pivot columns in $\mathbf{R}$. □

4.2.1 Full Rank

For an $m \times n$ matrix $\mathbf{A}$, its rank satisfies the inequality

$$\text{rank}(\mathbf{A}) \leq \min\{m, n\}.$$

(1): If $\text{rank}(\mathbf{A}) = \min\{m, n\}$, $\mathbf{A}$ has full rank
(2): A square matrix $\mathbf{A}$ is of full rank if and only if $\det(\mathbf{A}) \neq 0$
(3): $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^{\mathrm{T}})$ for any matrix $\mathbf{A}$ since the row space of $\mathbf{A}$ is the column space of $\mathbf{A}^{\mathrm{T}}$

Some other properties related to the rank of a matrix are as follows. First, let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ and $n \times p$ matrices respectively.
(4): Rank inequality

$$\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$$

Proof: Let $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \end{pmatrix}$, where $\mathbf{a}_i$ and $\mathbf{b}_i$ are the i^{th} columns of $\mathbf{A}$ and $\mathbf{B}$ respectively. Then,

$$\mathbf{AB} = \begin{pmatrix} \mathbf{Ab}_1 & \mathbf{Ab}_2 & \ldots & \mathbf{Ab}_p \end{pmatrix},$$

where $\mathbf{Ab}_i$ is the i^{th} column of $\mathbf{AB}$. Hence,

$$\mathbf{Ab}_i \in \text{column space of } \mathbf{A} = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}.$$

Also, the

$$\begin{aligned}\text{column space of } \mathbf{AB} &= \text{span}\{\mathbf{Ab}_1, \mathbf{Ab}_2, \ldots, \mathbf{Ab}_p\} \\ &\subseteq \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_\text{n} \\ &= \text{column space of } \mathbf{A}\end{aligned}$$

Thus, the column space of $\mathbf{AB}$ is less than or equal to the column space of $\mathbf{A}$, implying that

$$\text{rank}(\mathbf{AB}) \le \text{rank}(\mathbf{A}).$$

By considering that $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{AB})^\text{T}$, then

$$\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{AB})^\text{T} = \text{rank}(\mathbf{B}^\text{T}\mathbf{A}^\text{T}) \le \text{rank}(\mathbf{B}^\text{T}) = \text{rank}(\mathbf{B}).$$

The result follows. □

To conclude, property **(4)** implies that

$$\begin{aligned}\text{rank}(\mathbf{AB}) &\le \text{rank}(\mathbf{A}) \\ \text{rank}(\mathbf{AB}) &\le \text{rank}(\mathbf{B})\end{aligned}$$

(5): Sylvester's Rank Inequality

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n \le \text{rank}(\mathbf{AB})$$

4.3 Nullspaces and Nullities

Nullspace and Nullity

Let $\mathbf{A}$ be an $m \times n$ matrix. The solution space of the homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ is the nullspace of $\mathbf{A}$ and the dimension of its nullspace is called the nullity of $\mathbf{A}$. The nullity is denoted by $\text{nullity}(\mathbf{A})$.

Since the nullspace is a subspace of $\mathbb{R}^n$, then

$$\text{nullity}(\mathbf{A}) \le n.$$

Example: Consider the matrix $\mathbf{A}$, where

$$\mathbf{A} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -1 & -1 & 2 & -3 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 & -1 \end{pmatrix}.$$

We wish to find a basis for the nullspace of $\mathbf{A}$ and compute its nullity.

Solution: The RREF of the augmented matrix $(\mathbf{A}|\mathbf{0})$ is

$$\left(\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

To obtain a basis for the nullspace of $\mathbf{A}$, it is equivalent to solving the following system of equations:

$$\begin{aligned}x_1 + x_2 + x_5 &= 0 \\ x_3 + x_5 &= 0 \\ x_4 &= 0\end{aligned}$$

For the first two equations, observe that x_5 is the common variable. Let it be some arbitrary parameter s. Then,

$$x_1 + x_2 + s = 0$$
$$x_3 = -s$$

Now, we have found x_3. For the first equation, we have two unknowns, x_1 and x_2. We can set either one to be the other parameter (say t). Without a loss of generality, set $x_2 = t$. Hence, $x_1 = -s - t$. To conclude,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = s \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} + t \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

where $s, t \in \mathbb{R}$. Thus, a basis for the nullspace of $\mathbf{A}$ is

$$\left\{ \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right\}.$$

Since there are two vectors in the nullspace of $\mathbf{A}$, then $\text{nullity}(\mathbf{A}) = 2$. □

We make an observation, observe that the RREF of $\mathbf{A}$ has three pivot columns, implying that $\text{rank}(\mathbf{A}) = 3$. As mentioned, $\text{nullity}(\mathbf{A}) = 2$. Note that the number of columns of $\mathbf{A}$ is 5, which is, of course, the sum of 3 and 2. Is that a coincidence? It turns out that there is a theorem related to this, which is known as the Rank-Nullity Theorem. We will mention this in a while.

4.3.1 Invertible Matrix Theorem

Now, we will state four more properties of the Invertible Matrix Theorem, which are just a continuation of the previous seven mentioned.

Invertible Matrix Theorem

For an $n \times n$ matrix $\mathbf{A}$,
(8): the column space of $\mathbf{A} = \mathbb{R}^n$
(9): $\text{rank}(\mathbf{A}) = n$
(10): $\text{nullity}(\mathbf{A}) = 0$
(11): The nullspace of $\mathbf{A}$ is the zero vector. That is, $\{\mathbf{0}\}$.

4.3.2 Rank-Nullity Theorem

Rank-Nullity Theorem

For a matrix $\mathbf{A}$ with n columns,
$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n.$$

We provide two proofs of the rank-nullity theorem.

Proof: The first proof is simple. Consider a matrix $\mathbf{A}$. Then, the rank of its RREF is the same as $\text{rank}(\mathbf{A})$ and the kernel of the RREF of $\mathbf{A}$ is equal to the nullspace of $\mathbf{A}$.

This should be clear because the rank is invariant under EROs and the Gauss-Jordan form of $\mathbf{A}$ is obtained through row operations. Next, to prove the statement regarding the kernel, suppose there exists $\mathbf{x} \in \text{null}(\mathbf{A})$, which implies that $\mathbf{Ax} = \mathbf{0}$. Suppose that $\mathbf{B} = (\mathbf{E}_1\mathbf{E}_2 \ldots \mathbf{E}_k)\mathbf{A}$, where $\mathbf{B}$ is the Gauss-Jordan form of $\mathbf{A}$ and $\mathbf{E}_1, \mathbf{E}_2, \ldots \mathbf{E}_k$ are elementary matrices. This implies that $\mathbf{E}_1\mathbf{E}_2 \ldots \mathbf{E}_k$ is some invertible matrix. Post multiplying both sides of the equation by $\mathbf{x}$, we have

$$\mathbf{Bx} = (\mathbf{E}_1\mathbf{E}_2 \ldots \mathbf{E}_k)\mathbf{Ax} = \mathbf{0},$$

so $\mathbf{x} \in \text{null}(\mathbf{B})$. This means that $\text{null}(\mathbf{A}) \subseteq \text{null}(\mathbf{B})$.

Now, assume that $\mathbf{x} \in \text{null}(\mathbf{B})$. Then, since $\mathbf{Bx} = \mathbf{0}$,

$$\mathbf{Ax} = (\mathbf{E}_1\mathbf{E}_2 \ldots \mathbf{E}_k)^{-1}\mathbf{Bx} = \mathbf{0},$$

implying that $\text{null}(\mathbf{B}) \subseteq \text{null}(\mathbf{A})$.

Combining $\text{null}(\mathbf{A}) \subseteq \text{null}(\mathbf{B})$ and $\text{null}(\mathbf{B}) \subseteq \text{null}(\mathbf{A})$, we assert that $\text{null}(\mathbf{A}) = \text{null}(\mathbf{B})$. $\square$

For the second proof, we consider the cases where $\text{rank}(\mathbf{A}) = n$ and $\text{rank}(\mathbf{A}) = r < n$.

Proof: Suppose $\text{rank}(\mathbf{A}) = n$. Then, the only solution to the equation $\mathbf{Ax} = \mathbf{0}$ is the trivial solution. Hence, $\text{null}(\mathbf{A}) = \mathbf{0}$, so $\text{nullity}(\mathbf{A}) = 0$ and the result follows.

Now, if $\text{rank}(\mathbf{A}) = r$, where $r < n$, there are $n - r$ free variables in the solution $\mathbf{Ax} = \mathbf{0}$. Let $t_1, t_2, \ldots, t_{n-r}$ denote these free variables and let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-r}$ denote the solutions obtained by sequentially setting each free variable to 1 and the remaining free variables to zero. As the set $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-r}$ is linearly independent and every solution to $\mathbf{Ax} = \mathbf{0}$ can be expressed as a linear combination of all the x_i's for $1 \le i \le n - r$, then

$$\mathbf{x} = \sum_{i=1}^{n-r} t_i x_i.$$

Hence, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-r}$ spans the nullspace of $\mathbf{A}$. Since all the x_i's are linearly independent, then they form a basis for the nullspace, so $(\mathbf{A}) = n - r$. $\square$

4.3.3 Ordinary Differential Equations

This section is for the interested and to see how certain concepts in Linear Algebra are related to Calculus. Consider the ordinary differential equation

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x),$$

where $a, b, c \in \mathbb{R}$. To some who are taking up Engineering courses, the study of second order differential equations (which is the above mentioned) is crucial. Note that the solution to the differential equation, y, comprises the homogeneous solution for the case where

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0$$

and a particular solution for

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x).$$

Example: Solve the differential equation

$$\frac{d^2y}{dx^2} + \frac{dy}{dx} - 12y = 2\sin x.$$

Solution: First, we find the homogeneous solution. That is, we set

$$\frac{d^2y}{dx^2} + \frac{dy}{dx} - 12y = 0.$$

Its auxiliary equation is $m^2 + m - 12 = 0$, and so $m = 3$ or $m = -4$. The homogenous solution is

$$y = Ae^{3x} + Be^{-4x},$$

where A and B are constants. Note that this solution will satisfy the differential equation when the right side of the equation is zero. Since the n^{th} derivatives of sine and cosine functions are also sine and cosine functions for $n \in \mathbb{N}$, then the particular solution is of the form

$$y = C \sin x + D \cos x.$$

Unlike the homogeneous case, there is only one particular solution which will satisfy the differential equation

$$\frac{d^2y}{dx^2} + \frac{dy}{dx} - 12y = 2\sin x.$$

Substituting $y = C \sin 2x + D \cos 2x$ yields

$$\begin{aligned} -C \sin x - D \cos x + C \cos x - D \sin x - 12C \sin x - 12D \cos x &= 2 \sin x \\ (-13C - D) \sin x + (C - 13D) \cos x &= 2 \sin x \\ -13C - D = 2 \text{ and } C - 13D &= 0 \end{aligned}$$

We obtain the solutions $C = -\frac{13}{85}$ and $D = -\frac{1}{85}$, which implies that the particular solution is

$$y = -\frac{13}{85} \sin x - \frac{1}{85} \cos x.$$

Combining the homogeneous solution and the particular solution yields the general solution to the original differential equation, which is

$$y = Ae^{3x} + Be^{-4x} - \frac{13}{85} \sin x - \frac{1}{85} \cos x.$$

□

5 Orthogonality

5.1 Dot Product and L^p Norm

In general, let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ be a vector in $\mathbb{R}^n$. Then, the length of $\mathbf{u}$ is

$$\|u\| = \sqrt{u_1^2 + u_2^2 + \ldots + u_n^2}.$$

We call this the Euclidean L^2 norm or 2-norm. For the interested, especially those who have prior Mathematical Olympiad experience, the Cauchy-Bunyakovsky-Schwarz (or simply Cauchy-Schwarz) Inequality, Hölder's Inequality and Minkowski's Inequality are no alien to you. The last two inequalities are for general p in L^p space, so we have to deal with the concept of L^p norm.

The L^p norm of a vector $\mathbf{u}$ is

$$\|u\|_p = \left(|x_1|^p + |x_2|^p + \ldots + |x_n|^p\right)^{\frac{1}{p}}.$$

The absolute value is a norm on the one-dimensional vector spaces formed by the real or complex numbers.

Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ and θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Then, the distance between $\mathbf{u}$ and $\mathbf{v}$ is denoted by

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

By the cosine rule,

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

Since $-1 \le \cos\theta \le 1$, then

$$\mathbf{u} \bullet \mathbf{v} \le \|\mathbf{u}\|\|\mathbf{v}\|.$$

The inequality we just established is called the Cauchy-Schwarz Inequality. The inequality has different forms, such as in sums and integrals but always yields the same results. This case deals with the Cauchy-Schwarz Inequality for dot products.

The commutative, distributive and associative property of the dot product will not be covered as it was once taught in H2 Mathematics. However, we state an interesting fundamental result and its proof to start off the section on orthogonality. The result is that

$$\mathbf{u} \bullet \mathbf{u} \ge 0; \text{ and } \mathbf{u} \bullet \mathbf{u} = 0 \text{ if and only if } \mathbf{u} = \mathbf{0}.$$

Proof: We shall prove the first statement that $\mathbf{u} \bullet \mathbf{u} \ge 0$. Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$. Then,

$$\mathbf{u} \bullet \mathbf{u} = u_1^2 + u_2^2 + \ldots + u_n^2.$$

Regardless of whether each u_i is positive, zero, or negative for $1 \le i \le n$, its square, u_i^2, is always non-negative. Hence, the first result follows. □

Proof: For the second result, it is clear that when $\mathbf{u} = \mathbf{0}$, then $\mathbf{u} \bullet \mathbf{u} = 0$. Hence, it suffices to prove the other direction. Similarly, let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$. Then, taking the dot product of it with itself again, if $\mathbf{u} \bullet \mathbf{u} = 0$, it implies that all the u_i's must be zero for all $1 \le i \le n$. The result follows. □

5.1.1 Dot Product as Transpose

Suppose $\mathbf{u}$ and $\mathbf{v}$ are written as column vectors. That is,

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \text{ and } \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

Observe that

$$\mathbf{u} \bullet \mathbf{v} = \mathbf{u}^{\mathrm{T}}\mathbf{v}.$$

5.1.2 Some Classical Inequalities

Let $\mathbf{u}$ and $\mathbf{v}$ be any two vectors in $\mathbb{R}^n$. Then, we establish two classical inequalities which are the Triangle Inequality and the Cauchy-Schwarz Inequality.

Triangle Inequality

The Triangle Inequality states that

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Proof:

$$\begin{aligned}(\mathbf{u}+\mathbf{v}) \bullet (\mathbf{u}+\mathbf{v}) &= \mathbf{u} \bullet \mathbf{u} + 2(\mathbf{u} \bullet \mathbf{v}) + \mathbf{v} \bullet \mathbf{v} \\ \|\mathbf{u}+\mathbf{v}\|^2 &= \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta + \|\mathbf{v}\|^2 \\ \|\mathbf{u}+\mathbf{v}\|^2 &\leq (\|\mathbf{u}\| + \|\mathbf{v}\|)^2\end{aligned}$$

Taking square roots on both sides and the result follows. □

Cauchy-Schwarz Inequality

The Cauchy-Schwarz Inequality states that

$$|\mathbf{u} \bullet \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

Proof: Note that $\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$ and $|\cos\theta| \leq 1$. □

5.2 Orthogonal and Orthonormal Bases

Two vectors, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if their dot product is zero. That is, $\mathbf{u} \bullet \mathbf{v} = 0$. In other words, the concept of orthogonality in $\mathbb{R}^n$ is the same as perpendicularity in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Define a set S of vectors in $\mathbb{R}^n$. S is called an orthogonal set if every pair of distinct vectors in S are orthogonal. Moreover, if every vector in S is a unit vector (is of norm 1), then S is called an orthonormal set. The process of multiplying each vector $\mathbf{u}$ by $\frac{1}{\|u\|}$ is called normalising.

Let V be a vector space. A basis S of V is an orthogonal basis if S is orthogonal and similarly, a basis S of V is an orthonormal basis if S is orthonormal.

Example: Let $\mathbf{u}_1 = (2,0,0), \mathbf{u}_2 = (0,1,1)$ and $\mathbf{u}_3 = (0,1,-1)$. We wish to prove that $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ is an orthogonal set and construct an orthonormal set based on $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$.

Solution: Observe that

$$\begin{aligned}\mathbf{u}_1 \bullet \mathbf{u}_2 &= 0 \\ \mathbf{u}_1 \bullet \mathbf{u}_3 &= 0 \\ \mathbf{u}_2 \bullet \mathbf{u}_3 &= 0\end{aligned}$$

or simply, $\mathbf{u}_i \bullet \mathbf{u}_j = 0$ for $i \neq j$. Hence, $\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\}$ is an orthogonal set.

To find an orthonormal set, we have to ensure that each vector in S is a unit vector. Setting $\mathbf{v}_1 = (1,0,0), \mathbf{v}_2 = \left(0,\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right)$ and $\mathbf{v}_3 = \left(0,\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right)$, we observe that $\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ is an orthonormal set. □

The standard basis $E = \{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}$ of $\mathbb{R}^n$ is an orthonormal set because $\|\mathbf{e}_i\| = 1$ for all $1 \leq i \leq n$ and $\mathbf{e}_i \bullet \mathbf{e}_j = 0$ for all $i \neq j$. Moreover, it is both an orthogonal basis and orthonormal basis.

THEOREM

Let S be an orthogonal set of non-zero vectors in a vector space V. Then, S is linearly independent.

Proof: Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$. Consider the vector equation

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k = \mathbf{0}.$$

We wish to prove that $c_1 = c_2 = \ldots = c_k = 0$. Since S is orthogonal, then $\mathbf{u}_i \bullet \mathbf{u}_j = 0$ for all $i \neq j$. Then, for $1 \leq i \leq k$,

$$\begin{aligned}(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k) \bullet \mathbf{u}_i &= (c_1\mathbf{u}_1) \bullet \mathbf{u}_i + (c_2\mathbf{u}_2) \bullet \mathbf{u}_i + \ldots + (c_k\mathbf{u}_k) \bullet \mathbf{u}_i \\ &= (c_1\mathbf{u}_1) \bullet \mathbf{u}_i + (c_2\mathbf{u}_2) \bullet \mathbf{u}_i + \ldots + (c_i\mathbf{u}_i) \bullet \mathbf{u}_i + \ldots + (c_k\mathbf{u}_k) \bullet \mathbf{u}_i \\ &= (c_i\mathbf{u}_i) \bullet \mathbf{u}_i\end{aligned}$$

Hence, replacing $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k$ with $\mathbf{0}$, then

$$\begin{aligned}\mathbf{0} \bullet \mathbf{u}_i &= c_i(\mathbf{u}_i \bullet \mathbf{u}_i) \\ \mathbf{0} &= c_i\|\mathbf{u}_i\|^2\end{aligned}$$

Since $\mathbf{u}_i \neq 0$, then the norm of $\mathbf{u}_i \neq 0$ too. This implies that $c_i = 0$ for all $1 \leq i \leq n$. □

REMARK

To determine whether a set S of non-zero vectors in a vector space V of dimension k is an orthogonal basis, we need to check if S is orthogonal and $|S| = k$.

THEOREM

(1): If $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthogonal basis for a vector space V, then for any vector $\mathbf{w}$ in V,

$$\begin{aligned}\mathbf{w} &= \frac{\mathbf{w} \bullet \mathbf{u}_1}{\mathbf{u}_1 \bullet \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{w} \bullet \mathbf{u}_2}{\mathbf{u}_2 \bullet \mathbf{u}_2}\mathbf{u}_2 + \ldots + \frac{\mathbf{w} \bullet \mathbf{u}_k}{\mathbf{u}_k \bullet \mathbf{u}_k}\mathbf{u}_k \\ &= \sum_{i=1}^{k} \frac{\mathbf{w} \bullet \mathbf{u}_i}{\|\mathbf{u}_i\|}\mathbf{u}_i\end{aligned}$$

This also means that

$$(\mathbf{w})_S = \left(\frac{\mathbf{w} \bullet \mathbf{u}_1}{\mathbf{u}_1 \bullet \mathbf{u}_1}, \frac{\mathbf{w} \bullet \mathbf{u}_2}{\mathbf{u}_2 \bullet \mathbf{u}_2}, \ldots, \frac{\mathbf{w} \bullet \mathbf{u}_k}{\mathbf{u}_k \bullet \mathbf{u}_k}\right).$$

(2): If $T = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthonormal basis for a vector space V, then for any vector $\mathbf{w}$ in V,

$$\mathbf{w} = (\mathbf{w} \bullet \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{w} \bullet \mathbf{v}_2)\mathbf{v}_2 + \ldots + (\mathbf{w} \bullet \mathbf{v}_k)\mathbf{v}_k.$$

This also means that

$$(\mathbf{w})_T = (\mathbf{w} \bullet \mathbf{v}_1, \mathbf{w} \bullet \mathbf{v}_2, \ldots, \mathbf{w} \bullet \mathbf{v}_k).$$

Proof: We will only provide a proof for the first theorem because the second theorem follows from the definition of an orthonormal basis. To prove **(1)**, let

$$(\mathbf{w})_S = (c_1, c_2, \ldots, c_k),$$

meaning that we can write $\mathbf{w}$ as a linear combination of the $\mathbf{u}_i$'s for $1 \leq i \leq k$. Hence,

$$\mathbf{w} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_k\mathbf{u}_k.$$

Then, $\mathbf{w} \bullet \mathbf{u}_i = c_i(\mathbf{u}_i \bullet \mathbf{u}_i)$. Rearranging the equation yields the result. □

5.2.1 Orthogonality

Let V be a subspace of $\mathbb{R}^n$. A vector $\mathbf{u} \in \mathbb{R}^n$ is orthogonal to V if $\mathbf{u}$ is orthogonal to all vectors in V. If V is a plane in $\mathbb{R}^3$, then the normal vector $\mathbf{n}$ is orthogonal to V.

In gneeral, if $V = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is a subspace of $\mathbb{R}^n$, then a vector $\mathbf{v} \in \mathbb{R}^n$ is orthogonal to V if and only if $\mathbf{v} \bullet \mathbf{u}_i = 0$ for $1 \le i \le k$.

Example: Let $V = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2\}$ be a subspace of $\mathbb{R}^4$, where $\mathbf{u}_1 = (1,1,1,0)$ and $\mathbf{u}_2 = (0,-1,-1,1)$. We wish to find all vectors that are orthogonal to V.

Solution: Let $\mathbf{v} = (w,x,y,z)$ be a vector in $\mathbb{R}^4$. Then,

$$\begin{aligned}\mathbf{v} \bullet (a\mathbf{u}_1 + b\mathbf{u}_2) = 0 \text{ for all } a,b \in \mathbb{R}\\ \mathbf{v} \bullet \mathbf{u}_1 = 0 \text{ and } \mathbf{v} \bullet \mathbf{u}_2 = 0\end{aligned}$$

From here, we can form the following system of equations:

$$\begin{aligned} w + x + y &= 0 \\ -x - y + z &= 0 \end{aligned}$$

Hence, $(w,x,y,z) = (-t,-s+t,s,t)$ for some $s,t \in \mathbb{R}$. This implies that $\mathbf{v}$ is orthogonal to V if and only if

$$\mathbf{v} = (-t,-s+t,s,t) = s(0,-1,1,0) + t(-1,1,0,1)$$

for some $s,t \in \mathbb{R}$. □

5.2.2 Orthogonal Projection

Let V be a subspace of $\mathbb{R}^n$. Every vector $\mathbf{u} \in \mathbb{R}^n$ can be written uniquely as

$$\mathbf{u} = \mathbf{n} + \mathbf{p}$$

such that $\mathbf{n}$ is a vector orthogonal to V and $\mathbf{p}$ is a vector in V. The vector $\mathbf{p}$ is called the orthogonal projection of $\mathbf{u}$ onto V. In certain textbooks, $\mathbf{p}$ is usually written as

$$\mathbf{p} = \operatorname{proj}_V \mathbf{u}.$$

The following figure is taken from a textbook, for which some of the notations are different from what we use in this set of notes. It presents readers a geometric interpretation of orthogonal projection with reference to a subspace of $\mathbb{R}^n$. In this case, $\mathbf{z}$ and $\hat{\mathbf{y}}$ represent the normal vector and the projection vector respectively, and $\mathbf{y} = \mathbf{z} + \hat{\mathbf{y}}$ is an arbitrary vector in $\mathbb{R}^n$.

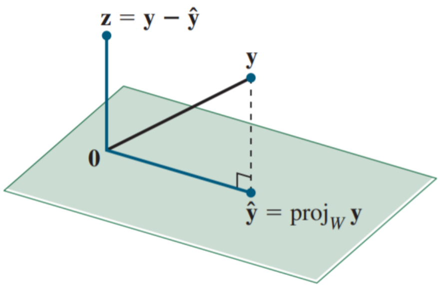

Figure 20: Geometric interpretation of orthogonal projection

THEOREM

Let V be a subspace of $\mathbb{R}^n$ and $\mathbf{w}$ a vector in $\mathbb{R}^n$.
(1): If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthogonal basis for V, then

$$\mathbf{w} = \frac{\mathbf{w} \bullet \mathbf{u}_1}{\mathbf{u}_1 \bullet \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{w} \bullet \mathbf{u}_2}{\mathbf{u}_2 \bullet \mathbf{u}_2}\mathbf{u}_2 + \ldots + \frac{\mathbf{w} \bullet \mathbf{u}_k}{\mathbf{u}_k \bullet \mathbf{u}_k}\mathbf{u}_k$$

is the projection of $\mathbf{w}$ onto V.

(2): If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthonormal basis for V, then

$$(\mathbf{w} \bullet \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{w} \bullet \mathbf{v}_2)\mathbf{v}_2 + \ldots + (\mathbf{w} \bullet \mathbf{v}_k)\mathbf{v}_k$$

is the projection of $\mathbf{w}$ onto V.

Proof: We only prove the first part as the second part is a consequence.

Given that

$$\mathbf{p} = \frac{\mathbf{w} \bullet \mathbf{u}_1}{\mathbf{u}_1 \bullet \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{w} \bullet \mathbf{u}_2}{\mathbf{u}_2 \bullet \mathbf{u}_2}\mathbf{u}_2 + \ldots + \frac{\mathbf{w} \bullet \mathbf{u}_k}{\mathbf{u}_k \bullet \mathbf{u}_k}\mathbf{u}_k \text{ and } \mathbf{n} = \mathbf{w} - \mathbf{p},$$

then for $1 \le i \le k$,

$$\begin{aligned}
\mathbf{n} \bullet \mathbf{u}_i &= \mathbf{w} \bullet \mathbf{u}_i - \mathbf{p} \bullet \mathbf{u}_i \\
&= \mathbf{w} \bullet \mathbf{u}_i - \frac{\mathbf{w} \bullet \mathbf{u}_1}{\mathbf{u}_1 \bullet \mathbf{u}_1}\mathbf{u}_1 - \frac{\mathbf{w} \bullet \mathbf{u}_2}{\mathbf{u}_2 \bullet \mathbf{u}_2}\mathbf{u}_2 - \ldots - \frac{\mathbf{w} \bullet \mathbf{u}_k}{\mathbf{u}_k \bullet \mathbf{u}_k}\mathbf{u}_k \\
&= \mathbf{w} \bullet \mathbf{u}_i - \frac{\mathbf{w} \bullet \mathbf{u}_i}{\mathbf{u}_i \bullet \mathbf{u}_i}(\mathbf{u}_i \bullet \mathbf{u}_i) \\
&= 0
\end{aligned}$$

□

5.2.3 Gram-Schmidt Process

Gram-Schmidt Process

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ be a basis for a vector space V. We can construct an orthogonal basis for V, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, where

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1 \\
\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\mathbf{u}_2 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{u}_3 - \frac{\mathbf{u}_3 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{u}_3 \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_k &= \mathbf{u}_k - \frac{\mathbf{u}_k \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{u}_k \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 - \ldots - \frac{\mathbf{u}_k \bullet \mathbf{v}_{k-1}}{\mathbf{v}_{k-1} \bullet \mathbf{v}_{k-1}}\mathbf{v}_{k-1}
\end{aligned}$$

In general,

$$\mathbf{v}_k = \mathbf{u}_k - \sum_{i=1}^{k-1} \frac{\mathbf{u}_k \bullet \mathbf{v}_i}{\|\mathbf{v}_i\|^2}\mathbf{v}_i.$$

It is thus clear that

$$\left\{\frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2, \ldots, \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k\right\}$$

is an orthonormal basis for V.

5.3 Best Approximations

Let V be a subspace of $\mathbb{R}^n$. Take any $\mathbf{u} \in \mathbb{R}^n$ and let $\mathbf{p}$ be the projection of $\mathbf{u}$ onto V. Then,

$$d(\mathbf{u}, \mathbf{p}) \leq d(\mathbf{u}, \mathbf{v})$$

for all $\mathbf{v} \in V$, implying that $\mathbf{p}$ is the best approximation of $\mathbf{u}$ in V.

Proof: Let $\mathbf{p}$ and $\mathbf{w}$ be the projections of $\mathbf{u}$ and $\mathbf{x}$ onto V respectively. For any vector $\mathbf{v}$ in V,

$$\mathbf{n} = \mathbf{u} - \mathbf{p}$$
$$\mathbf{w} = \mathbf{p} - \mathbf{v}$$
$$\mathbf{x} = \mathbf{u} - \mathbf{v}$$

where $\mathbf{n}$ is the normal to V. Note that $\mathbf{x} = \mathbf{n} + \mathbf{w}$ and the vectors $\mathbf{n}$ and $\mathbf{w}$ are orthogonal. Then,

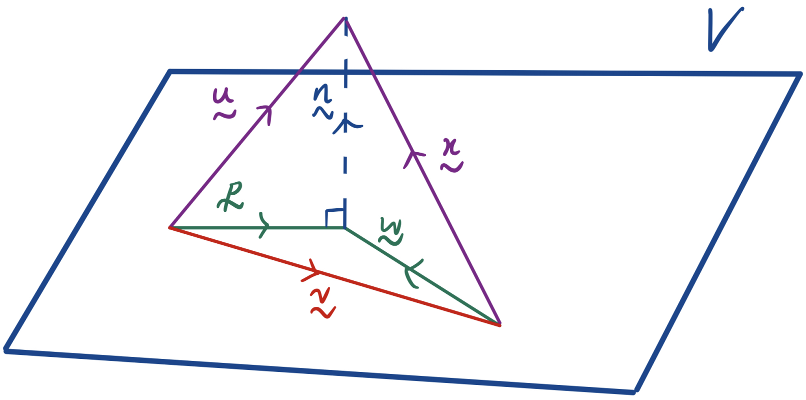

Figure 21: Geometrical Interpretation of the Best Approximation

$$\begin{aligned}
\|\mathbf{x}\|^2 &= \mathbf{x} \bullet \mathbf{x} \\
&= (\mathbf{n} + \mathbf{w}) \bullet (\mathbf{n} + \mathbf{w}) \\
&= \mathbf{n} \bullet \mathbf{n} + 2(\mathbf{n} \bullet \mathbf{w}) + \mathbf{w} \bullet \mathbf{w} \\
&= \|\mathbf{n}\|^2 + \|\mathbf{w}\|^2 \\
&\geq \|\mathbf{n}\|^2
\end{aligned}$$

Thus, $\|\mathbf{x}\| \geq \|\mathbf{n}\|$. Since $d(\mathbf{u}, \mathbf{p}) = \|\mathbf{u} - \mathbf{p}\| = \|\mathbf{n}\|$ and $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \|\mathbf{x}\|$, then we are done. □

5.3.1 Least Squares Problem

Solution to the Least Squares Problem

Let $\mathbf{Ax} = \mathbf{b}$ be a linear system, where $\mathbf{A}$ is an $m \times n$ matrix, and let $\mathbf{p}$ be the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$. Then,

$$\|\mathbf{b} - \mathbf{p}\| \leq \|\mathbf{b} - \mathbf{Av}\|$$

for all $\mathbf{v} \in \mathbb{R}^n$. This implies that $\mathbf{u}$ is a least squares solution to $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{Au} = \mathbf{p}$.

5.3.2 Method of Least Squares

THEOREM

Let $\mathbf{Ax} = \mathbf{b}$ be a linear system. Then, $\mathbf{u}$ is a least squares solution to $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{u}$ is a solution to $\mathbf{A}^{\mathrm{T}}\mathbf{Ax} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$.

Proof: Let $\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix}$, where $\mathbf{a}_i$ is the i^{th} column of $\mathbf{A}$, and let V be the column space of $\mathbf{A}$. Then, if $\mathbf{u}$ is a solution to $\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$, we have

$$\mathbf{A}^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\mathbf{u}) = \mathbf{0}.$$

Thus, $\mathbf{a}_i \bullet (\mathbf{b} - \mathbf{A}\mathbf{u}) = 0$ for all $1 \le i \le n$, and so $\mathbf{b} - \mathbf{A}\mathbf{u}$ is orthogonal to all the $\mathbf{a}_i$'s, implying that $\mathbf{b} - \mathbf{A}\mathbf{u}$ is orthogonal to V. Hence, $\mathbf{A}\mathbf{u}$ is the projection of $\mathbf{b}$ onto $\mathbf{V}$, and the result follows. The proof for the other directions is simply a reversal of the steps mentioned. □

5.3.3 Least Squares Lines

A common task in Science and Engineering is to analyse and understand relationships between several quantities which vary. This section describes a variety of situations in which data are used to build or verify a formula that predicts the value of one variable as a function of other variables. In each case, the problem will amount to solving a least squares problem.

Here, instead of the conventional matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, we write it as $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$, where $\mathbf{X}$, $\boldsymbol{\beta}$ and $\mathbf{y}$ are the design matrix, parameter vector and observation vector respectively. Experimental data often produce points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ such that when graphed, seem to lie close to a line, modelled by the equation $y = \beta_0 + \beta_1 x$. We call this line the line of regression, and recall from H2 Mathematics that the line of regression of y on x is

$$y - \overline{y} = b\left(x - \overline{x}\right), \text{ where } b = \frac{\sum\left(x - \overline{x}\right)\left(y - \overline{y}\right)}{\sum\left(x - \overline{x}\right)^2}.$$

Unlike the conventional equation of a line $y = mx + c$, the equation $y = \beta_0 + \beta_1 x$ is commonly used for least squares lines.

Suppose β_0 and β_1 are fixed and consider the line $y = \beta_0 + \beta_1 x$ as shown. Corresponding to each data point (x_j, y_j), there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinate. y_j is the observed value of y and $\beta_0 + \beta_1 x_j$ is the predicted value of y, which is determined by the regression line. The difference between an observed y-value and a predicted y-value is called a *residual*. Recall that for an observed y-value y_j, where

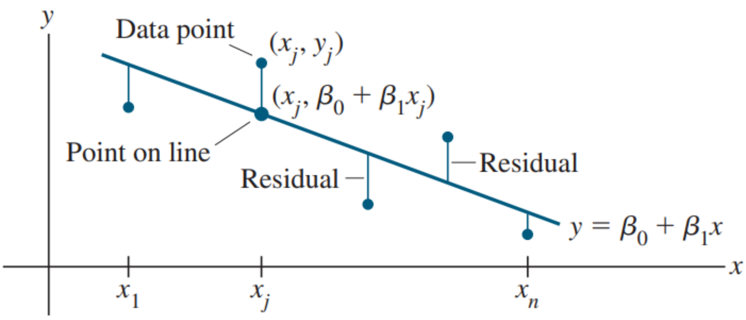

Figure 22: Fitting a line to experimental data

$1 \le j \le n$, its predicted y-value is $\beta_0 + \beta_1 x_j$. We can express this system of equations as a matrix equation, which is

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

which resembles $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$. The $\boldsymbol{\beta}$ that minimises this sum also minimises the distance between $\mathbf{X}\boldsymbol{\beta}$ and $\mathbf{y}$. Hence, finding the least-squares solution of $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$ is equivalent to finding the $\boldsymbol{\beta}$ that determines the equation of the least squares line.

5.4 Orthogonal Matrices

A square matrix $\mathbf{A}$ is orthogonal if $\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}}$. Hence, if we wish to prove if a square matrix $\mathbf{A}$ is orthogonal, it suffices to show that $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{I}$ or $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{I}$.

The following statements are equivalent:
(1): $\mathbf{A}$ is orthogonal
(2): The rows of $\mathbf{A}$ form an orthonormal basis for $\mathbb{R}^n$
(3): The columns of $\mathbf{A}$ form an orthonormal basis for $\mathbb{R}^n$

THEOREM

Let S and T be two orthonormal bases for a vector space and let $\mathbf{P}$ be the transition matrix from S to T. Then, $\mathbf{P}$ is orthogonal and $\mathbf{P}^{\mathrm{T}}$ is the transition matrix from T to S.

6 Diagonalisation

6.1 Eigenvalues and Eigenvectors

Let $\mathbf{A}$ be a square matrix of order n. A non-zero column vector $\mathbf{u} \in \mathbb{R}^n$ is called an eigenvector of $\mathbf{A}$ if

$$\mathbf{Au} = \lambda\mathbf{u}$$

for some scalar λ. The scalar λ is called an eigenvalue of $\mathbf{A}$ and $\mathbf{u}$ is an eigenvector of $\mathbf{A}$ associated with the eigenvalue λ. Now, we will state a technique used to find the eigenvalues and eigenvectors of a square matrix.

Let $\mathbf{A}$ be a square matrix of order n. Then,

$$\begin{aligned}
&\lambda \text{ is an eigenvalue of } \mathbf{A} \\
\iff\ & \mathbf{Au} = \lambda\mathbf{u} \text{ for some non-zero column vector } \mathbf{u} \in \mathbb{R}^n \\
\iff\ & (\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0} \text{ for some non-zero column vector } \mathbf{u} \in \mathbb{R}^n \\
\iff\ & (\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0} \text{ has non-trivial solutions} \\
\iff\ & \det(\mathbf{A} - \lambda\mathbf{I}) = 0
\end{aligned}$$

If expanded, $\det(\mathbf{A} - \lambda\mathbf{I})$ is a polynomial in λ of degree n. The equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ is called the characteristic equation of $\mathbf{A}$ and the polynomial $\det(\mathbf{A} - \lambda\mathbf{I})$ is called the characteristic polynomial of $\mathbf{A}$.

Example: For example, we wish to find the eigenvalues and eigenvectors of the following matrix:

$$\mathbf{A} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix}$$

A fun fact is that the above matrix is a Markov Matrix since its entries are between 0 and 1 inclusive and the sum of column entries is 1. Of course, such matrices play a pivotal role in the branch of Statistics.

Solution: The characteristic polynomial, $\det(\mathbf{A} - \lambda\mathbf{I})$ is

$$(\lambda - 1)(\lambda - 0.95)$$

after some algebraic manipulation. Setting it equal to 0 yields $\lambda = 1$ or $\lambda = 0.95$, which are the eigenvalues of $\mathbf{A}$. Next, we find the eigenvectors of $\mathbf{A}$. For $\lambda = 1$, substituting it into $(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}$, where $\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$ yields

$$\begin{gathered}
(\mathbf{A} - \mathbf{I})\mathbf{u} = \mathbf{0} \\
\begin{pmatrix} -0.04 & 0.01 \\ 0.04 & -0.01 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{0} \\
\begin{pmatrix} -0.04x + 0.01y \\ 0.04x - 0.01y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{gathered}$$

Note that the two rows on the left side of the equation differ by a scalar multiple of -1. Hence, $0.04x - 0.01y = 0 \implies 4x = y$. Hence, the corresponding eigenvector is $\begin{pmatrix} 1 \\ 4 \end{pmatrix}$. The eigenvector corresponding to the eigenvalue $\lambda = 0.95$ will be left as an exercise to the reader. □

Example: Next, we wish to find the eigenvalues of the following matrix:

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

Solution: The characteristic polynomial of $\mathbf{B}$ is

$$\lambda^2(\lambda - 3).$$

The eigenvalues are 0 and 3. □

REMARK

In the previous example, even though $\mathbf{B}$ is of order 3, we expect it to have 3 eigenvalues, and thus 3 eigenvectors corresponding to the respective eigenvalues. However, we only obtained two eigenvalues, namely 0 and 3. So, is there a problem?

There is actually no issue because the polynomial $\lambda^2(\lambda - 3)$ is of degree 3, so by the Fundamental Theorem of Algebra, we expect it to have exactly 3 complex roots, counting multiplicity. Of course, the imaginary parts of the eigenvalues are zero so the roots are real.

In general, given any square matrix $\mathbf{A}$, when finding its eigenvalues, we can use EROs to reduce $\mathbf{A} - \lambda\mathbf{I}$ to a triangular matrix to find $\det(\mathbf{A} - \lambda\mathbf{I})$ but cannot reduce $\mathbf{A}$ to a triangular matrix using EROs.

To those who have watched Avengers: Endgame, there is a scene where Tony Stark talked about the eigenvalue of a Möbius Strip. This actually makes sense but is out of our discussion due to its complexity!

THEOREM

In relation to determinant and trace, for a square matrix $\mathbf{A}$ of order n with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then

$$\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$$
$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$$

6.1.1 Invertible Matrix Theorem

Now, we will state one more property of the Invertible Matrix Theorem, which is just a continuation of the previous four mentioned.

Invertible Matrix Theorem

For an $n \times n$ matrix $\mathbf{A}$,
(12): 0 is not an eigenvalue of $\mathbf{A}$

6.1.2 Triangular Matrices

Suppose $\mathbf{A}$ is an $n \times n$ (upper) triangular matrix, where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nn} \end{pmatrix}.$$

Note that

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda - a_{11})(\lambda - a_{22})\ldots(\lambda - a_{nn}),$$

which implies that the eigenvalues of $\mathbf{A}$ are $a_{11}, a_{22}, \ldots, a_{nn}$. The same claim can be made for lower triangular matrices.

6.1.3 Eigenspaces

Let $\mathbf{A}$ be a square matrix of order n and λ an eigenvalue of $\mathbf{A}$. Then, the solution space of the linear system

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}$$

is called the eigenspace of $\mathbf{A}$ associated with the eigenvalue λ and is denoted by E_λ or $E_\lambda(\mathbf{A})$. If $\mathbf{u}$ is a non-zero vector in E_λ, then $\mathbf{u}$ is an eigenvector of $\mathbf{A}$ associated with λ.

Example: Earlier, we found the eigenvalues of the matrix

$$\mathbf{A} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix}$$

which are 1 and 0.95, as well as the eigenvector corresponding to the eigenvalue 1, which is $\begin{pmatrix} 1 \\ 4 \end{pmatrix}$. The eigenspace of $\mathbf{A}$ is easy to obtain.

Solution: Since

$$\begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} 1 \\ 4 \end{pmatrix}$$

for some $t \in \mathbb{R}$, then it is clear that $E_1 = \operatorname{span}\left\{(1,4)^{\mathrm{T}}\right\}$. □

6.1.4 Diagonalisation and Power of Matrices

A square matrix $\mathbf{A}$ is diagonalisable if there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is a diagonal matrix. That is,

$$\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} \text{ or } \mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

$\mathbf{P}$ is said to diagonalise $\mathbf{A}$. However, note not all matrices are diagonalisable.

THEOREM

Let $\mathbf{A}$ be a square matrix of order n. Then $\mathbf{A}$ is diagonalisable if and only if $\mathbf{A}$ has linearly independent eigenvectors.

Proof: Suppose $\mathbf{A}$ is diagonalisable. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix}$ be an invertible matrix such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is a diagonal matrix. That is,

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

$$\mathbf{A}\mathbf{P} = \mathbf{P}\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

$$\mathbf{A}\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \ldots & \mathbf{A}\mathbf{u}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \ldots & \lambda_n\mathbf{u}_n \end{pmatrix}$$

As such, $\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i$ for $1 \le i \le n$. Since $\mathbf{P}$ is invertible, the columns of $\mathbf{P}$, namely $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ form a basis for $\mathbb{R}^n$.

Next, suppose $\mathbf{A}$ has linearly independent eigenvectors. Then, $\mathbf{A}\mathbf{u}_i = \lambda_i\mathbf{u}_i$ for $1 \le i \le n$, where all the λ's are the eigenvalues of $\mathbf{A}$. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix}$, which is n $n \times n$ matrix. Since the columns of $\mathbf{P}$ are linearly independent and $\dim(\mathbb{R}^n) = n$, then the columns of $\mathbf{P}$ form a basis for $\mathbb{R}^n$, implying that $\mathbf{P}$ is invertible. □

Now, we will state a method to diagonalise a square matrix $\mathbf{A}$ of order n.

Diagonalisation Process

Step 1: Find all the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ by solving $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
Step 2: For each eigenvalue λ_i for $1 \le i \le k$, find a basis S_{λ_i} for the eigenspace E_{λ_i}
Step 3: Let

$$S = \bigcup_{i=1}^{k} S_{\lambda_i},$$

which is always linearly independent. If $|S| < n$, then $\mathbf{A}$ is not diagonalisable. On the other hand, if $|S| = n$, say $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, then $\mathbf{A}$ is diagonalisable and $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \end{pmatrix}$ is an invertible matrix that diagonalises $\mathbf{A}$.

THEOREM

Let $\mathbf{A}$ be a square matrix of order n. If $\mathbf{A}$ has n distinct eigenvalues, then $\mathbf{A}$ is diagonalisable. However, the converse of the theorem is not true. That is, a diagonalisable matrix of order n need not have n distinct eigenvalues.

At the start of this section, we mentioned that that if $\mathbf{P}$ diagonalises $\mathbf{A}$, then $\mathbf{A} = \mathbf{PDP}^{-1}$, where the diagonal entries of $\mathbf{D}$ are the eigenvalues of $\mathbf{A}$ and the columns of $\mathbf{P}$ are the corresponding eigenvectors of $\mathbf{A}$. Using this method, we can find the power of any diagonalisable square matrix. Suppose

$$\mathbf{A} = \mathbf{PDP}^{-1}.$$

Then, for $n \in \mathbb{Z}$,

$$\mathbf{A}^n = \mathbf{PD}^n\mathbf{P}^{-1}.$$

Proof: Using induction, we assume that the statement is true for some $k \in \mathbb{N}$. Once we prove that it is true for $k \in \mathbb{N}$, then it is true for all $n \in \mathbb{N}$, and we can repeat this process for the negative integers. Assuming that

$$\mathbf{A}^k = \mathbf{PD}^k\mathbf{P}^{-1}$$

is true, we wish to prove that

$$\mathbf{A}^{k+1} = \mathbf{PD}^{k+1}\mathbf{P}^{-1}$$

is true. This is true because

$$\begin{aligned}
\mathbf{A}^{k+1} &= \mathbf{A}^k\mathbf{A} \\
&= \mathbf{PD}^k\mathbf{P}^{-1}\mathbf{PDP}^{-1} \text{ using induction hypothesis} \\
&= \mathbf{PD}^k\mathbf{IDP}^{-1} \\
&= \mathbf{PD}^k\mathbf{DP}^{-1} \\
&= \mathbf{PD}^{k+1}\mathbf{P}^{-1}
\end{aligned}$$

which concludes the proof. $\square$

Example: Suppose we wish to diagonalise the matrix

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

and find $\mathbf{A}^n$ for arbitrary $n \in \mathbb{N}$.

Solution: Note that

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{pmatrix} -4-\lambda & 0 & -6 \\ 2 & 1-\lambda & 2 \\ 3 & 0 & 5-\lambda \end{pmatrix}.$$

Setting $\det(\mathbf{A}-\lambda\mathbf{I})=0$, we have

$$\begin{aligned}(-4-\lambda)(1-\lambda)(5-\lambda)-6[-3(1-\lambda)]&=0\\(-4-\lambda)(1-\lambda)(5-\lambda)+18(1-\lambda)&=0\\(1-\lambda)[(-4-\lambda)(5-\lambda)+18]&=0\\(1-\lambda)(\lambda+1)(\lambda-2)&=0\end{aligned}$$

Thus, the eigenvalues are -1, 1 and 2. For the eigenvalue -1,

$$\begin{aligned}(\mathbf{A}+\mathbf{I})\begin{pmatrix}x\\y\\z\end{pmatrix}&=\mathbf{0}\\\begin{pmatrix}-3&0&-6\\2&2&2\\3&0&6\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}&=\begin{pmatrix}0\\0\\0\end{pmatrix}\end{aligned}$$

We only need to solve the following two equations:

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

Substituting $x=-2z$ into the second equation yields $y=z$, and hence

$$-\frac{x}{2}=y=z,$$

implying that the corresponding eigenvector is $\begin{pmatrix}-2\\1\\1\end{pmatrix}$. The eigenvectors corresponding to $\lambda=1$ and $\lambda=2$ are $\begin{pmatrix}0\\1\\0\end{pmatrix}$ and $\begin{pmatrix}-1\\0\\1\end{pmatrix}$ respectively. Hence, $\mathbf{D}=\begin{pmatrix}-1&0&0\\0&1&0\\0&0&2\end{pmatrix}$ and $\mathbf{P}=\begin{pmatrix}-2&0&-1\\1&1&0\\1&0&1\end{pmatrix}$, implying that

$$\mathbf{A}=\begin{pmatrix}-2&0&-1\\1&1&0\\1&0&1\end{pmatrix}\begin{pmatrix}-1&0&0\\0&1&0\\0&0&2\end{pmatrix}\begin{pmatrix}-2&0&-1\\1&1&0\\1&0&1\end{pmatrix}^{-1}.$$

Thus, $\mathbf{A}^n$ can be computed using simple matrix multiplication as shown, but the rest is left as an exercise:

$$\mathbf{A}^n=\begin{pmatrix}-2&0&-1\\1&1&0\\1&0&1\end{pmatrix}\begin{pmatrix}(-1)^n&0&0\\0&1&0\\0&0&2^n\end{pmatrix}\begin{pmatrix}-2&0&-1\\1&1&0\\1&0&1\end{pmatrix}^{-1}.$$

□

6.1.5 Fibonacci Sequence

Check out the video titled 'The applications of eigenvectors and eigenvalues' by Zach Star for an analysis of this section.

I believe that we are all familiar with the sequence:

$$0,1,1,2,3,5,8,13,21,34,55,89,\ldots$$

It is called the Fibonacci Sequence and it can be modelled by a second order linear recurrence relation with constant coefficients. That is,

$$F_n=F_{n-1}+F_{n-2},$$

where $F_0 = 0$ and $F_1 = 1$. By repeatedly applying this recursion, we can obtain the n^{th} term of the sequence. In terms of matrix multiplication, the recurrence relation can be expressed, also, as

$$\begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_{n-1} \\ F_{n-2} \end{pmatrix}.$$

As such, we have

$$\begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^2 \begin{pmatrix} F_{n-2} \\ F_{n-3} \end{pmatrix}$$

and hopefully, one can spot the pattern that

$$\begin{aligned} \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-1} \begin{pmatrix} F_1 \\ F_0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{aligned}$$

Hence, after diagonalisation,

$$\begin{aligned} \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix} &= \begin{pmatrix} 1+\sqrt{5} & 1-\sqrt{5} \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}^{n-1} \begin{pmatrix} \frac{1}{2\sqrt{5}} & -\frac{1-\sqrt{5}}{4\sqrt{5}} \\ -\frac{1}{2\sqrt{5}} & \frac{1+\sqrt{5}}{4\sqrt{5}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1+\sqrt{5} & 1-\sqrt{5} \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}^{n-1} \begin{pmatrix} \frac{1}{2\sqrt{5}} \\ -\frac{1}{2\sqrt{5}} \end{pmatrix} \end{aligned}$$

Fom here, we can find the value of F_n for arbitrary $n \in \mathbb{N}$. It can be verified that

$$\begin{aligned} F_n &= \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n \\ &= \frac{1}{\sqrt{5}} \left[\phi^n - (1-\phi)^n \right] \end{aligned}$$

where $\phi = \frac{1+\sqrt{5}}{2}$ is known as the golden ratio and has great significance in the study of the Fibonacci Sequence.

6.2 Orthogonal Diagonalisation

A square matrix $\mathbf{A}$ is called orthogonally diagonalisable if there exists an orthogonal matrix $\mathbf{P}$ (that is $\mathbf{P}^{\mathrm{T}} = \mathbf{P}^{-1}$) such that $\mathbf{P}^{\mathrm{T}}\mathbf{A}\mathbf{P}$ is a diagonal matrix. The matrix $\mathbf{P}$ is said to orthogonally diagonalise $\mathbf{A}$.

Note that every 2×2 symmetric matrix has real eigenvalues. We will not discuss the case for general $n \times n$, $n > 2$ since we require more information beyond our current knowledge.

Proof: Any 2×2 symmetric matrix $\mathbf{A}$ can be expressed as $\mathbf{A} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$. Then,

$$\begin{aligned} \det(\mathbf{A} - \lambda\mathbf{I}) &= (a-\lambda)(d-\lambda) - b^2 \\ &= \lambda^2 - (a+d)\lambda + ad - b^2 \end{aligned}$$

As

$$\begin{aligned} [-(a+d)]^2 - 4(ad - b^2) &= (a+d)^2 - 4ad + 4b^2 \\ &= (a-d)^2 + 4b^2 \geq 0 \end{aligned}$$

then the equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ has real roots, and the result follows. $\square$

6.2.1 Spectral Theorem

Spectral Theorem

A square matrix $\mathbf{A}$ is orthogonally diagonalisable if and only if $\mathbf{A}$ is symmetric. The proof is trivial since it is clear that diagonal matrices are symmetric.

We state an algorithm to orthogonally diagonalise a matrix.

Orthogonal Diagonalisation Process

Let $\mathbf{A}$ be a symmetric matrix of order n.
Step 1: Find all the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ by solving $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
Step 2: For each eigenvalue λ_i for $1 \le i \le k$, find a basis S_{λ_i} for the eigenspace E_{λ_i} and then use the Gram-Schmidt Process to transform S_{λ_i} into an orthonormal basis T_{λ_i}
Step 3: Let

$$T = \bigcup_{i=1}^{k} T_{\lambda_i},$$

and say $T = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Then, $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{pmatrix}$ is an orthogonal matrix that orthogonally diagonalises $\mathbf{A}$.

6.3 Quadratic Forms and Conic Sections

A quadratic form is a polynomial with terms all of degree two, for instance $4x^2 + 2xy - 3y^2$. Other than in Linear Algebra, they play a pivotal role in Number Theory, in particular, in the study of diophantine equations. Those interested can read up more about it!

The expression

$$\begin{aligned} Q(x_1, x_2, \ldots, x_n) &= \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij} x_i x_j \\ &= q_{11}x_1^2 + q_{12}x_1x_2 + \ldots + q_{1n}x_1x_n + q_{22}x_2^2 + \ldots + q_{2n}x_2x_n + \ldots + q_{nn}x_n^2 \end{aligned}$$

where the q_{ij}'s are real numbers, is called a quadratic form in n variables, namely $x_1, x_2, \ldots, x_n$. We define an $n \times n$ symmetric matrix $\mathbf{A} = (a_{ij})$ such that

$$a_{ij} = \begin{cases} q_{ii} & \text{if } i = j \\ \frac{1}{2}q_{ij} & \text{if } i < j \\ \frac{1}{2}q_{ji} & \text{if } i > j \end{cases}$$

and let $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^{\mathrm{T}}$. It can be verified that

$$Q(x_1, x_2, \ldots, x_n) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}.$$

Alternatively, the quadratic form can be regarded as a mapping $Q : \mathbb{R}^n \to \mathbb{R}$ defined by

$$Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$$

for $\mathbf{x} \in \mathbb{R}^n$.

Let $Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ be a quadratic form in n variables (i.e. $x_1, x_2, \ldots, x_n$), where $\mathbf{A}$ is an $n \times n$ symmetric matrix and $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^{\mathrm{T}}$. We adopt the following technique to simplify the quadratic form. First, we find an orthogonal matrix $\mathbf{P}$ that diagonalises the symmetric matrix $\mathbf{A}$. That is,

$$\mathbf{P}^{\mathrm{T}}\mathbf{A}\mathbf{P} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

Define new variables $y_1, y_2, \ldots, y_n$ such that $\mathbf{y} = \mathbf{P}^{\mathrm{T}}\mathbf{x} = \mathbf{P}^{-1}\mathbf{x}$, where $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \ldots & y_n \end{pmatrix}^{\mathrm{T}}$. Note that $\mathbf{x} = \mathbf{P}\mathbf{y}$. Then, the quadratic form becomes

$$\begin{aligned}
Q(\mathbf{x}) &= Q(\mathbf{Py}) \\
&= (\mathbf{Py})^{\mathrm{T}}\mathbf{A}(\mathbf{Py}) \\
&= \mathbf{y}^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}\mathbf{APy} \\
&= \begin{pmatrix} y_1 & y_2 & \ldots & y_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} \begin{pmatrix} y_1 & y_2 & \ldots & y_n \end{pmatrix}^{\mathrm{T}} \\
&= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2
\end{aligned}$$

Example: Consider the quadratic form $Q_1(x, y) = x^2 - xy + y^2$. We shall simplify the quadratic form and prove that it becomes

$$\frac{1}{4}(x+y)^2 + \frac{3}{4}(y-x)^2.$$

Solution: Considering $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$, where $\mathbf{x}^{\mathrm{T}} = \begin{pmatrix} x & y \end{pmatrix}$ and $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we have

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \begin{pmatrix} a^2x + (b+c)xy + dy^2 \end{pmatrix}.$$

It is clear that $a = d = 1$ and $b + c = -1$. Since $\mathbf{A}$ is symmetric, then $b = c = -\frac{1}{2}$. Alternatively, one can observe that the entries of $\mathbf{A}$ are as such without wasting much time on the matrix multiplication. Next, we use the algorithm that is used to orthogonally diagonalise the matrix $\begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}$.

$$\begin{aligned}
\det(\mathbf{A} - \lambda\mathbf{I}) &= 0 \\
\det\begin{pmatrix} 1-\lambda & -\frac{1}{2} \\ -\frac{1}{2} & 1-\lambda \end{pmatrix} &= 0 \\
(1-\lambda)^2 - \frac{1}{4} &= 0
\end{aligned}$$

Hence, the eigenvalues are $\frac{1}{2}$ and $\frac{3}{2}$ and their corresponding eigenvectors are $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Note that

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$$

is an orthogonal basis so we just have to normalise the vectors to make it an orthonormal basis. We see that the Gram-Schmidt Process is not necessary in this case. Hence, writing everything in the form $\mathbf{P}^{\mathrm{T}}\mathbf{AP} = \mathbf{D}$, we have

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

Next, we set $\mathbf{y} = \mathbf{P}^{\mathrm{T}}\mathbf{x}$, where $\mathbf{y} = \begin{pmatrix} x' \\ y' \end{pmatrix}$, which gives us

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(x+y) \\ \frac{1}{\sqrt{2}}(y-x) \end{pmatrix}.$$

Hence,

$$\begin{aligned}
Q_1(x, y) &= \mathbf{y}^{\mathrm{T}}\mathbf{Dy} \\
&= \frac{1}{2}(x')^2 + \frac{3}{2}(y')^2 \\
&= \frac{1}{4}(x+y)^2 + \frac{3}{4}(y-x)^2
\end{aligned}$$

□

6.3.1 Matrix Representation of Conic Sections

This section brings back some lovely memories of NYJC H2 Further Mathematics Preliminary Examination paper in 2021 and coincidentally, that year's A-Level had two question on the matrix equation of conic sections!

Conic sections are the sets of points whose coordinates satisfy a second-degree polynomial equation in two variables, namely

$$Q(x,y) = ax^2 + bxy + cy^2 + dx + ey = f,$$

where a, b, c, d, e and f are real numbers and a, b, c are not all zero. With reference to matrices, we can rewrite the above as a matrix equation! That is,

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d & e \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = f.$$

By setting $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\mathbf{A} = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} d \\ e \end{pmatrix}$, the matrix equation becomes

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} + \mathbf{b}^{\mathrm{T}}\mathbf{x} = f$$

and the term $ax^2 + bxy + cy^2$ (or $\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$) is known as the quadratic form associated with the quadratic equation.

The graph of a quadratic equation is known as a conic section. We call a conic *degenerate* if it is the empty set, a point, a line or a pair of lines; and it is called *non-degenerate* if it is either a circle, ellipse, hyperbola or a parabola.

First, we introduce the standard form (or canonical form) of conic sections.

Circle:

$$x^2 + y^2 = a^2$$

where a is the radius of the circle

Ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

and if $a > b$, we obtain a horizontal ellipse with semi-major axis a and semi-minor axis b. On the other hand, if $a < b$, we obtain a vertical ellipse with semi-major axis b and semi-minor axis a. Note that the circle is a special case of the ellipse and it is achieved if $a = b$.

With reference to the matrix representation of conic sections, an ellipse of equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

can be expressed as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

Hyperbola:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \text{ or } \frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$$

depending on whether the hyperbola opens left and right or upward and downward. For the first hyperbola which opens left and right, in matrix equation form, it can be expressed as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & -\frac{1}{b^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

For the second hyperbola which opens upwards and downwards, it can be expressed as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} -\frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

Parabola:

$$x^2 = ky \text{ or } y^2 = kx,$$

where $k < 0$ or $k > 0$. In matrix equation form, it can be expressed as either

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 & -k \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \text{ or } \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -k & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

respectively.

Recall that $ax^2 + bxy + cy^2$ is the quadratic form associated with the quadratic equation. The matrix of the quadratic form can be written as $\begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}$. Note that this matrix is symmetric. We define the discriminant, Δ, of a conic section to be the following expression:

$$\Delta = b^2 - 4ac,$$

which is the same as the discriminant of a quadratic equation. Note that this can be obtained via

$$b^2 - 4ac = -4 \det \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}.$$

We make the following remarks:
(i): Q is an ellipse if and only if $\Delta < 0$
(ii): Q is a parabola if and only if $\Delta = 0$
(iii): Q is a hyperbola if and only if $\Delta > 0$

Example: Consider the quadratic equation $x^2 - xy + y^2 - x - y = 1$. We wish to prove that this is an ellipse that is centered at $(1, 1)$.

Solution: Observe that the quadratic form resembles the one discussed in the earlier example, $Q_1(x, y)$. Since

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(x+y) \\ \frac{1}{\sqrt{2}}(y-x) \end{pmatrix}$$

and substituting x' and y' into the quadratic equation yields

$$\begin{aligned} \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} -1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} &= 1 \\ \frac{\left(x' - \sqrt{2}\right)^2}{2^2} + \frac{(y')^2}{\left(\frac{2}{\sqrt{3}}\right)^2} &= 1 \end{aligned}$$

which resembles the standard form of an ellipse. However, the centre of the ellipse is not $(\sqrt{2}, 0)$ as we regard (x', y') as the coordinates of the point (x, y) using a new coordinate system with the x'-axis in the direction of $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and the y'-axis in the direction of $-\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. Substituting $x' = \sqrt{2}$ and $y' = 0$ into the system of equations

$$\begin{aligned} x' &= \frac{1}{\sqrt{2}}(x+y) \\ y' &= \frac{1}{\sqrt{2}}(y-x) \end{aligned}$$

yields $x = y = 1$, which is the centre of the ellipse. □

6.4 Applications

6.4.1 Introduction to Stochastic Processes: Markov Chains

Throwback to the last question of NYJC Further Mathematics Common Test 1 2021!

A Markov Chain is a system which experiences transitions from one state to another according to certain probabilistic rules. The defining characteristic of a Markov Chain is that no matter how the process arrived at its present state, the possible future states are fixed. In other words, the probability of transitioning to any particular state is dependent solely on the current state and time elapsed. The state space, or set of all possible states, can be anything like letters, numbers, weather conditions, etc.

For this section, we will only discuss Discrete-Time Markov Chains (DTMC).

The Markov Property

For any $n \in \mathbb{N}$ and possible states $i_0, i_1, \ldots, i_n$ of the random variables, the Markov Property states that

$$\mathrm{P}\left(X_n = i_n | X_{n-1} = i_{n-1}\right) = \mathrm{P}\left(X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots X_{n-1} = i_{n-1}\right).$$

A transition matrix $\mathbf{P}_t$ for Markov chain $\{X\}$ at time t is a matrix containing information on the probability of transitioning between states. In particular, given an ordering of a matrix's rows and columns by the state space S, the (i,j)-entry of $\mathbf{P}_t$ is

$$(\mathbf{P}_t)_{i,j} = \mathrm{P}(X_{t+1} = j | X_t = i).$$

We define the k-step transition matrix as $\mathbf{P}_t^{(k)}$, where

$$\mathbf{P}_t^{(k)} = \mathbf{P}_t \mathbf{P}_{t+1} \ldots \mathbf{P}_{t+k-1}.$$

It can be shown that the (i,j)-entry of a 2-step transition matrix is

$$(\mathbf{P}_t \mathbf{P}_{t+1})_{i,j} = \mathrm{P}(X_{t+2} = j | X_t = i).$$

Proof:

$$\begin{aligned}
(\mathbf{P}_t \mathbf{P}_{t+1})_{i,j} &= \sum_{k=1}^{n} (P_t)_{i,k} (P_{t+1})_{k,j} \\
&= \sum_{k=1}^{n} \mathrm{P}\left(X_{t+1} = k | X_t = i\right) \mathrm{P}\left(X_{t+2} = j | X_{t+1} = k\right) \\
&= \mathrm{P}\left(X_{t+2} = j | X_t = i\right)
\end{aligned}$$

where the final equality follows from conditional probability □

Hence, the k-step transition matrix is

$$\mathbf{P}_t^{(k)} = \begin{pmatrix} \mathrm{P}\left(X_{t+k} = 1 | X_t = 1\right) & \mathrm{P}\left(X_{t+k} = 2 | X_t = 1\right) & \ldots & \mathrm{P}\left(X_{t+k} = n | X_t = 1\right) \\ \mathrm{P}\left(X_{t+k} = 1 | X_t = 2\right) & \mathrm{P}\left(X_{t+k} = 2 | X_t = 2\right) & \ldots & \mathrm{P}\left(X_{t+k} = n | X_t = 2\right) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{P}\left(X_{t+k} = 1 | X_t = n\right) & \mathrm{P}\left(X_{t+k} = 2 | X_t = n\right) & \ldots & \mathrm{P}\left(X_{t+k} = n | X_t = n\right) \end{pmatrix}.$$

Since the total of transition probability from a state i to all other states must be 1, then the sum of the row entries is equal to 1. We say that such a matrix is *right stochastic*.

Example: The following is a simple Markov Chain involving two states, namely A and B. The probability that A transits to itself and to B are 0.2 and 0.8 respectively, whereas the probability that B transits to itself and to A are 0.6 and 0.4 respectively. In the study of Markov Chains, we are interested in the probability that a process beginning at either state A or B would end up at either A or B after k moves.

We will investigate the probability that a process beginning at state A would end up at B after 2 moves, and after k moves.

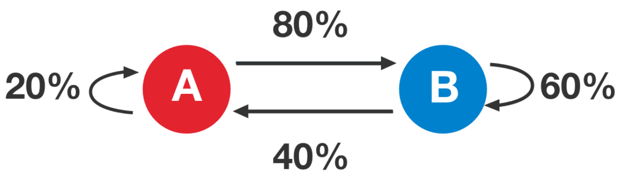

Figure 23: Simple Markov Chain involving two states

Solution: The probability that a process beginning at state A would end up at B after 2 moves is easy to calculate. It is simply

$$0.2(0.8) + 0.8(0.6) = 0.64.$$

It would be tedious to list the cases for general $n \in \mathbb{N}$ since there are many permutations from the first to the $(k-1)^{\text{th}}$ transition that would yield the same outcome. This is where the power of matrices (both literally and figuratively) comes in.

We consider a matrix

$$\mathbf{A} = \begin{pmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{pmatrix}$$

and we make a couple of observations. Since $\mathbf{A}$ is a right stochastic matrix, then the sum of each row entry is 1. Also, the diagonal entries correspond to the respective probabilities that a state transits to itself.

It is easy to compute

$$\mathbf{A}^2 = \begin{pmatrix} 0.36 & 0.64 \\ 0.32 & 0.68 \end{pmatrix}.$$

Note that the a_{12} entry is 0.64, which corresponds to the probability that a process beginning at state A would end up at B after 2 moves. Hence, to find $\mathbf{A}^k$, we need to diagonalise $\mathbf{A}$ by first finding its eigenvalues and eigenvectors.

$$\mathbf{A}^k = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{5} \end{pmatrix}^k \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

and the simplification will be left as an exercise. Recall that each (i,j)-entry for $i = 1,2$ and $j = 1,2$ represents a probability corresponding to a certain scenario. $\square$

Of course, the curious reader would ask that if this process were to continue indefinitely, what would occur? Thus, we set $k \to \infty$ and the diagonal matrix will tend to $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, and so

$$\lim_{k \to \infty} \mathbf{A}^k = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

We say that the system approaches its *steady state* as $k \to \infty$.

Markov Chains have a ton of usages in our everyday lives. Search the Internet for articles and detailed explanations on some of these interesting topics (not exhaustive):
(i): Genetics: The Hardy-Weinberg Principle
(ii): Population Dynamics: Matrix Population Model
(iii): Page Rank: The Mathematics of Google Search

7 Linear Transformations

7.1 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

A linear transformation is a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ of the form

$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}$$

for $\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^n$.

If $n = m$, then T is also called a linear operator on $\mathbb{R}^n$.

We can rewrite the formula of T as

$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

and the matrix $(a_{ij})_{m \times n}$ is called the standard matrix for/matrix representation of T.

Now, we provide a definition of a linear transformation. That is, for vector spaces V and W, a mapping $T : V \to W$ is called a linear transformation if and only if

$$T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v})$$

for all $\mathbf{u}, \mathbf{v} \in V$ and $\alpha, \beta \in \mathbb{R}$. The two definitions of a linear transformation are the same if $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$.

Properties of Linear Transformations

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then,
(1): $T(\mathbf{0}) = \mathbf{0}$
(2): If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ and $c_1, c_2, \ldots, c_k \in \mathbb{R}$, then

$$T\left(\sum_{i=1}^{k} c_i\mathbf{u}_i\right) = \sum_{i=1}^{k} c_i T(\mathbf{u}_i)$$

which shows that a linear transformation preserves linear combinations. The second property is analogous to derivatives and integrals, which is why they are referred to as linear operators.

Note that to prove that T is a linear transformation, we need to prove the two properties mentioned but to disprove the statement, we just need to provide a counterexample.

Example: Let $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T_1\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+1 \\ y+3 \end{pmatrix} \text{ for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

It is clear that T_1 is not a linear transformation since

$$T_1\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

so the output vector is non-zero, contradicting the first property.

7.1.1 Identity Transformation

The identity transformation (or identity mapping) $I : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$I(\mathbf{x}) = I\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{x}$$

for $\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^n$. I is a linear operator on $\mathbb{R}^n$ and the standard matrix for I is the identity matrix $\mathbf{I}_n$.

7.1.2 Zero Transformation

The zero transformation (or zero mapping) $O : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$O(\mathbf{x}) = O\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$$

for $\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}^{\mathrm{T}} \in \mathbb{R}^n$. O is a linear transformation and the standard matrix for O is the zero matrix $\mathbf{0}_{m\times n}$.

7.1.3 Bases for $\mathbb{R}^n$ and Standard Matrices

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for $\mathbb{R}^n$. Given any vector $\mathbf{v} \in \mathbb{R}^n$, we can write $\mathbf{v}$ as a linear combination of the $\mathbf{u}_i$'s. That is,

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_n\mathbf{u}_n$$

for some $c_1, c_2, \ldots, c_n \in \mathbb{R}$. For a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we have

$$T(\mathbf{v}) = c_1T(\mathbf{u}_1) + c_2T(\mathbf{u}_2) + \ldots + c_nT(\mathbf{u}_n).$$

In other words, the image $T(\mathbf{v})$ of $\mathbf{v}$ is completely determined by the images $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$ of the basis vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear transformation such that

$$T\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \; T\left(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \text{ and } T\left(\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ -1 \end{pmatrix}.$$

Note that $\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}\right\}$ is a basis for $\mathbb{R}^3$, and thus the image $T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right)$ of every $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ is completely determined by the images of the three basis vectors. A question we would like to ask is what is the matrix representation of T?

Solution: Aforementioned, any vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ can be expressed as a linear combination of the basis vectors. Hence,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = c_1\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c_2\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + c_3\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}.$$

We solve for the unknowns c_1, c_2 and c_3 and express them in terms of x, y and z. Note that

$$\begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix} \xrightarrow{\text{REF}} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & -1 \end{pmatrix}.$$

which implies that

$$\begin{aligned} c_1 + 2c_3 &= x \\ c_2 - 2c_3 &= y \\ -c_3 &= z \end{aligned}$$

Thus,

$$\begin{aligned} c_1 &= x - 2y + 2z \\ c_2 &= -x + 3y - 2z \\ c_3 &= y - z \end{aligned}$$

Now, we can obtain the matrix representation, which is

$$\begin{aligned} T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) &= (x-2y+2z)T\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right) + (-x+3y-2z)T\left(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right) + (y-z)T\left(\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}\right) \\ &= (x-2y+2z)\begin{pmatrix} 1 \\ 3 \end{pmatrix} + (-x+3y-2z)\begin{pmatrix} -1 \\ 2 \end{pmatrix} + (y-z)\begin{pmatrix} 4 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} 2x - y \\ x - y + 3z \end{pmatrix} \\ &= \begin{pmatrix} 2 & -1 & 0 \\ 1 & -1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} \end{aligned}$$

□

We make an observation. Recall that the aforementioned linear transformation is a mapping from $\mathbb{R}^3$ to $\mathbb{R}^2$ and its matrix representation is a 2×3 matrix. In general, for a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, its matrix representation is an $m \times n$ matrix.

Instead of computing the formula for T directly, we can find the standard matrix using the images of the basis vectors of the standard basis. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $\mathbf{A} = (a_{ij})_{m \times n}$ be the standard matrix for T. Take the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$, where each $\mathbf{e}_i$ is a column vector with 1 in the i^{th} row and 0 for the rest of the rows, where $1 \le i \le n$.

We have the following result:

$$T(\mathbf{e}_i) = \mathbf{A}\mathbf{e}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix},$$

which is the i^{th} column of $\mathbf{A}$. Hence, $\mathbf{A} = \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \ldots & T(\mathbf{e}_n) \end{pmatrix}$.

Example: We use the example stated earlier. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear transformation such that

$$T\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \; T\left(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \text{ and } T\left(\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ -1 \end{pmatrix}.$$

Solution: It is easy to find $T(\mathbf{e}_1)$, which is

$$\begin{aligned} T\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right) &= T\left(\begin{pmatrix}1\\1\\1\end{pmatrix}\right) - T\left(\begin{pmatrix}0\\1\\1\end{pmatrix}\right) \\ &= \begin{pmatrix}1\\3\end{pmatrix} - \begin{pmatrix}-1\\2\end{pmatrix} \\ &= \begin{pmatrix}2\\1\end{pmatrix} \end{aligned}$$

The rest of the $T(\mathbf{e}_i)$'s are not difficult to calculate too. It will be left as an exercise. Through this method, we will have the same conclusion as earlier. □

If you are observant enough, you will be able to see that the following matrix

$$\begin{pmatrix}1 & -1 & 4\\3 & 2 & -1\end{pmatrix}\begin{pmatrix}1 & 0 & 2\\1 & 1 & 0\\1 & 1 & -1\end{pmatrix}^{-1},$$

upon evaluation, is the matrix representation of T as well.

7.1.4 Composition of Linear Transformations

Recall that in H2 Mathematics, under the topic of Functions, for the composite function (or simply the composition of) $f \circ g$ to exist, then the range of g must be a subset of the domain of f. We have a similar, but more formal idea for this at a higher level. It involves mapping. Given functions $f : X \to Y$ and $g : A \to B$, then a sufficient condition for the composition $f \circ g$ to exist is that $B \subseteq X$.

As linear transformations are regarded as mappings, we shall now introduce the idea of the composition of linear transformations.

Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations. The composition of T with S, denoted by $T \circ S$, is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^k$ such that

$$(T \circ S)(\mathbf{u}) = T(S(\mathbf{u})) \text{ for } \mathbf{u} \in \mathbb{R}^n.$$

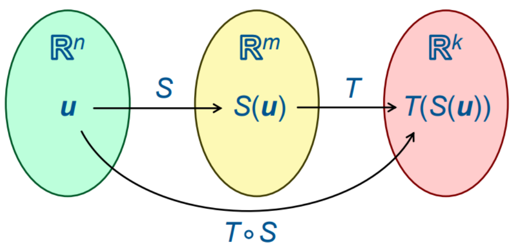

Figure 24: A composition of linear transformations

Moreover, $T \circ S : \mathbb{R}^n \to \mathbb{R}^k$ is a linear transformation too. If $\mathbf{A}$ and $\mathbf{B}$ are the standard matrices for the linear transformations S and T respectively, then the standard matrix for $T \circ S$ is $\mathbf{BA}$.

Proof: For all $\mathbf{u} \in \mathbb{R}^n$, $(T \circ S)(\mathbf{u}) = T(S(\mathbf{u})) = T(\mathbf{Au}) = \mathbf{BAu}$. □

7.2 Range and Rank

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The range of T, denoted by $\mathrm{R}(T)$, is the set of images T. That is,

$$\mathrm{R}(T) = \{T(\mathbf{u})|\mathbf{u} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for $\mathbb{R}^n$. As the image of every vector $\mathbf{v} \in \mathbb{R}^n$ under T is a linear combination of $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$, then it is clear that $\mathrm{R}(T) \subseteq \operatorname{span}\{T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)\}$. Also, every linear combination of $T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)$ is an element of $\mathrm{R}(T)$, then considering the subset relationships, we conclude that

$$\mathrm{R}(T) = \operatorname{span}\{T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)\}.$$

Of course, this section has significant links with vector spaces. We shall establish a connection between $\mathrm{R}(T)$ and the column space of a matrix representation of the linear transformation.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $\mathbb{A}$ be the standard matrix for T. Then,

$$\mathrm{R}(T) = \text{ column space of } \mathbf{A},$$

which is a subspace of $\mathbb{R}^m$. In relation to the rank of a matrix, for a linear transformation T, the dimension of $\mathrm{R}(T)$ is called the rank of T, denoted by $\operatorname{rank}(T)$. As $\mathbf{A}$ is the standard matrix for T, then $\operatorname{rank}(T) = \operatorname{rank}(\mathbf{A})$.

Example: Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be a linear transformation defined by

$$T\left(\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+2y+z \\ x+3y \\ x+4y-z \\ y-z \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} \in \mathbb{R}^4.$$

We wish to find a basis for the range of T and compute its rank.

Solution: It is clear that the matrix representation of the linear transformation is

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

and after Gaussian Elimination, we obtain the matrix

$$\begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

As the pivots are in the second and third columns, a basis for $\mathrm{R}(T)$ is $\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 4 \\ 1 \end{pmatrix}\right\}$ and so $\operatorname{rank}(T) = 2$. $\square$

7.3 Kernel and Nullity

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The kernel of T, denoted by $\ker(T)$, is the set of vectors in $\mathbb{R}^n$ whose image is the zero vector in $\mathbb{R}^m$. That is,

$$\ker(T) = \{\mathbf{u} | T(\mathbf{u}) = \mathbf{0}\} \subseteq \mathbb{R}^n.$$

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $\mathbf{A}$ the standard matrix for T. Then,

$$\ker(T) = \text{ the nullspace of } \mathbf{A},$$

which is a subspace of $\mathbb{R}^n$. Let T be a linear transformation. The dimension of $\ker(T)$ is called the nullity of T and is denoted by $\operatorname{nullity}(T)$. If $\mathbf{A}$ is the standard matrix for T, then $\operatorname{nullity}(T) = \operatorname{nullity}(\mathbf{A})$.

Example: We use the example that was stated under range and rank. Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be a linear transformation defined by

$$T\left(\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+2y+z \\ x+3y \\ x+4y-z \\ y-z \end{pmatrix} \text{ for } \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} \in \mathbb{R}^4.$$

We wish to find a basis for the kernel of T and compute its nullity.

Solution: Recall that the row-echelon form of the matrix representation of T is

$$\begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This is equivalent to the following system of equations:

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

As there are 2 equations and 4 unknowns, then the degree of freedom is 2. We set $w = \lambda$ and $z = \mu$. It is easy to see that $y = \mu$ and $x = -3\mu$, implying that a basis for the kernel is $\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \\ 1 \\ 1 \end{pmatrix}\right\}$ and since there are two vectors in the basis, then it implies that $\text{nullity}(T) = 2$. $\square$

7.3.1 Invertible Matrix Theorem

Now, we will state the last property of the Invertible Matrix Theorem, which is just a continuation of the previous one mentioned. For an $n \times n$ matrix $\mathbf{A}$,
(13): The linear transformation $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is one-one and maps $\mathbb{R}^n$ onto $\mathbb{R}^n$

Invertible Matrix Theorem

To conclude, these are the thirteen properties of the Invertible Matrix Theorem. For an $n \times n$ matrix $\mathbf{A}$,
(1): $\mathbf{A}$ is invertible
(2): The linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution
(3): The RREF of $\mathbf{A}$ is an identity matrix
(4): $\mathbf{A}$ can be expressed as a product of elementary matrices
(5): $\det(\mathbf{A}) \neq 0$
(6): The rows of $\mathbf{A}$ form a basis for $\mathbb{R}^n$
(7): The columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$
(8): the column space of $\mathbf{A} = \mathbb{R}^n$
(9): $\text{rank}(\mathbf{A}) = n$
(10): $\text{nullity}(\mathbf{A}) = 0$
(11): The nullspace of $\mathbf{A}$ is the zero vector. That is, $\{\mathbf{0}\}$.
(12): 0 is not an eigenvalue of $\mathbf{A}$
(13): The linear transformation $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is one-one and maps $\mathbb{R}^n$ onto $\mathbb{R}^n$

7.4 Geometric Transformations in $\mathbb{R}^2$

7.4.1 Translation

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the translation

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+a \\ y+b \end{pmatrix} \text{ for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2,$$

where a and b are real constants. The transformation involves translating a figure on the xy-plane a units and b units in the positive x-direction and positive y-direction respectively. If a and b are both non-zero, then T is not a linear transformation since if $x = y = 0$, then the output vector is non-zero.

7.4.2 Scaling

Let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that

$$S\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} \text{ and } S\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ \lambda_2 \end{pmatrix}$$

for some positive real numbers λ_1 and λ_2. The standard matrix for S is

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

and

$$S\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda_1 x \\ \lambda_2 y \end{pmatrix} \text{ for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

The effect of S is to scale by a factor of λ_1 along the x-axis and by a factor of λ_2 along the y-axis. Hence, S is a scaling along the x- and y-axes by factors of λ_1 and λ_2 respectively. We state two special cases, which occur when $\lambda_1 = \lambda_2 = \lambda$.

(i): S is a dilation if $\lambda > 1$

(ii): S is a contraction if $\lambda < 1$

7.4.3 Reflection

Let $F_1 : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that

$$F_1\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } F_1\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

The standard matrix for F_1 is

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

and

$$F_1\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix} \text{ for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

F_1 is the reflection about the x-axis.

Similarly, the reflection $F_2 : \mathbb{R}^2 \to \mathbb{R}^2$ about the y-axis has the standard matrix

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

implying that

$$F_2\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} -x \\ y \end{pmatrix} \text{ for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

Let $F_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection about the line $y = x$. Then,

$$F_3\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix} \text{ and } F_3\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}1\\0\end{pmatrix}.$$

The standard matrix for F_3 is

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

and

$$F_3\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \begin{pmatrix}y\\x\end{pmatrix} \text{ for } \begin{pmatrix}x\\y\end{pmatrix} \in \mathbb{R}^2.$$

Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection about the line $y = mx$, with $m = \tan\theta$, where θ is the angle between the x-axis and the line. The standard matrix of F is

$$\begin{pmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{pmatrix}.$$

Proof: It is because

$$F(\mathbf{e}_1) = \begin{pmatrix}\cos 2\theta\\ \sin 2\theta\end{pmatrix} \text{ and } F(\mathbf{e}_2) = \begin{pmatrix}\sin 2\theta\\ -\cos 2\theta\end{pmatrix}.$$

□

The formula for F can also be written as

$$F(\mathbf{u}) = \mathbf{u} - 2(\mathbf{u} \bullet \mathbf{n})\mathbf{n} \text{ for } \mathbf{u} \in \mathbb{R}^2,$$

where $\mathbf{n} = \begin{pmatrix}\sin\theta\\ -\cos\theta\end{pmatrix}$.

7.4.4 Shear

A shear mapping is a linear map that displaces each point in a fixed direction, by an amount proportional to its signed distance from the line that is parallel to that direction and goes through the origin. The following diagram depicts that of a horizontal shear.

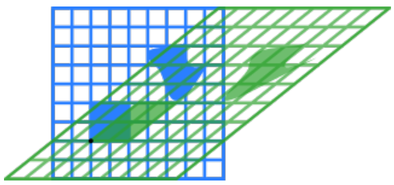

Figure 25: Horizontal shear

A mapping $H : \mathbb{R}^2 \to \mathbb{R}^2$ is called a shear in the x-direction by a factor of k if

$$H\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \begin{pmatrix}x + ky\\y\end{pmatrix} \text{ for } \begin{pmatrix}x\\y\end{pmatrix} \in \mathbb{R}^2.$$

The standard matrix for H is

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

7.4.5 Rotation

Let $R : \mathbb{R}^2 \to \mathbb{R}^2$ be an anti-clockwise rotation about the origin through an angle θ. The standard matrix for R is

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}.$$

Proof: We use polar coordinates. Suppose $(x, y) = (r\cos\alpha, r\sin\alpha)$ and (x', y') are the coordinates of the point after the rotation by θ. Then,

$$\begin{aligned}(x', y') &= (r\cos(\alpha+\theta), r\sin(\alpha+\theta)) \\ &= (r\cos\alpha\cos\theta - r\sin\alpha\sin\theta, r\sin\alpha\cos\theta + r\cos\alpha\sin\theta) \\ &= (x\cos\theta - y\sin\theta, y\cos\theta + x\sin\theta)\end{aligned}$$

□

Note that the rotation matrix is orthogonal.

7.5 Geometric Transformations in $\mathbb{R}^3$

7.5.1 Translation

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the translation

$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+a \\ y+b \\ z+c \end{pmatrix} \text{ for } \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3,$$

where a, b and c are real constants. The transformation involves translating a figure a units, b units and c units in the positive x-direction, positive y-direction and positive z-direction respectively. If a, b and c are all non-zero, then T is not a linear transformation since if $x = y = z = 0$, then the output vector is non-zero.

7.5.2 Scaling

The standard matrix for the scaling along the x, y and z-axes in $\mathbb{R}^3$ by factors of λ_1, λ_2 and λ_3 respectively is

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Similarly, we state two special cases, which occur when $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$.
(i): the scaling is a dilation if $\lambda > 1$
(ii): the scaling is a contraction if $\lambda < 1$

7.5.3 Reflection

The standard matrices for reflections about the xy-plane, xz-plane and yz-plane in $\mathbb{R}^3$ are

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

respectively.

Moreover, if $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a mapping such that

$$T(\mathbf{u}) = \mathbf{u} - 2\left(\frac{\mathbf{u} \bullet \mathbf{n}}{\mathbf{n} \bullet \mathbf{n}}\right)\mathbf{n} \text{ for } \mathbf{u} \in \mathbb{R}^3,$$

where $\mathbf{n} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is a non-zero vector, then T is the reflection about the plane $ax + by + cz = 0$ in $\mathbb{R}^3$.

7.5.4 Shear

A mapping $H' : \mathbb{R}^3 \to \mathbb{R}^3$ is called a shear in the x-direction by a factor of k_1 and in the y-direction by a factor of k_2 if

$$H'\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x + k_1 z \\ y + k_2 z \\ z \end{pmatrix} \text{ for all } \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

H' is a linear transformation with the standard matrix

$$\begin{pmatrix} 1 & 0 & k_1 \\ 0 & 1 & k_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

7.5.5 Rotation

Let $R : \mathbb{R}^3 \to \mathbb{R}^3$ be the anti-clockwise rotation about the z-axis through an angle θ. As the z-axis is fixed under the rotation, then $R(\mathbf{e}_3) = \mathbf{e}_3$. The xy-plane is rotated in the same manner as the rotation in $\mathbb{R}^2$ as discussed previously. Thus, the standard matrix for R is

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Similarly, the standard matrices for rotations in $\mathbb{R}^3$ about the x-axis and the y-axis are

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \text{ and } \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$$

respectively.

7.6 Special Matrices and Linear Transformations

7.6.1 Isometric Transformation

An isometric transformation (also known as a rigid transformation or a Euclidean transformation) is a transformation in Euclidean Space that preserves the Euclidean Distance between every pair of points. Isometric transformations include rotations, translations, reflections or any sequence of these. It is an example of an *affine transformation*, which simply said is a linear transformation which preserves points, straight lines, and planes. The transformation is *invariant*.

Suppose a linear operator T on $\mathbb{R}^n$ is called an isometry. Then, for all $\mathbf{u} \in \mathbb{R}^n$,

$$\|T(\mathbf{u})\| = \|\mathbf{u}\|.$$

Firstly, we would like to prove the following result:

THEOREM

If T is an isometry on $\mathbb{R}^n$, for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$T(\mathbf{u}) \bullet T(\mathbf{v}) = \mathbf{u} \bullet \mathbf{v}.$$

Proof:

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) \bullet T(\mathbf{u}+\mathbf{v}) &= \|T(\mathbf{u}+\mathbf{v})\|^2 \\ &= \|\mathbf{u}+\mathbf{v}\|^2 \text{ by the definition of an isometry} \\ &= \|\mathbf{u}\|^2 + 2(\mathbf{u} \bullet \mathbf{v}) + \|\mathbf{v}\|^2 \end{aligned}$$

Next, we start the left side of the equation by using the linearity property of a linear transformation.

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) \bullet T(\mathbf{u}+\mathbf{v}) &= [T(\mathbf{u}) + T(\mathbf{v})] \bullet [T(\mathbf{u}) + T(\mathbf{v})] \\ &= \|T(\mathbf{u})\|^2 + 2[T(\mathbf{u}) \bullet T(\mathbf{v})] + \|T(\mathbf{v})\|^2 \end{aligned}$$

Lastly, we make the following comparison:

$$\|\mathbf{u}\|^2 + 2(\mathbf{u} \bullet \mathbf{v}) + \|\mathbf{v}\|^2 = \|T(\mathbf{u})\|^2 + 2[T(\mathbf{u}) \bullet T(\mathbf{v})] + \|T(\mathbf{v})\|^2$$

and the result follows. □

7.6.2 Nilpotent Matrix and Transformation

A nilpotent matrix is a square matrix $\mathbf{N}$ such that $\mathbf{N}^k = \mathbf{0}$ for some $k \in \mathbb{N}$. A nilpotent transformation is a linear transformation L of a vector space such that $L^k = \mathbf{0}$ for some $k \in \mathbb{N}$.

THEOREM

$\mathbf{I} - \mathbf{N}$ and $\mathbf{I} + \mathbf{N}$ are invertible matrices.

Proof: We only prove that $\mathbf{I} - \mathbf{N}$ is invertible since the proof that $\mathbf{I} + \mathbf{N}$ is invertible is similar.

$$(\mathbf{I} - \mathbf{N})(\mathbf{I} + \mathbf{N} + \mathbf{N}^2 + \ldots + \mathbf{N}^k) = \mathbf{I}$$

□

THEOREM

A nilpotent transformation L on $\mathbb{R}^n$ naturally determines a flag of subspaces. Let K_j denote the kernel of $\mathbf{N}^j$ and $n_j = \dim(K_j)$. Then,

$$0 = n_0 < n_1 < \ldots < n_{k-1} < n_k = n.$$

7.6.3 Idempotent Matrix

A square matrix $\mathbf{A}$ is idempotent if and only if $\mathbf{A}^2 = \mathbf{A}$. The eigenvalues of an idempotent matrix are 0 or 1. Idempotent matrices arise frequently in Regression Analysis and Econometrics, for example, in the study of the least squares problem.

7.7 Applications

7.7.1 Computer Graphics: The Mathematics behind Animations

Throwback to my Mathematics Club presentation which I did in the middle of J2!

A subdivision surface is a curved surface represented by a *high poly mesh*. The curved surface is the functional limit of an iterative process of subdividing each polygonal face into smaller surfaces that better approximate the underlying final curved surface. Accredited to Edwin Catmull (founder of Pixar) and James Clark, the Catmull-Clark Algorithm is a technique used in 3D computer graphics to create curved surfaces by using subdivision surface modeling.

We start with a mesh of an arbitrary polyhedron (like a cuboid on the left for instance). All the vertices in this mesh shall be called original points. Upon using the Catmull-Clark Algorithm (which will not be discussed in detail), the new mesh will consist only of quadrilaterals, which in general will not be planar. The new mesh will generally look smoother than the old mesh. Repeated subdivision results in meshes that are more and more rounded. This is similar to the idea of Riemann Integration where we subdivide a region into n equally spaced rectangles and take the sum of areas of the rectangles as the number of rectangles tends to infinity.

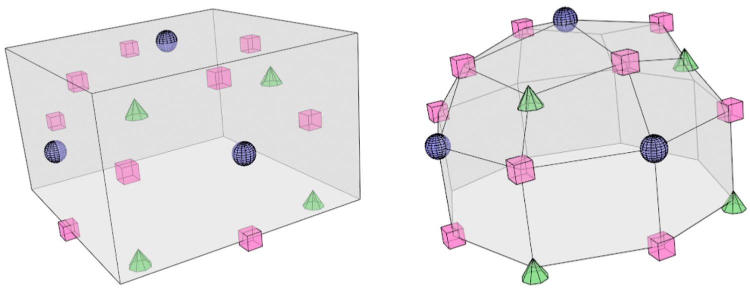

Figure 26: Effect of the Catmull-Clark Algorithm

The limit surface of Catmull–Clark subdivision surfaces can also be evaluated directly, without any recursive refinement. In 1998, a Dutch researcher in the field of computer graphics proposed a method which reformulates the recursive refinement process into a matrix exponential problem, which can be solved directly by means of matrix diagonalisation.

It seems pretty strange that matrices and exponentiation have a relation, but this involves series expansion. Let $\mathbf{A}$ be a square matrix which is diagonalisable. Then, $e^{\mathbf{A}}$ is defined to be

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \ldots + \frac{\mathbf{A}^n}{n!} + \ldots = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}.$$

Since $\mathbf{A}$ is diagonalisable, then we can write $\mathbf{A}^n$ as $\mathbf{P}\mathbf{D}^n\mathbf{P}^{-1}$ for $n \in \mathbb{N}$.

Bézier Curves are also used in computer graphics. A Bézier Curve is a parametric curve comprising a set of discrete control points which defines a smooth, continuous curve by means of a formula. For example, due to the curvature of a blade of grass, it can be modelled by a quadratic Bézier Curve.

A quadratic Bézier Curve is the path traced by the function $\mathbf{B}(t)$, given points $\mathbf{P}_0$, $\mathbf{P}_1$, and $\mathbf{P}_2$. $\mathbf{B}(t)$ can be expressed as

$$\mathbf{B}(t) = (1-t)[(1-t)\mathbf{P}_0 + t\mathbf{P}_1] + t[(1-t)\mathbf{P}_1 + t\mathbf{P}_2],\ 0 \le t \le 1$$

which can be interpreted as the linear interpolant of corresponding points on the linear Bézier Curves from $\mathbf{P}_0$ to $\mathbf{P}_1$ and from $\mathbf{P}_1$ to $\mathbf{P}_2$ respectively.

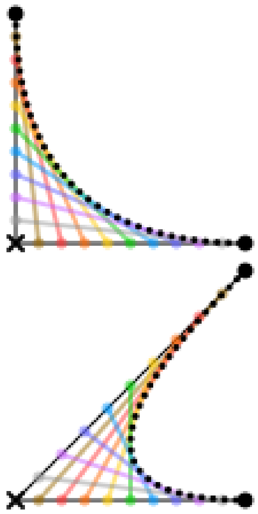

Figure 27: Quadratic Bézier Curve

Sceneries can be generated using fractals. A fractal landscape is a surface generated using a stochastic algorithm to produce a fractal behaviour which mimics the appearance of natural terrain. In April 2022, I came across a video by Inigo Quilez titled 'Painting a Landscape with Maths'. Albeit 42 minutes long, it surprisingly uses substantial amount of Linear Algebra, as well as Calculus, to paint a landscape of a mountainous terrain with clouds.

8 General Vector Spaces

Welcome to Linear Algebra II (module code MA2101)! Here, we shall first talk about a more general, and abstract, framework for vector spaces. However, we must first introduce the idea of a field, denoted by $\mathbb{F}$. We make some remarks, including historical ones, just to lay the foundations of this section.

$\mathbb{R}$ is a complete ordered field since it satisfies the nine field axioms. Sidetrack to a bit of Abstract Algebra, the best known fields are those of $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. Many other fields, such as p-adic fields are commonly used and studied in Mathematics, particularly in Number Theory. There are certain properties which are deemed trivial and shall not be discussed. For example, the trichotomy property states that if $a, b \in \mathbb{R}$, then either $a < b$, $a > b$ or $a = b$, which is intuitive.

The relation of two fields is expressed by the notion of a field extension. It turned out to be a very young French man by the name of Évariste Galois, who laid the foundations of the pinnacle of Abstract Algebra, which is known today as Galois Theory. This was coined sometime in the late 1820s to early 1830s, just before his death in 1832 due to a gunshot wound. He died at the age of 20. As a political activist, he was heavily involved in the political turmoil that revolved around the French Revolution of 1830 (Second French Revolution).

Galois Theory is devoted to understanding the symmetries of field extensions. Among other results, this theory shows that angle trisection and squaring the circle cannot be done with a compass and straightedge. These were problems that were studied by ancient Greek Mathematicians. Galois Theory also shows that quintic equations are, in general, algebraically unsolvable. In fact, Galois proved this result which remained unsolved for over 350 years, just like how Fermat's Last Theorem was unsolved for 358 years. The aforementioned theorem is known as the Abel-Ruffini Theorem, named after Henrik Abel and Paolo Ruffini, for which the former shared a tragic life story, like Galois, and died at 29.

8.1 Fields

A field consists of the following:
(i) a non-empty set $\mathbb{F}$
(ii) an operation of addition $a + b$ between every pair of elements $a, b \in \mathbb{F}$
(iii) an operation of multiplication ab between every pair of elements $a, b \in \mathbb{F}$

We say that the operations satisfy the following 11 field axioms:
(F1) Closure under addition: For all $a, b \in \mathbb{F}$, $a + b \in \mathbb{F}$
(F2) Commutative law for addition: For all $a, b \in \mathbb{F}$, $a + b = b + a$
(F3) Closure under addition: For all $a, b, c \in \mathbb{F}$, $(a + b) + c = a + (b + c)$
(F4) Existence of additive identity: There exists an element $0 \in F$ such that $a + 0 = a$ for all $a \in \mathbb{F}$. 0 is the zero element of $\mathbb{F}$ and all other elements in $\mathbb{F}$ are non-zero elements of $\mathbb{F}$
(F5) Closure under addition: For every $a \in \mathbb{F}$, there exists $b \in \mathbb{F}$ such that $a + b = 0$. b is the additive inverse of a, and we write $b = -a$.
(F6) Closure under multiplication: For all $a, bin\mathbb{F}$, $ab \in \mathbb{F}$
(F7) Commutative law for multiplication: For all $a, b \in \mathbb{F}$, $ab = ba$
(F8) Associative law for multiplication: For all $a, b, c \in \mathbb{F}$, $(ab)c = a(bc)$
(F9) Existence of the multiplicative identity: There exists a non-zero element 1 in $\mathbb{F}$ such that $1a = a$ for all $a \in \mathbb{F}$
(F10) Existence of multiplicative inverse: For every non-zero element $a \in \mathbb{F}$, there exists $c \in \mathbb{F}$ such that $ac = 1$. c is the multiplicative inverse of a, and we write $c = a^{-1}$.
(F11) Distributive law: For all $a, b, c \in \mathbb{F}$, $a(b + c) = ab + ac$

In relation to (F5) and (F10), the additive inverse and multiplicative inverse of a, respectively, are unique.

Note that

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}.$$

It is clear that $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$ satisfy the 11 field axioms, so they are regarded as fields. $\mathbb{N}$ does not satisfy (F4), (F5) and (F10), whereas $\mathbb{Z}$ does not satisfy (F10), hence, $\mathbb{N}$ and $\mathbb{Z}$ are not fields.

Example: Let $\mathbb{F}_2 = \{0, 1\}$. Define the addition and multiplication on $\mathbb{F}_2$ as follows:

$+$	0	1
0	0	1
1	1	0

and

$\times$	0	1
0	0	0
1	0	1

It can be verified that $\mathbb{F}_2$ is a field. However, since the steps are lengthy, we omit the proof. A field which has only finitely many elements is called a finite field. The field $\mathbb{F}_2$ in this example is that of a finite field.

THEOREM

A finite field of q elements exists if and only if $q = p^s$ for some prime p and $s \in \mathbb{N}$.

Now, we state some properties of a field, $\mathbb{F}$, with some related to uniqueness.

(1): Uniqueness of additive identity
If $b, c \in \mathbb{F}$ satisfy $a + b = a + c = a$ for all $a \in \mathbb{F}$, then $b = c$

(2): Uniqueness of additive inverse
For any $a \in \mathbb{F}$, if there exist $b, c \in \mathbb{F}$ such that $a + b = a + c = 0$, then $b = c$

(3): Uniqueness of multiplicative identity
If b, c are non-zero elements in $\mathbb{F}$ satisfying $ba = ca = a$ for all $a \in \mathbb{F}$, then $b = c$

(4): Uniqueness of multiplicative inverse
For any $a \in \mathbb{F}$ and $a \neq 0$, if there exist $b, c \in \mathbb{F}$ such that $ab = ac = 1$, then $b = c$

(5): For any $a \in \mathbb{F}$, $a0 = 0$ and $(-1)a = a$

(6): For any $a, b \in \mathbb{F}$, if $ab = 0$, then $a = 0$ or $b = 0$
(7): Defining subtraction
For any $a, b \in \mathbb{F}$, the subtraction of a by b is defined by $a + (-b)$ and denoted by $a - b$

(8): Defining division
For any $a, b \in \mathbb{F}$, where $b \neq 0$, the division of a by b is defined by ab^{-1}

Back when I was in Year One of Junior College, after the end of our Project Work cycle, my college's Mathematics Club organised an end-of-year project for us to take part it. One of the projects I chose was on inequalities, but before delving into the classical inequalities like AM-GM and Cauchy-Schwarz, the creator of the project, who was a friend of mine by the name of Joseph, gave an introduction to ordered sets and fields so as to define *order between numbers*. I did not know that certain results, which were considered trivial, had rather lengthy proofs. Let us state an example by proving (2) which is on the uniqueness of the additive inverse.

Proof:

$$\begin{aligned} b &= b+0 \because (F4) \\ &= b+(a+c) \because \text{by given assumption} \\ &= (b+a)+c \because (F3) \\ &= (a+b)+c \because (F2) \\ &= 0+c \because \text{by given assumption} \\ &= c+0 \because (F2) \\ &= c \because (F4) \end{aligned}$$

□

For a field $\mathbb{F}$, a linear system with coefficients taken from $\mathbb{F}$ is a linear system over $\mathbb{F}$. A linear system over $\mathbb{R}$ is a real linear system and a linear system over $\mathbb{C}$ is a complex linear system. A matrix with all entries taken from $\mathbb{F}$ is a matrix over $\mathbb{F}$. In particular, a matrix over $\mathbb{R}$ is a real matrix and a matrix over $\mathbb{C}$ is a complex matrix.

The usual Gaussian Elimination discussed in MA2001 also works in this case. We do not have any fancy results when it comes to reducing a matrix containing complex entries to row echelon form.

8.1.1 Trace

Let $\mathbb{F}$ be a field and $\mathbf{A} = (a_{ij})$ an $n \times n$ matrix over $\mathbb{F}$. The trace of $\mathbf{A}$, denoted by $\operatorname{tr}(\mathbf{A})$, is defined to be the sum of the diagonal entries of $\mathbf{A}$. That is,

$$\operatorname{tr}(\mathbf{A}) = a_{11} + a_{22} + \ldots + a_{nn} = \sum_{i=1}^{n} a_{ii}.$$

Note that $\operatorname{tr}(\mathbf{A}) = \operatorname{tr}(\mathbf{A}^{\mathrm{T}})$. We state some other properties of the trace of a matrix.

(1): If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices over $\mathbb{F}$, then $\operatorname{tr}(\mathbf{A}+\mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B})$
(2): If $c \in \mathbb{F}$ and $\mathbf{A}$ is an $n \times n$ matrix over $\mathbb{F}$, then $\operatorname{tr}(c\mathbf{A}) = c \operatorname{tr}(\mathbf{A})$
(3): If $\mathbf{C}$ and $\mathbf{D}$ are $m \times n$ and $n \times m$ matrices respectively over $\mathbb{F}$, then $\operatorname{tr}(\mathbf{CD}) = \operatorname{tr}(\mathbf{DC})$
Proof: Suppose

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \ldots & c_{mn} \end{pmatrix} \text{ and } \mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \ldots & d_{1m} \\ d_{21} & d_{22} & \ldots & d_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \ldots & d_{nm} \end{pmatrix}.$$

Let $\mathbf{P} = \mathbf{CD}$ and $\mathbf{Q} = \mathbf{DC}$. For $\mathbf{P}$, note that

$$p_{ij} = \sum_{k=1}^{n} c_{ik} d_{kj},$$

where $1 \le i, j \le m$, so

$$\operatorname{tr}(\mathbf{P}) = \sum_{i=1}^{m} p_{ii} = \sum_{i=1}^{m} \sum_{k=1}^{n} c_{ik} d_{ki}.$$

In a similar fashion,

$$q_{ij} = \sum_{k=1}^{m} d_{ik} c_{kj},$$

where $1 \le i, j \le n$, so

$$\operatorname{tr}(\mathbf{Q}) = \sum_{i=1}^{n} q_{ii} = \sum_{i=1}^{n} \sum_{k=1}^{m} d_{ik} c_{ki} = \sum_{i=1}^{m} \sum_{k=1}^{n} c_{ik} d_{ki}.$$

Thus, the result follows. □

REMARK

However, we cannot extend the third result to three matrices, say $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ even if the matrices can be multiplied accordingly. That is, in general, $\text{tr}(\mathbf{XYZ}) \neq \text{tr}(\mathbf{YXZ})$.

8.2 Vector Spaces

A vector space comprises the following:
(i) A field $\mathbb{F}$, where the elements in $\mathbb{F}$ are scalars
(ii) A non-empty set V, where the elements in V are vectors
(iii) An operation of vector addition $\mathbf{u}+\mathbf{v}$ between every pair of vectors $\mathbf{u}, \mathbf{v} \in V$
(iv) An operation of scalar multiplication $c\mathbf{u}$ between every $c \in \mathbb{F}$ and every vector $\mathbf{u} \in V$

Moreover, the operations satisfy the following axioms:
(V1) Closure under vector addition: For all $\mathbf{u}, \mathbf{v} \in V, \mathbf{u}+\mathbf{v} \in V$
(V2) Commutative law for vector addition: For all $\mathbf{u}, \mathbf{v} \in V, \mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$
(V3) Associative law for vector addition: For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V, \mathbf{u}+(\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v})+\mathbf{w}$
(V4) Existence of the zero vector: There exists $\mathbf{0} \in V$, called the zero vector, such that $\mathbf{u}+\mathbf{0} = \mathbf{u}$ for all $\mathbf{u} \in V$
(V5) Existence of additive inverse: For every $\mathbf{u} \in V$, there exists $\mathbf{v} \in V$ such that $\mathbf{u}+\mathbf{v} = \mathbf{0}$. $\mathbf{v}$ is the negative of $\mathbf{u}$ and we denote $\mathbf{v}$ by $-\mathbf{u}$.
(V6) Closure under scalar multiplication: For all $c \in \mathbb{F}$ and $\mathbf{u} \in V$, $c\mathbf{u} \in V$
(V7): For all $b, c \in \mathbb{F}$ and $\mathbf{u} \in V$, $b(c\mathbf{u}) = (bc)\mathbf{u}$
(V8): For all $\mathbf{u} \in V$, $1\mathbf{u} = \mathbf{u}$
(V9) Distributive law: For all $c \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$, $c(\mathbf{u}+\mathbf{v}) = c\mathbf{u}+c\mathbf{v}$
(V10) Distributive law: For all $b, c \in \mathbb{F}$ and $\mathbf{u} \in V, (b+c)\mathbf{u} = b\mathbf{u}+c\mathbf{u}$

Some examples of vector spaces are $\mathbb{Q}^n, \mathbb{R}^n, \mathbb{C}^n, \mathbb{F}_2^n$, $\mathcal{M}_{m\times n}$ (the set of all $m \times n$ matrices over $\mathbb{F}$), the set of all infinite sequences, $\mathcal{P}(\mathbb{F})$ (the set of all polynomials over $\mathbb{F}$) and $\mathcal{F}(A, \mathbb{F})$ (the set of all functions $f : A \to \mathbb{F}$).

In addition, in relation to polynomials, we say that $p(x)$ is a polynomial over $\mathbb{F}$ if

$$p(x) = a_0 + a_1x + a_2x^2 + \ldots + a_mx^m,$$

where $a_0, a_1, \ldots, a_m \in \mathbb{F}$. In particular, if $\mathbb{F} = \mathbb{R}$, $p(x)$ is a real polynomial, and it is complex otherwise. That is, $\mathbb{F} = \mathbb{C}$. Let us state more examples of vector spaces.

Example: Let $\mathbb{F}$ be a field and $V = \{\mathbf{0}\}$. Define

$$\mathbf{0}+\mathbf{0} = \mathbf{0} \text{ and } c\mathbf{0} = \mathbf{0} \text{ for } c \in \mathbb{F}.$$

Then, V is a vector space over $\mathbb{F}$ which is called the zero space.

Example: $\mathbb{C} = \{a+bi | a, b \in \mathbb{R}\}$ is a vector sapce over $\mathbb{R}$ using the usual addition of complex numbers as the vector addition and usual multiplication of real numbers to complex numbers as the scalar multiplication. That is,

$$(a+bi)+(c+di) = (a+c)+(b+d)i \text{ for } a+bi, c+di \in \mathbb{C} \text{ with } a, b, c, d \in \mathbb{R}$$

and

$$c(a+bi) = (ca)+(cb)i \text{ for } c \in \mathbb{R} \text{ and } a+bi \in \mathbb{C} \text{ with } a, b \in \mathbb{R}.$$

Example: Let V be the set of all positive real numbers, i.e. $V = \{a \in \mathbb{R} | a > 0\}$. Define the vector addition $\dagger$ by

$$a \dagger b = ab \text{ for } a, b \in V$$

and the scalar multiplication $*$ by

$$m * a = a^m \text{ for } m \in \mathbb{R} \text{ and } a \in V.$$

We shall prove that V is a vector space over $\mathbb{R}$.

Solution: As $0 \in V$, then V is non-empty. Next, we must show that V is closed under addition and scalar multiplication. Thus,

$$(m * a) \dagger b = a^m \dagger b = a^m b.$$

Since $a, b > 0$ by definition, then $a^m b$ is also a vector in V. The result follows. □

Let V be a vector space over a field $\mathbb{F}$. We have the following properties:
(1) Uniqueness of the zero vector: If $\mathbf{v}, \mathbf{w} \in V$ satisfy $\mathbf{u}+\mathbf{v} = \mathbf{u}+\mathbf{w} = \mathbf{u}$ for all $\mathbf{u} \in V$, then $\mathbf{v} = \mathbf{w}$
(2) Uniqueness of the additive inverse: For any $\mathbf{u} \in V$, if there exist $\mathbf{v}, \mathbf{w} \in V$ such that $\mathbf{u}+\mathbf{v} = \mathbf{u}+\mathbf{w} = \mathbf{0}$, then $\mathbf{v} = \mathbf{w}$
(3): For all $\mathbf{u} \in V$, $0\mathbf{u} = \mathbf{0}$ and $(-1)\mathbf{u} = -\mathbf{u}$
(4): For all $c \in \mathbb{F}$, $c\mathbf{0} = \mathbf{0}$
(5): If $c\mathbf{u} = \mathbf{0}$, where $c \in \mathbb{F}$ and $\mathbf{u} \in V$, then $c = 0$ or $\mathbf{u} = \mathbf{0}$

8.3 Subspaces

A subset W of a vector space V is a subspace of V if W itself is a vector space using the same vector addition and scalar multiplication as in V. For a vector space V and letting $\mathbf{0}$ denote the zero vector, the subspaces $\{\mathbf{0}\}$ and V, are called trivial subspaces of V. We say that $\{\mathbf{0}\}$ is the zero space. Other subspaces of V are proper subspaces of V.

Let $\mathbb{F}$ be a field, $V = \mathbb{F}^2$ and $W = \{(a,a) | a \in \mathbb{F}\} \subseteq V$. We have the following properties:
(V1): For any two vectors $(a,a), (b,b) \in W$, the sum $(a,a)+(b,b) = (a+b, a+b)$ is a vector in W. Thus, W is closed under vector addition.
(V4): The zero vector $(0,0)$ of V is contained in W
(V5): For any vector $(a,a) \in W$, the negative, $(-a,-a)$ is also a vector in W
(V6): For any vector $(a,a) \in W$ and scalar $c \in \mathbb{F}$, the scalar multiple $c(a,a) = (ca,ca)$ is also a vector in W. Thus, W is closed under scalar multiplication.

Recall from MA2001 that to verify that a subset W of V is a subspace of V, we need to show that W contains the zero vector, and W is closed under both vector addition and scalar multiplication. Hence, W is a subspace of V if and only if for all $c \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in W$, $\mathbf{u} + c\mathbf{v} \in W$.

Recall that the intersection of two subspaces is also a subspace. We can extend this result to n subspaces.

THEOREM

If $W_1, W_2, \ldots, W_n$ are subspaces of V, then

$$\bigcap_{i=1}^{n} W_i$$

is also a subspace of V.

Proof: It is clear that the intersection is non-empty. Suppose $a, b \in \bigcap_{i=1}^{n} W_i$. Then, $a \in W_i$ for all $1 \le i \le n$. The same can be said for b. By the property of subspaces, for $c \in \mathbb{F}$, $c\mathbf{a} + \mathbf{b} \in W_i$ for all $1 \le i \le n$ and the result follows. □

Sum of Subspaces

The sum of two subspaces, say W_1 and W_2 is also a subspace of a vector space V. We define the sum to be the set

$$W_1 + W_2 = \{\mathbf{w}_1 + \mathbf{w}_2 | \mathbf{w}_1 \in W_1 \text{ and } \mathbf{w}_2 \in W_2\}.$$

Proof: It is clear that $\mathbf{0} \in W_1 + W_2$ so the sum of subspaces is non-empty. Let $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{v}_1, \mathbf{w}_1 \in W_1$ and $\mathbf{v}_2, \mathbf{w}_2 \in W_2$. Let $c \in \mathbb{F}$ too. Then,

$$c\mathbf{v} + \mathbf{w} = c(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{w}_1 + \mathbf{w}_2 = c\mathbf{v}_1 + \mathbf{w}_1 + c\mathbf{v}_2 + \mathbf{w}_2 \in W_1 + W_2.$$

The last line follows due to $c\mathbf{v}_1 + \mathbf{w}_1 \in W_1$ and $c\mathbf{v}_2 + \mathbf{w}_2 \in W_2$ as W_1 and W_2 are subspaces of V. □

Moreover, $W_1 + W_2$ is the smallest subspace of V that contains both W_1 and W_2. Precisely, if U is a subspace of V such that $W_1 \subseteq U$ and $W_2 \subseteq U$, then $W_1 + W_2 \subseteq U$.

8.4 Linear Span and Linear Independence

Let V be a vector space over a field $\mathbb{F}$ and B be a non-empty subset of V. The set of all linear combinations of vectors taken from B is a subspace of V. We say that

$$W = \{\mathbf{u} \in V | \mathbf{u} \text{ is a linear combination of some vectors from } B\}$$

is a subspace of V spanned by B. We write $W = \operatorname{span}(B)$, where it is known that $B \subseteq W$.

8.5 Bases and Dimensions

A subset B of a vector space V is a basis for V if B is linearly independent and B spans V. V is finite dimensional if it has a basis consisting of finitely many vectors, and it is infinite dimensional otherwise.

8.5.1 Zorn's Lemma

Zorn's Lemma

A partially ordered set P has the property that every chain in P has an upper bound in P. Then, P contains at least one maximal element.

Zorn's Lemma can be used to show that every vector space has a basis.

8.5.2 Ordered Basis

Let V be a finite dimensional vector space over a field $\mathbb{F}$, where $\dim(V) = n \geq 1$. A basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for V is an ordered basis if the vectors in B have a fixed order such that $\mathbf{v}_1$ is the first vector, $\mathbf{v}_2$ is the second and so on.

If $\mathbf{u} \in V$ such that

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n,$$

then the coefficients $c_1, c_2, \ldots, c_n$ are the coordinates of $\mathbf{v}$ relative to B. The vector

$$(\mathbf{u})_B = \begin{pmatrix} c_1 & c_2 & \ldots & c_n \end{pmatrix} \text{ or } [\mathbf{u}]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

in $\mathbb{F}^n$ is called the coordinate vector of $\mathbf{u}$ relative to B.

In relation to the coordinate vector relative to a given basis, we have the following two properties:
(i): For any $\mathbf{u}, \mathbf{v} \in V$, $\mathbf{u} = \mathbf{v}$ if and only if $(\mathbf{u})_B = (\mathbf{v})_B$
(ii): For any $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ and $c_1, c_2, \ldots, c_r \in \mathbb{R}$,

$$(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_r\mathbf{v}_r)_B = c_1(\mathbf{v}_1)_B + c_2(\mathbf{v}_2)_B + \ldots + c_r(\mathbf{v}_r)_B.$$

Example: Let us state an example involving matrices. Suppose $B_2 = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4\}$, where $\mathbf{A}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathbf{A}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathbf{A}_1 = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$ and $\mathbf{A}_4 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ are real matrices. Using B_2 as an ordered basis, we wish to find the coordinate of $C = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}$ relative to B_2.

Solution: Consider the system of equations

$$\begin{gathered} c_1\mathbf{A}_1 + c_2\mathbf{A}_2 + c_3\mathbf{A}_3 + c_4\mathbf{A}_4 = \mathbf{C} \\ \begin{pmatrix} c_1 + c_3 + c_4 & c_1 + c_2 + c_3 \\ c_1 + c_2 - c_3 & c_1 - c_3 - c_4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} \end{gathered}$$

Solving the above system will yield the solutions for c_1, c_2, c_3 and c_4 and the required answer would be $(\mathbf{C})_{B_2} = (2, 1, -1, 0)$. □

The dimension of a vector space V over a field $\mathbb{F}$, denoted by $\dim(V)$, is defined to be the number of vectors in a basis for V. In addition, we define the dimension of the zero space to be zero.

Example:

$$\begin{aligned} \dim_{\mathbb{F}}(\mathbb{F}^n) &= n \\ \dim_{\mathbb{C}}(\mathbb{C}^n) &= n \\ \dim_{\mathbb{R}}(\mathbb{C}^n) &= 2n \\ \dim_{\mathbb{F}}(\mathcal{M}_{m\times n}(\mathbb{F})) &= mn \\ \dim_{\mathbb{F}}(\mathcal{P}_n(\mathbb{F})) &= n+1 \end{aligned}$$

THEOREM

If W is a subspace of a finite dimensional vector space V, then
(i): $\dim(W) \le \dim(V)$
(ii): if $\dim(W) = \dim(V)$, then $W = V$

Proof: We shall prove **(i)** first. Note that if W contains only the zero vector, then $\dim(W) = 0$, which of course, satisfies the inequality. Suppose otherwise, that is W contains a non-zero vector $\mathbf{u}_1$ and the set $\{\mathbf{u}_1\}$ is linearly independent. If $\operatorname{span}(\{\mathbf{u}_1\}) \neq W$, then there exists $\mathbf{u}_2 \in W \backslash \operatorname{span}(\{\mathbf{u}_1\})$. Thus, the set $\{\mathbf{u}_1, \mathbf{u}_2\} = \{\mathbf{u}_1\} \cup \{\mathbf{u}_2\}$ is linearly independent set of W. We can repeat this process till we can construct the following linearly independent set of W:

$$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\} = \bigcup_{i=1}^{m} \{\mathbf{u}_i\}$$

Thus, $m \le \dim(V)$, implying that W is finite dimensional and $\dim(W) \le \dim(V)$. □

Next, we prove **(ii)**. Note that any basis B of W is a linearly independent set in V containing $\dim(V)$ vectors, so B is a basis for V. As such, $V = \operatorname{span}(B) = W$. □

8.6 Direct Sums of Subspaces

Let W_1 and W_2 be subspaces of a vector space V. Define $W_1 + W_2$ to be the direct sum of W_1 and W_2 if every vector $\mathbf{u} \in W_1 + W_2$ can be expressed uniquely as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2, \ \mathbf{w}_1 \in W_1 \text{ and } \mathbf{w}_2 \in W_2.$$

We say that the direct sum of W_1 and W_2 is $W_1 \oplus W_2$.

THEOREM

Let W_1 and W_2 be subspaces of a vector space v. Then, $W_1 + W_2$ is a direct sum if and only if $W_1 \cap W_2 = \{\mathbf{0}\}$.

Proof: First, we prove that if $W_1 \cap W_2 = \{\mathbf{0}\}$, then $W_1 + W_2$ is a direct sum. Suppose a vector $\mathbf{u} \in W_1 + W_2$ can be written as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 \text{ and } \mathbf{u} = \mathbf{w}_1' + \mathbf{w}_2',$$

where $\mathbf{w}_1, \mathbf{w}_1' \in W_1$ and $\mathbf{w}_2, \mathbf{w}_2' \in W_2$. Note that we wish to prove that there is a unique representation for $\mathbf{u}$. It is clear that

$$\mathbf{w}_1 - \mathbf{w}_1' = \mathbf{w}_2' - \mathbf{w}_2.$$

Since W_1 is a subspace, then $\mathbf{w}_1 - \mathbf{w}_1' \in W_1$. Also, $\mathbf{w}_2 - \mathbf{w}_2' \in W_2$. Thus, it follows that $\mathbf{w}_1 - \mathbf{w}_1' = \mathbf{w}_2' - \mathbf{w}_2 \in W_1 \cap W_2$. As $W_1 \cap W_2 = \{\mathbf{0}\}$, then

$$\mathbf{w}_1 - \mathbf{w}_1' = \mathbf{w}_2' - \mathbf{w}_2 = \mathbf{0},$$

which implies that $\mathbf{w}_1 = \mathbf{w}_1'$ and $\mathbf{w}_2 = \mathbf{w}_2'$, implying that the sum is unique.

Next, we prove that if $W_1 + W_2$ is a direct sum, then $W_1 \cap W_2 = \{\mathbf{0}\}$. Suppose $\mathbf{w} \in W_1 \cap W_2$. Then, $\mathbf{0} = \mathbf{w} + (-\mathbf{w})$, where $\mathbf{w} \in W_1 \cap W_2 \subseteq W_1$, and $-\mathbf{w} \in W_1 \cap W_2 \subseteq W_2$. On the other hand, $\mathbf{0} = \mathbf{0} + \mathbf{0}$, where the first zero vector is in W_1 and the second zero vector is in W_2. Since $W_1 + W_2$ is a direct sum, then $\mathbf{0}$ can only be written uniquely as $\mathbf{w}_1 + \mathbf{w}_2$ for $\mathbf{w}_1 \in W_1$ and $\mathbf{w}_2 \in W_2$. Hence, $\mathbf{w} = \mathbf{0}$ and the result follows. □

Example: Let W_1 and W_2 denote the sets of symmetric and skew-symmetric $n \times n$ real matrices respectively. That is,

$$W_1 = \left\{\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) | \mathbf{A}^{\mathrm{T}} = \mathbf{A}\right\} \text{ and } W_2 = \left\{\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{R}) | \mathbf{A}^{\mathrm{T}} = -\mathbf{A}\right\}.$$

Note that $W_1$and W_2 are subspaces of $\mathcal{M}_{n\times n}(\mathbb{R})$. Thus, their sum is a subset of the set of $n \times n$ real matrices. Recall in MA2001 that every matrix $\mathbf{B}$ can be written as the sum of a symmetric and a skew-symmetric matrix. That is,

$$\mathbf{B} = \frac{1}{2}(\mathbf{B} + \mathbf{B}^{\mathrm{T}}) + \frac{1}{2}(\mathbf{B} - \mathbf{B}^{\mathrm{T}}),$$

where it is clear that $\frac{1}{2}(\mathbf{B} + \mathbf{B}^{\mathrm{T}}) \in W_1$ and $\frac{1}{2}(\mathbf{B} - \mathbf{B}^{\mathrm{T}}) \in W_2$. Also, $W_1 + W_2 = \mathcal{M}_{n\times n}(\mathbb{R})$ and $W_1 \cap W_2 = \{\mathbf{0}\}$. As such, $W_1 + W_2$ is a direct sum. That is, $\mathcal{M}_{n\times n}(\mathbb{R}) = W_1 \oplus W_2$.

Example: Let W_1 be the subspace of $C([0, 2\pi])$ spanned by $g \in C([0, 2\pi])$, where $g(x) = \sin x$ for $x \in [0, 2\pi]$. Define W_2 to be

$$W_2 = \left\{ f \in C([0, 2\pi]) \,\middle|\, \int_0^{2\pi} f(t) \sin t \, dt = 0 \right\}.$$

Then, W_2 is a subspace of $C([0, 2\pi])$ and $C([0, 2\pi]) = W_1 \oplus W_2$.

Proof: First, we prove that the zero function, O, denoted by $O(x) = 0$ for $x \in [0, 2\pi]$, is in W_2. The zero function is defined by $O : C([0, 2\pi]) \to \mathbb{R}$. It is clear that

$$\int_0^{2\pi} O(t) \sin t \, dt = 0.$$

Next, let $c \in \mathbb{R}$ and take any $f_1, f_2 \in W_2$. Then,

$$\int_0^{2\pi} (cf_1(t) + f_2(t)) \sin t \, dt = c \int_0^{2\pi} f_1(t) \sin t \, dt + \int_0^{2\pi} f_2(t) \sin t \, dt = c(0) + 0 = 0.$$

Thus, we asserted that W_2 is a subspace of $C([0, 2\pi])$.

Next, we wish to prove that $C([0, 2\pi])$ is a direct sum of W_1 and W_2. Suppose $h \in ([0, 2\pi])$. Then, define $f = h - cg$, where

$$c = \frac{1}{\pi} \int_0^{2\pi} h(t) \sin t \, dt.$$

We shall prove that $f \in W_2$. This is true because

$$\int_0^{2\pi} f(t) \sin t \, dt = \int_0^{2\pi} h(t) - cg(t) \, dt = 0,$$

and so $h = f + cg$, which implies that $C([0, 2\pi]) = W_1 + W_2$. Next, let $f \in W_1 \cap W_2$. We wish to prove that $f = O$. Since $f \in W_1$, then $f = ag$ for some $a \in \mathbb{R}$. However, $f \in W_2$ too, so

$$0 = \int_0^{2\pi} f(t) \sin t \, dt = \int_0^{2\pi} a \sin^2 t \, dt = a\pi,$$

therefore, $a = 0$. Hence, $W_1 \cap W_2 = \{O\}$ and the direct sum property follows. □

Let W_1 and W_2 be subspaces of a vector space V. Suppose $W_1 + W_2$ is a direct sum. We have the following properties:

THEOREM

(i): If B_1 and B_2 are bases for W_1 and W_2 respectively, then $B_1 \cup B_2$ is a basis for $W_1 \oplus W_2$
(ii): If W_1 and W_2 are finite dimensional, then

$$\dim(W_1 \oplus W_2) = \dim(W_1) + \dim(W_2)$$

Next, we state a result for subspaces $W_1, W_2, \ldots, W_k$ of V.
(i): The sum

$$W_1 + W_2 + \ldots + W_k = \{\mathbf{w}_1 + \mathbf{w}_2 + \ldots + \mathbf{w}_k | \mathbf{w}_i \in W_i \text{ for } i = 1, 2, \ldots, k\}$$

is a subspace of V

(ii): The subspace $W_1 + W_2 + \ldots + W_k$ is a direct sum of $W_1, W_2, \ldots, W_k$ if every vector $\mathbf{u} \in W_1 + W_2 + \ldots + W_k$ can be expressed uniquely as

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 + \ldots + \mathbf{w}_k$$

with $\mathbf{w}_i \in W_i$ for $1 \le i \le k$. Then, the sum of the W_i's can be expressed as $W_1 \oplus W_2 \oplus \ldots \oplus W_k$.

8.7 Cosets and Quotient Spaces

8.7.1 Cosets

Let W be a subspace of a vector space V. For $\mathbf{u} \in V$, the set

$$W + \mathbf{u} = \{\mathbf{w} + \mathbf{u} | \mathbf{w} \in W\}$$

is called the coset of W containing $\mathbf{u}$.

Example: For example, let W be the subspace of $\mathbb{F}_2^3$ spanned by $(1, 0, 1)$ and $(0, 1, 1)$. Then,

$$W = \{a(1,0,1) + b(0,1,1) | a, b \in \mathbb{F}_2\} = \{(0,0,0), (1,0,1), (0,1,1), (1,1,0)\}.$$

The following are all the cosets of W:

$$\begin{aligned}
W + (0,0,0) &= \{(0,0,0), (1,0,1), (0,1,1), (1,1,0)\} \\
W + (0,0,1) &= \{(0,0,1), (1,0,0), (0,1,0), (1,1,1)\} \\
W + (0,1,0) &= \{(0,1,0), (1,1,1), (0,0,1), (1,0,0)\} \\
W + (0,1,1) &= \{(0,1,1), (1,1,0), (0,0,0), (1,0,1)\} \\
W + (1,0,0) &= \{(10,0,0), (0,0,1), (1,1,1), (0,1,0)\} \\
W + (1,0,1) &= \{(1,0,1), (0,0,0), (1,1,0), (0,1,1)\} \\
W + (1,1,0) &= \{(1,1,0), (0,1,1), (1,0,1), (0,0,0)\} \\
W + (1,1,1) &= \{(1,1,1), (0,1,0), (1,0,0), (0,0,1)\}
\end{aligned}$$

Observe that

$$W + (0,0,0) = W + (0,1,1) = W + (1,0,1) = W + (1,1,0) = W$$

and

$$W+(0,0,1)=W+(0,1,0)=W+(1,0,0)=W+(1,1,1)=\mathbb{F}_2^3-W.$$

Let us state some properties of cosets. Let W be a subspace of a vector space V.

THEOREM

For any $\mathbf{v},\mathbf{w}\in V$, the following are equivalent:

$$\mathbf{v}\in V+\mathbf{w} \iff \mathbf{w}\in W+\mathbf{v} \iff \mathbf{v}-\mathbf{w}\in W \iff W+\mathbf{v}=W+\mathbf{w}$$

Another property is that either $W+\mathbf{v}=W+\mathbf{w}$ or $(W+\mathbf{v})\cap(W+\mathbf{w})=\varnothing$.

We shall prove the first equivalence relation for a start.
Proof: Starting with the first statement that $\mathbf{v}\in V+\mathbf{w}$, note that $V+\mathbf{w}$ is the coset of V containing $\mathbf{w}$. Thus, $\mathbf{v}=\mathbf{u}+\mathbf{w}$ for some $\mathbf{u}\in W$. Then, $\mathbf{w}=(-\mathbf{u})+\mathbf{v}$, implying that $\mathbf{w}\in W+\mathbf{v}$ since we know that W is a subspace implies that the negative of a vector will still be contained within the subspace. □

With the first statement, we shall prove that $\mathbf{v}-\mathbf{w}\in W$. Since $\mathbf{v}-\mathbf{w}=\mathbf{u}$, the result follows. □

Thirdly, starting with the third statement, we shall prove $W+\mathbf{v}=W+\mathbf{w}$. Note that $\mathbf{v}-\mathbf{w}\in W$. Suppose $\mathbf{u}\in W$, then $\mathbf{u}-\mathbf{w}$ is also a vector in W. Thus, $\mathbf{u}\in W+\mathbf{w}$. Since $-\mathbf{w}\in W$, then $\mathbf{u}-\mathbf{v}\in W$, so $\mathbf{u}\in W+\mathbf{v}$. Hence, $W+\mathbf{v}=W+\mathbf{w}$. □

Lastly, given $W+\mathbf{v}=W+\mathbf{w}$, note that $\mathbf{0}\in W$ implies that $\mathbf{v}\in W+\mathbf{v}$, and so $\mathbf{v}\in W+\mathbf{w}$. □

Now, we shall prove the final property.
Proof: Suppose $(W+\mathbf{v})\cap(W+\mathbf{w})\neq\varnothing$. If we can show that $W+\mathbf{v}=W+\mathbf{w}$, then we are done. We know that there exists $\mathbf{x}\in(W+\mathbf{v})\cap(W+\mathbf{w})$. In other words, $\mathbf{x}\in W+\mathbf{v}$ and $\mathbf{x}\in W+\mathbf{w}$. The result follows. □

8.7.2 Addition and Scalar Multiplication of Cosets

Let V be a vector space over a field $\mathbb{F}$ and let W be a subspace of V. Define the addition of two cosets by

$$(W+\mathbf{u})+(W+\mathbf{v})=W+(\mathbf{u}+\mathbf{v})$$

for $\mathbf{u},\mathbf{v}\in V$.

The scalar multiplication of a coset is defined by

$$c(W+\mathbf{u})=W+c\mathbf{u},$$

where $c\in\mathbb{F}$ and $\mathbf{u}\in V$.

8.7.3 Quotient Spaces

Let V be a vector space over a field $\mathbb{F}$ and W a subspace of V. Denote the set of all cosets of W by V/W, that is

$$V/W=\{W+\mathbf{u}|\mathbf{u}\in V\}.$$

We say that V/W is the quotient space of V modulo W and it is a vector space over $\mathbb{F}$ using the addition and scalar multiplication defined previously.

In Abstract Algebra, the term *quotient* is used to define modulo arithmetic for algebraic structures. Take for instance $n\mathbb{Z}=\{0,\pm n,\pm 2n,\ldots\}\subseteq\mathbb{Z}$. We define the quotient space of $\mathbb{Z}$ modulo $n\mathbb{Z}$ by

$$\mathbb{Z}/n\mathbb{Z}=\{n\mathbb{Z}+a|a\in\mathbb{Z}\},$$

where $n\mathbb{Z}+a=\{a,\pm n+a,\pm 2n+a,\ldots\}$. Thus, for $a,b\in\mathbb{Z}$, $n\mathbb{Z}+a=n\mathbb{Z}+b$ if and only if $a\equiv b \pmod n$. We say that a leaves a remainder of b when divided by n. Moreover, the operations of addition and multiplication are defined by

$$(n\mathbb{Z}+a)+(n\mathbb{Z}+b)=n\mathbb{Z}+(a+b) \text{ and } (n\mathbb{Z}+a)(n\mathbb{Z}+b)=n\mathbb{Z}+ab$$

for $n\mathbb{Z}+a, n\mathbb{Z}+b\in\mathbb{Z}/n\mathbb{Z}$. These indeed have a strong semblance with the arithmetic of integer addition and multiplication modulo n.

THEOREM

Let V be a finite dimensional vector space and W a subspace of V. Let $\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m\}$ be a basis for W. Then,

(a): For $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\in V$, $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k,\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_m\}$ is a basis for V if and only if

$$\{W+\mathbf{v}_1,W+\mathbf{v}_2,\ldots,W+\mathbf{v}_k\}$$

is a basis for V/W

(b): $\dim(V/W)=\dim(V)-\dim(W)$

9 General Linear Transformations

9.1 Linear Transformations

Let V and W be two vector spaces over a field $\mathbb{F}$. A linear transformation $T : V \to W$ is a mapping from V to W that satisfies the following two axioms:

(T1) Additivity: For all $\mathbf{u}, \mathbf{v} \in V$, $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
(T2) Scalar multiplication: For all $c \in \mathbb{F}$ and $\mathbf{u} \in V$, $T(c\mathbf{u}) = cT(\mathbf{u})$

If $W = V$, the linear transformation $T : V \to V$ is a linear operator on V. Examples include are the derivative and the integral. If $W = \mathbb{F}$, the linear transformation $T : V \to \mathbb{F}$ is a linear functional on V.

9.2 Examples of Linear Functionals and Linear Operators

9.2.1 Sequences

Let V be the set of all convergent subsequences over $\mathbb{R}$. Note that V is a subspace of $\mathbb{R}^n$. Define a mapping $T : V \to \mathbb{R}$ by

$$T((a_n)_{n\in\mathbb{N}}) = \lim_{n\to\infty} a_n$$

for $(a_n)_{n\in\mathbb{N}} \in V$.

By the linearity property of sequences, for any convergent sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ and any $c \in \mathbb{R}$,

$$\lim_{n\to\infty}(a_n+b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n \text{ and } \lim_{n\to\infty} ca_n = c\lim_{n\to\infty} a_n,$$

so we say that T is a linear functional.

9.2.2 Derivatives and Integrals

Let $[a,b]$, with $a < b$, be a closed interval of the real line. We use the real vector space $C^\infty([a,b])$. Let $D : C^\infty([a,b]) \to C^\infty([a,b])$ be the differential operator such that for every $f \in C^\infty([a,b])$, $D(f)$ is a function in $C^\infty([a,b])$ defined by

$$D(f)(x) = \frac{df(x)}{dx}$$

for $a \le x \le b$.

Let $F : C^\infty([a,b]) \to C^\infty([a,b])$ be the integral opeator such that for every $f \in C^\infty([a,b])$, $F(f)$ is a function in $C^\infty([a,b])$ defined by

$$F(f)(x) = \int_a^x f(t)\ dt$$

for $a \le x \le b$. We say that both D and F are linear operators.

Let V and W be vector spaces over the same field. Suppose V has a basis B. Let $T : V \to W$ be a linear transformation. For every $\mathbf{u} \in V$,

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_m\mathbf{v}_m$$

for some scalars $a_1, a_2, \ldots, a_m$ and some $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in B$. By the properties of linear transformations, it is clear that

$$T(\mathbf{u}) = a_1T(\mathbf{v}_1) + a_2T(\mathbf{v}_2) + \ldots + a_mT(\mathbf{v}_m)$$

and so T is completely determined by the images of the vectors from B.

9.3 Matrices for Linear Transformations

Let $T : V \to W$ be a linear transformation, where V and W are vector spaces over a field $\mathbb{F}$. Suppose V is finite dimensional and has an ordered basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, where $n = \dim(V) \geq 1$. Then, every vector $\mathbf{u} \in V$ can be expressed uniquely as

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_m\mathbf{v}_n$$

for some scalars $a_1, a_2, \ldots, a_n$. Using the idea of coordinate vectors,

$$[\mathbf{u}]_B = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Furthermore, suppose W is finite dimensional and has an ordered basis C with $\dim(W) \geq 1$. Then,

$$[T(\mathbf{u})]_C = \left([T(\mathbf{v}_1)_C \quad T(\mathbf{v}_2)_C \quad \ldots \quad T(\mathbf{v}_n)_C\right)\mathbf{u}]_B.$$

The matrix $\left([T(\mathbf{v}_1)_C \quad T(\mathbf{v}_2)_C \quad \ldots \quad T(\mathbf{v}_n)_C\right)$ completely and uniquely determines the linear transformation T. We call this matrix $[T]_{C,B}$ the matrix for T relative to the ordered bases B and C. Note that

$$[T(\mathbf{u})]_C = [T]_{C,B}[\mathbf{u}]_B$$

for all $\mathbf{u} \in V$. If $W = V$ and $C = B$, $[T]_{B,B}$ is written as $[T]_B$ and the matrix is called the matrix for T relative to the ordered basis B.

9.3.1 Transition Matrices

Suppose V is a finite dimensional vector space with $\dim(V) \geq 1$. Let B and C be two ordered bases for V and consider the identity operator $I_V : V \to V$. Then,

$$[\mathbf{u}]_C = [I_V(\mathbf{u})]_C = [I_V]_{C,B}[\mathbf{u}]_B$$

for all $\mathbf{u} \in V$. As such, for any $\mathbf{u} \in V$, $[I_V]_{C,B}$ is converting the coordinate vector of $\mathbf{u}$ relative to B to the coordinate vector of $\mathbf{u}$ relative to C. Thus, $[I_V]_{C,B}$ is the transition matrix from B to C.

Moreover, $[I_V]_{C,B}$ is invertible and its inverse is the transition matrix from C to B. That is,

$$([I_V]_{C,B})^{-1} = [I_V]_{B,C}.$$

9.4 Compositions of Linear Transformations

Let $S : U \to V$ and $T : V \to W$ be linear transformations. Then, the composition mapping $T \circ S : U \to W$, defined by

$$(T \circ S)(\mathbf{u}) = T(S(\mathbf{u})),$$

for $\mathbf{u} \in U$, is also a linear transformation.

Suppose U, V and W are finite dimensional with $\dim(U), \dim(V), \dim(W) \geq 1$. Let A, B and C be ordered bases for U, V and W respectively. Then,

$$[T \circ S]_{C,A} = [T]_{C,B}[S]_{B,A}.$$

For a linear operator $T : V \to V$ and any non-negative integer m, we define T^m as follows:

$$T^m = \begin{cases} I_V & \text{if } m = 0 \\ \underbrace{T \circ T \circ \ldots \circ T}_{m \text{ times}} & \text{if } m \geq 1. \end{cases}$$

If V is finite dimensional and $\dim(V) \geq 1$, then for an ordered basis B, we have $[T^m]_B = ([T]_B)^m$.

Let $T : V \to V$ be a linear operator, where V is a finite dimensional vector space with $\dim(V) \geq 1$. Suppose B and C are two ordered bases for V. Then,

$$[I_V]_{C,B}[T]_B = [T]_C[I_V]_{C,B}.$$

Proof:

$$[I_V]_{C,B}[T]_B = [I_V]_{C,B}[T]_{B,B} = [I_V \circ T]_{C,B} = [T]_{C,B} = [T \circ I_V]_{C,B} = [T]_{C,C}[I_V]_{C,B} = [T]_C[I_V]_{C,B}$$

and we are done. □

9.4.1 Similar Matrices

Let $\mathbb{F}$ be a field and $\mathbf{A}, \mathbf{D} \in \mathcal{M}_{n\times n}(\mathbb{F})$. Then, $\mathbf{D}$ is similar to $\mathbf{A}$ if there exists an invertible matrix $\mathbf{P} \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that $\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

Alternatively, we can say that $\mathbf{D}$ is similar to $\mathbf{A}$ if and only if there exists a linear operator T on a vector space V, and over $\mathbb{F}$ and of dimension n, such that $\mathbf{D} = [T]_B$ and $\mathbf{A} = [T]_C$ for some ordered bases B and C for V.

As such, we have an alternative interpretation of diagonalisation. To diagonalise a square matrix $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$, it is the same as to finding an ordered basis B for $\mathbb{F}^n$ such that the matrix $[L_{\mathbf{A}}]_B$ is a diagonal matrix.

9.5 The Vector Space $\mathcal{L}(V, W)$

Let V and W be vector spaces over the same field $\mathbb{F}$ and let $\mathcal{L}(V, W)$ be the set of all linear transformations from V to W. Then, $\mathcal{L}(V, W)$ forms a vector space over $\mathbb{F}$ with addition and scalar multiplication defined the same way as how we do so for linear transformations. Moreover, if V and W are finite dimensional, then

$$\dim(\mathcal{L}(V, W)) = \dim(V)\dim(W).$$

9.5.1 Dual Space

If V is a vector space over $\mathbb{F}$, the vector space $\mathcal{L}(V, \mathbb{F})$ of all linear functionals on V is the dual space of V and is denoted by V^*. Hence, if V is finite dimensional,

$$\dim(V) = \dim(V^*).$$

9.6 Kernels and Ranges

THEOREM

Let $T : V \to W$ be a linear transformation. Then, $\ker(T)$ is a subspace of V and $R(T)$ is a subspace of W.

Proof: Since $T(\mathbf{0}_V) = \mathbf{0}_W$, then $\ker(T)$ is non-empty. Suppose $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $c \in \mathbb{R}$. Then,

$$T(c\mathbf{v}_1 + \mathbf{v}_2) = cT(\mathbf{v}_1) + T(\mathbf{v}_2) = c(\mathbf{0}_W) + \mathbf{0}_W = \mathbf{0}_W.$$

Thus, $\ker(T)$ is a subspace of V. □

Let $\mathbf{w}_1, \mathbf{w}_2 \in R(T)$. Then, there exist $\mathbf{v}_1, \mathbf{v}_2 \in \ker(T)$. It is clear that $R(T)$ is non-empty. For $c \in \mathbb{R}$,

$$T(c\mathbf{v}_1 + \mathbf{v}_2) = cT(\mathbf{v}_1) + T(\mathbf{v}_2) = c\mathbf{w}_1 + \mathbf{w}_2 \in R(T),$$

and we are done. □

9.6.1 Coordinate Vectors and the Rank-Nullity Theorem

Let $T : V \to W$ be a linear transformation, where V and W are finite dimensional with $\dim(V) \geq 1$ and $\dim(W) \geq 1$. For any ordered bases B and C for V and W respectively, we have

$$\{[\mathbf{u}]_B | \mathbf{u} \in \ker(T)\} \text{ is in the nullspace of } [T]_{C,B} \text{ and } \text{nullity}(T) = \text{nullity}([T]_{C,B})$$

and

$$\{[\mathbf{v}]_C | \mathbf{v} \in \mathrm{R}(T)\} \text{ is in the column space of } [T]_{C,B} \text{ and } \text{rank}(T) = \text{nullity}([T]_{C,B}).$$

Hence,

$$\text{rank}([T]_{C,B}) = \text{nullity}([T]_{C,B}) = \text{ number of columns in } [T]_{C,B},$$

and so

$$\text{rank}(T) + \text{nullity}(T) = \dim(V).$$

Now, suppose B and C are subsets of V such that B is a basis for $\ker(T)$, $\{T(\mathbf{v}) | \mathbf{v} \in C\}$ is a basis for $R(T)$ and for any $\mathbf{v}, \mathbf{v}' \in C$, if $\mathbf{v} \neq \mathbf{v}'$, then $T(\mathbf{v}) \neq T(\mathbf{v}')$. Then, $B \cup C$ is a basis for V.

9.6.2 Special Mappings

Let $f : A \to B$ be a mapping.

f is *injective* or *one-to-one* if for every $z \in B$, there exists at most one $x \in A$ such that $f(x) = z$.

f is *surjective* or *onto* if for every $z \in B$, there exists at least one $x \in A$ such that $f(x) = z$.

f is *bijective* if it is both injective and surjective.

THEOREM

Let $T : V \to W$ be a linear transformation. Then,
(1): T is injective if and only if $\ker(T) = \{\mathbf{0}\}$
(2): T is surjective if and only if $R(T) = W$

Proof: The second statement is obvious so we shall only prove the first. Suppose $\ker(T) = \{\mathbf{0}\}$. Assume that for some $\mathbf{w} \in W$, there exist $\mathbf{u}, \mathbf{v} \in V$ such that $T(\mathbf{u}) = T(\mathbf{v}) = \mathbf{w}$. Then, $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}$ so $\mathbf{u} - \mathbf{v} \in \ker(T)$. That is, $\mathbf{u} - \mathbf{v} = \mathbf{0}$ and so T is injective as $\mathbf{u} = \mathbf{v}$. The proof for the other direction is obvious. $\square$

We provide two corollaries as results of the above linear transformation.

COROLLARY

T is injective if and only if $\text{nullity}(T) = 0$. If W is finite dimensional, then T is surjective if and only if $\text{rank}(T) = \dim(W)$.

9.7 Isomorphisms

Let $T : V \to W$ be a linear transformation. Then, T is an isomorphism from V to W if T is bijective. We write $V \cong W$.

In Abstract Algebra, the term *isomorphic* is used to denote that two algebraic objects have the same *structure*. For example, consider the mapping $T_1 : \mathcal{P}_n(\mathbb{F}) \to \mathbb{F}^{n+1}$ defined by

$$T_1(a_0 + a_1x + a_2x^2 + \ldots + a_nx^n) = (a_0, a_1, a_2, \ldots, a_n),$$

where $a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \in \mathcal{P}_n(\mathbb{F})$. We say that $\mathcal{P}_n(\mathbb{F})$ is isomorphic to $\mathbb{F}^{n+1}$ and so the two are regarded the same as vector spaces. However, they are different in some other areas. For example, we can apply the techniques of polynomial multiplication on $\mathcal{P}_n(\mathbb{F})$ but vector multiplication does not exist for $\mathbb{F}^{n+1}$.

Example: Let $\mathbb{F}$ be a field and let $T : \mathbb{F}^3 \to \mathcal{P}_2(\mathbb{F})$ be the linear transformation defined by

$$T((a,b,c)) = a + (a+b)x + (a+b+c)x^2$$

for $(a,b,c) \in \mathbb{F}^3$. We shall prove that T is an isomorphism.

Solution: We shall prove that T is injective first. It suffices to check that $\ker(T) = \{(0,0,0)\}$, which is clear because this would imply that $a = 0$, $a+b = 0$ and $a+b+c = 0$ and therefore, $a = b = c = 0$. Hence, T is injective.

Next, suppose $d + ex + fx^2 \in \mathcal{P}_2(\mathbb{F})$. Then, $a + (a+b)x + (a+b+c)x^2 = d + ex + fx^2$, which implies that $a = d$, $b = e - d$ and $c = f - e$. Hence, $T((d, e-f, f-e)) = d + ex + fx^2$ for all $d + ex + fx^2 \in \mathcal{P}_2(\mathbb{F})$, implying that T is surjective.

Since T is bijective, then it is an isomorphism. $\square$

THEOREM

A mapping $T : V \to W$ is bijective if and only if there exists a mapping $S : W \to V$ such that $S \circ T = I_V$ and $T \circ S = I_W$, where I_V and I_W are identity operators on V and W respectively. S is called the inverse of T, and we write $S = T^{-1}$. Thus, a bijective mapping is also called an invertible mapping.

We therefore note that if T is an isomorphism, then T^{-1} is a linear transformation and hence, is also an isomorphism.

Suppose V and W are finite dimensional with $\dim(V) = \dim(W) \geq 1$. Let B and C be ordered bases for V and W respectively.
(1): T is an isomorphism if and only if $[T]_{C,B}$ is invertible
(2): If T is an isomorphism, then $[T^{-1}]_{B,C} = ([T]_{C,B})^{-1}$

Suppose $S : V \to W$ and $T : W \to V$ are linear transformations such that $T \circ S = I_W$, where I_W is the identity operator on W.
(1): S is injective and T is surjective
(2): If V and W are finite dimensional, $\dim(V) = \dim(W)$, then S and T are isomorphisms, with $S^{-1} = T$ and $T^{-1} = S$

We shall first prove the first property.
Proof: First, we show that $\ker(S) = \{\mathbf{0}\}$. Let $\mathbf{u} \in \ker(S)$. Then, $S(\mathbf{u}) = \mathbf{0}$. Hence, $T(S(\mathbf{u})) = T(\mathbf{0}) = \mathbf{0}$, which asserts that S is injective. $\square$

To show that T is surjective, we show that $R(T) = W$. Let $\mathbf{w} \in W$ and $\mathbf{v} \in V$, where $S(\mathbf{w}) = \mathbf{v}$. Then, $T(\mathbf{v}) = T(S(\mathbf{w})) = (T \circ S)(\mathbf{w}) = I_W(\mathbf{w}) = \mathbf{w}$, and so T is surjective. $\square$

Now, let us prove the second property.
Proof: Since T is surjective, then $\text{rank}(T) = \dim(W) = \dim(V)$. Thus, $\text{nullity}(T) = \dim(V) - \text{rank}(T) = 0$, so T is injective. Hence, T is bijective and invertible. As such,

$$T^{-1} \circ (T \circ S) = T^{-1} \circ I_W,$$

which implies that $S = T^{-1}$. Taking the inverse on both sides yields $T^{-1} = S$. $\square$

THEOREM

If V and W are finite dimensional vector spaces over the same field, $V \cong W$ if and only if $\dim(V) = \dim(W)$.

Proof: Suppose $V \cong W$. We wish to prove $\dim(V) = \dim(W)$. Note that $T : V \to W$. Let $S_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for V. We claim that $S_2 = \{T(\mathbf{v}_1), T(\mathbf{v}_2), \ldots, T(\mathbf{v}_n)\}$ is a basis for W. First, we show that $W = \operatorname{span}(S_2)$. Take any arbitrary vector $\mathbf{w} \in W$. Since T is surjective, there exists $\mathbf{u} \in V$ such that $T(\mathbf{u}) = \mathbf{w}$. As $\mathbf{u} \in \operatorname{span}(S_1)$, the result follows.

Next, we shall show that the vectors in S_2 are linearly independent. Suppose $c_1, c_2, \ldots, c_n$ are some scalars such that

$$c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \ldots + c_nT(\mathbf{v}_n) = \mathbf{0}.$$

By the linearity property of linear transformations, $T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n) = \mathbf{0}$, which implies that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n \in \ker(T)$, so T is injective and $\ker(T) = \{\mathbf{0}\}$. Hence,

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n = \mathbf{0}.$$

Since all the $\mathbf{v}_i$'s are linearly independent, where $1 \le i \le n$, then all the c_i's are equal to 0, and the result follows.

Now, suppose $\dim(V) = \dim(W) = n$. We shall prove that $V \cong W$. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ be bases for V and W respectively. Define a linear transformation $T : V \to W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for $1 \le i \le n$. Hence, for any $\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n \in V$, we have

$$T(\mathbf{u}) = a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \ldots + a_n\mathbf{w}_n.$$

T is hence surjective. Moreover, it is clear that

$$a_1T(\mathbf{v}_1) + a_2T(\mathbf{v}_2) + \ldots + a_nT(\mathbf{v}_n) = a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \ldots + a_n\mathbf{w}_n.$$

Since the elements of a basis are unique, then $a_i = 0$ for all $1 \le i \le n$, and so $\ker(T) = \{\mathbf{0}\}$. Hence, T is injective and we conclude that T is bijective. $\square$

We make some remarks about isomorphism related to matrices and polynomials. For a field $\mathbb{F}$,
(1): $\mathcal{M}_{m\times n} \cong \mathbb{F}^{mn}$
(2): $\mathcal{P}_n(\mathbb{F}) \cong \mathbb{F}^{n+1}$
(3): If V and W are finite dimensional vector spaces over $\mathbb{F}$ such that $\dim(V) = n$ and $\dim(W) = m$, then $\mathcal{L}(V, W) \cong \mathbb{F}^{mn} \cong \mathcal{M}_{m\times n}(\mathbb{F})$ (i.e. $V \cong V^*$, where V^* is the dual space of V)

Before we proceed to the last part on isomorphism where we would like to state the Three Isomorphism Theorems, let us relate the idea of isomorphism to Graph Theory.

9.7.1 Isomorphism in Graph Theory

Let us state some definitions first.

Define $V(G)$ to be the vertex set of G, $v(G)$ and $e(G)$ to be the number of vertices and edges of G respectively. Lastly, let $d(v)$ be the degree of the vertex v, or namely, the number of edges adjacent to v.

For two graphs G and H, we say that $G \cong H$ if there is a bijection between the vertex sets of G and H. That is, there exists a mapping

$$f : V(G) \to V(H).$$

An isomorphism from G to H ensures that any two vertices u and v that are adjacent in G are also adjacent in H. We call this an *edge-preserving bijection.*

Consider the figure above. Note that $d(a) = 1$, $d(b) = 6$, $d(c) = 8$, $d(d) = 3$, $d(g) = 5$, $d(h) = 2$, $d(i) = 4$ and $d(j) = 7$. Hence, we say that $G \cong H$ even though they look pretty different!

9.7.2 First Isomorphism Theorem

Let $T : V \to W$ be a linear transformation. Then,

$$V/\ker(T) \cong R(T).$$

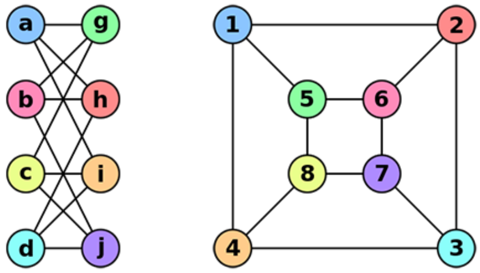

Figure 28: Left graph is G and right graph is H

This provides us an algebraic relation between the nullspace of $\mathbf{A}$ and the column space of $\mathbf{A}$, where $\mathbf{A}$ is an $m \times n$ matrix over $\mathbb{F}$. Suppose $L_{\mathbf{A}}(\mathbf{u}) = \mathbf{A}\mathbf{u}$, where $\mathbf{u} \in \mathbb{F}^n$. For each $\ker(L_{\mathbf{A}}) + \mathbf{v} \in \mathbb{F}^n / \ker(L_{\mathbf{A}})$, we have

$$\ker(L_{\mathbf{A}}) + \mathbf{v} = \{\mathbf{u} + \mathbf{v} | \mathbf{u} \in \ker(L_{\mathbf{A}})\}$$

gives the solution set of the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = \mathbf{A}\mathbf{v}$ is an element in the column space of $\mathbf{A}$.

9.7.3 Second Isomorphism Theorem

Let V and W be subspaces of a vector space U. Then,

$$(V + W)/W \cong V/(V \cap W).$$

9.7.4 Third Isomorphism Theorem

Let V be a subspace of a vector space U and let W be a subspace of V. Then,

$$(U/W)/(V/W) \cong U/V.$$

10 Multilinear Forms and Determinants

10.1 Permutations

Call a permutation σ of $S = \{1, 2, \ldots, n\}$ a bijective mapping from S to S. We represent σ by the following, and we refer to it as Cauchy's Two-Line Notation:

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}$$

The set of permutations of S is denoted by S_n and it is clear that $|S_n| = n!$.

Example: For example,

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

represents the mapping from $\{1, 2, 3\}$ to $\{1, 2, 3\}$ such that $\sigma(1) = 3$, $\sigma(2) = 1$ and $\sigma(3) = 2$.

This idea can be extended to compositions. For $\sigma, \tau \in S_n$, $\sigma \circ \tau$ is also a permutation and we can rewrite it as $\sigma\tau$.

10.1.1 Transposition

For $\alpha, \beta \in \{1, 2, \ldots, n\}$, we define $\phi_{\alpha,\beta}$ to be the permutation of $\{1, 2, \ldots, n\}$ such that

$$\phi_{\alpha,\beta}(k) = \begin{cases} k & \text{if } k \neq \alpha, \beta \\ \beta & \text{if } k = \alpha \\ \alpha & \text{if } k = \beta \end{cases}.$$

This permutation is known as the transposition of α and β. Note that ϕ is symmetric and self-inverse. That is,

$$\phi_{\alpha,\beta} = \phi_{\beta,\alpha} \text{ and } \phi_{\alpha,\beta}^{-1} = \phi_{\alpha,\beta}.$$

Example: For example, in S_4, we note that

$$\phi_{12} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}, \ \phi_{23} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} \text{ and } \phi_{34} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}.$$

Observe that

$$\phi_{12}\phi_{23}\phi_{12} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} = \phi_{13}.$$

REMARK

The decomposition of a permutation is not unique. For example,

$$\sigma = \phi_{13} = \phi_{14}\phi_{34}\phi_{14} = \phi_{12}\phi_{23}\phi_{12}.$$

REMARK

We note the following:
(1): $\{\sigma^{-1} | \sigma \in S_n\} = S_n$
(2): For any $\tau \in S_n$, $\{\tau\sigma | \sigma \in S_n\} = \{\sigma\tau | \sigma \in S_n\} = S_n$

Every permutation can be written as a product of transpositions. We can extend this result to say that for every $\sigma \in S_n$, there exists $\alpha_1, \alpha_2, \ldots, \alpha_k \in \{1, 2, \ldots, n-1\}$ such that

$$\sigma = \phi_{\alpha_1,\alpha_1+1}\phi_{\alpha_2,\alpha_2+1}\cdots\phi_{\alpha_k,\alpha_k+1}.$$

10.1.2 Inversion

Let $\sigma \in S_n$. An inversion is said to occur in σ if $\sigma(i) > \sigma(j)$ for $i < j$. If the total number of inversions in σ is even, σ is known as an even permutation. Otherwise, it is known as an odd permutation.

We define the sign, or parity, of σ by the signum function, $\operatorname{sgn}(\sigma)$, which is a piece-wise function given by

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}.$$

Moreover, if σ is a product of k transpositions, we have $\operatorname{sgn}(\sigma) = (-1)^k$. Thus, a permutation is even if it is the product of k transpositions and odd otherwise.

Example: Let

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} \text{ and } \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}.$$

In σ, inversions occur when $(i, j) = (1, 2), (1, 4)$ and $(3, 4)$. As such, σ is an odd permutation and $\operatorname{sgn}(\sigma) = -1$. In τ, inversions occur when $(i, j) = (1, 2), (1, 3), (1, 4)$ and $(3, 4)$. As such, τ is an even permutation and $\operatorname{sgn}(\tau) = 1$.

THEOREM

The signum function is a completely multiplicative. That is, for any $\sigma, \tau \in S_n$, $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$. Also, $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$.

10.2 Multilinear Forms

Let V be a vector space over $\mathbb{F}$ and let V^n be defined as the product of V n times. A mapping $T : V^n \to \mathbb{F}$ is a multilinear form on V if for each i, $1 \le i \le n$,

$$T(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, a\mathbf{v} + b\mathbf{w}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n) = aT(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{v}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n) + bT(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{w}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n)$$

for all $a, b \in \mathbb{F}$ and $\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_n, \mathbf{v}, \mathbf{w} \in V$.

If $n = 2$, we say that T is a bilinear form on V.

10.2.1 Bilinear Forms

If T is a bilinear form on V, where V is a vector space over $\mathbb{F}$, for $a \in \mathbb{F}$,

$$T(\mathbf{u} + \mathbf{v}, \mathbf{w}) = T(\mathbf{u}, \mathbf{w}) + T(\mathbf{v}, \mathbf{w}) \text{ and } T(a\mathbf{u}, \mathbf{v}) = aT(\mathbf{u}, \mathbf{v}).$$

Example: Let $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$. We define $T : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ by

$$T(\mathbf{u}, \mathbf{v}) = \mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{v}$$

for $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$. Then, T is a bilinear form on $\mathbb{F}^n$.

Proof: Suppose $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$. Consider $T(\mathbf{u} + \mathbf{v}, \mathbf{w}) = (\mathbf{u} + \mathbf{v})^{\mathrm{T}}\mathbf{A}\mathbf{w} = \mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{w} + \mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{w} = T(\mathbf{u}, \mathbf{w}) + T(\mathbf{v}, \mathbf{w})$. Next, let $a \in \mathbb{F}$. Then, $T(a\mathbf{u}, \mathbf{v}) = (a\mathbf{u})^{\mathrm{T}}\mathbf{A}\mathbf{v} = \mathbf{u}^{\mathrm{T}}a\mathbf{A}\mathbf{v} = a\mathbf{u}^{\mathrm{T}}\mathbf{A}\mathbf{v} = aT(\mathbf{u}, \mathbf{v})$. □

10.2.2 Alternating Multilinear Forms

A multilinear form T on V is alternative if $T(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = 0$ whenever $\mathbf{u}_\alpha \ne \mathbf{u}_\beta$ for some $\alpha \ne \beta$.

Every alternating multilinear map is anti-symmetric. That is, for all transpositions $\tau = \phi_{\alpha,\beta} \in S_n$ and $\mathbf{u}_1, \ldots, \mathbf{u}_n \in V$,

$$T(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = -T(\mathbf{u}_{\tau(1)}, \mathbf{u}_{\tau(2)}, \ldots, \mathbf{u}_{\tau(n)}).$$

Example: Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{F})$, then

$$T\left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix}\right) = ad - bc$$

for $\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \in \mathbb{F}^2$, which is an alternative bilinear form.

We shall define two functions, $P(\mathbf{A})$ and $D(\mathbf{A})$, which are called the permanent of $\mathbf{A}$ and determinant of $\mathbf{A}$ respectively, and for which we will formally introduce the latter in the next section on determinants too.

Define $P : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ by

$$P(\mathbf{A}) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{\sigma(i),i},$$

for $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. $P(\mathbf{A})$ is known as the permanent of $\mathbf{A}$. P is a multilinear form on $\mathbb{F}^n$.

Next, define $D : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ by

$$D(\mathbf{A}) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i}$$

for $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. D is also regarded as a multilinear form on $\mathbb{F}^n$. Moreover, it is alternative.

Let $T : V^n \to \mathbb{F}$ be an alternative multilinear form on a vector space V. Then, for all $\sigma \in S_n$ and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in V$,

$$T(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \operatorname{sgn}(\sigma) T(\mathbf{u}_{\sigma(1)}, \mathbf{u}_{\sigma(2)}, \ldots, \mathbf{u}_{\operatorname{sgn}(n)}).$$

In addition, if V is finite dimensional over a field, take a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ of V. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be defined as such:

$$\begin{aligned} \mathbf{u}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + \ldots + a_{m1}\mathbf{v}_m \\ \mathbf{u}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + \ldots + a_{m2}\mathbf{v}_m \\ \vdots &= \vdots \\ \mathbf{u}_n &= a_{1n}\mathbf{v}_1 + a_{2n}\mathbf{v}_2 + \ldots + a_{mn}\mathbf{v}_m \end{aligned}$$

where $a_{11}, a_{12}, \ldots, a_{mn} \in \mathbb{F}$. We state two properties as a consequence of this.

(1): Let $\mathcal{F}$ be the set of all mappings from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, m\}$. Then,

$$T(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \sum_{f \in \mathcal{F}} \prod_{i=1}^{n} a_{f(i),i} T(\mathbf{v}_{f(1)}, \mathbf{v}_{f(2)}, \ldots, \mathbf{v}_{f(n)}).$$

(2): If T is an alternative form, then $T(\mathbf{v}_{f(1)}, \mathbf{v}_{f(2)}, \ldots, \mathbf{v}_{f(n)}) = 0$ when f is not injective. That is, there exists $\alpha, \beta \in \{1, 2, \ldots, n\}$ such that $\alpha \neq \beta$ and $f(\alpha) = f(\beta)$. We have two cases, namely $m < n$ and $m \geq n$.

Case 1 ($m < n$)**:** Then, $T(\mathbf{v}_{f(1)}, \mathbf{v}_{f(2)}, \ldots, \mathbf{v}_{f(n)}) = 0$ for all $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in V$, i.e. T is a zero mapping

Case 2 ($m \geq n$)**:** Then, the equation

$$T(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \sum_{f \in \mathcal{F}} \prod_{i=1}^{n} a_{f(i),i} T(\mathbf{v}_{f(1)}, \mathbf{v}_{f(2)}, \ldots, \mathbf{v}_{f(n)})$$

still holds if we change $\mathcal{F}$ to the set of all injective mappings from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, m\}$. In particular, when $m = n$, $\mathcal{F}$ can be replaced by S_n.

10.3 Determinants

A mapping $D : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is called a determinant function on $\mathcal{M}_{n\times n}(\mathbb{F})$ if it satisfies the following axioms:
(D1): By regarding the columns of matrices in $\mathcal{M}_{n\times n}(\mathbb{F})$ as vectors in $\mathbb{F}^n$, D is a multilinear form on $\mathbb{F}^n$
(D2): $D(\mathbf{A}) = 0$ if $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$ has two identical columns
(D3): $D(\mathbf{I}_n) = 1$

10.3.1 Leibniz's Formula for Determinants

Leibniz's Formula

There exists one and only one determinant function on $\mathcal{M}_{n\times n}(\mathbb{F})$ and it is the function $\det : \mathcal{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ defined by

$$\det(\mathbf{A}) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i}$$

for $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. We call this Leibniz's Formula. The phrase *one and only one* hints to us that the idea of existence and uniqueness is involved.

Proof: We first proof the existence of the determinant function. (D1) and (D2) on det being an alternative multilinear form were mentioned earlier. Since $\mathbf{I}_n = (\delta_{ij})$, where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$, then for any $\sigma \in S_n$, $\prod_{i=1}^{n} \delta_{\sigma(i),i} = 0$ whenever σ is not the identity mapping. As such,

$$\det(\mathbf{I}_n) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \delta_{\sigma(i),i} = \prod_{i=1}^{n} \delta_{ii} = 1,$$

and so D satisfies (D3), and the proof of existence follows. □

To prove that D is unique, take any $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{n\times n}(\mathbb{F})$. By considering the standard basis vectors $\{\mathbf{e}_1\mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for $\mathbb{F}^n$,

$$D(\mathbf{A}) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{\sigma(i),i} D\left(\begin{pmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n\end{pmatrix}\right) = \det(\mathbf{A})D(\mathbf{I}_n) = \det(\mathbf{A})$$

and we are done. □

We shall now verify Leibniz's Formula for Determinants for 1×1, 2×2 and 3×3 matrices.

We start off with a 1×1 matrix $\mathbf{A} = (a_{11}) \in \mathcal{M}_{1\times 1}(\mathbb{F})$. In S_2, there is only one permutation $\sigma = \begin{pmatrix}1\\1\end{pmatrix}$. Then,

$$\det(\mathbf{A}) = \operatorname{sgn}(\sigma)a_{\sigma(1),1} = a_{11},$$

which is obviously true.

Next, consider a 2×2 matrix $\mathbf{A} = \begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{F})$. Then, as there are two permutations in S_2, namely

$$\sigma_1 = \begin{pmatrix}1 & 2\\ 1 & 2\end{pmatrix} \text{ and } \sigma_2 = \begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix},$$

then

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

For a 3×3 matrix $\mathbf{A} = \begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix} \in \mathcal{M}_{3\times 3}(\mathbb{F})$, note that there are six permutations in S_3 (simple to list them out), so

$$\begin{aligned}\det(\mathbf{A}) &= \sum_{i=1}^{6} \operatorname{sgn}(\sigma_i) a_{\sigma_i(1),1} a_{\sigma_i(2),2} a_{\sigma_i(3),3} \\ &= a_{11}a_{22}a_{33} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}\end{aligned}$$

10.3.2 Cofactor Expansion

Let $\mathbf{A} = (a_{ij} \in \mathcal{M}_{n\times n}(\mathbb{F})$. Define $\tilde{\mathbf{A}}_{ij}$ to be an $(n-1)\times(n-1)$ matrix obtained from $\mathbf{A}$ by deleting the i^{th} row and the j^{th} column. Then, for any $\alpha = 1, 2, \ldots, n$ and $\beta = 1, 2, \ldots, n$,

$$\begin{aligned}\det(\mathbf{A}) &= \sum_{i=1}^{n} a_{\alpha i} A_{\alpha i} \\ &= \sum_{i=1}^{n} a_{i\beta} A_{1\beta}\end{aligned}$$

where $A_{\alpha\beta} = (-1)^{\alpha+\beta} \det(\tilde{\mathbf{A}}_{\alpha\beta})$.

11 Diagonalisation and Jordan Canonical Forms

11.1 Eigenvalues and Diagonalisation

Recall that if V is a vector space and $T : V \to V$ is a linear operator, a non-zero vector $\mathbf{u} \in V$ is an eigenvector of T if $T(\mathbf{u}) \in \operatorname{span}\{\mathbf{u}\}$. That is, $T(\mathbf{u}) = \lambda\mathbf{u}$ for some $\lambda \in \mathbb{R}$. The scalar λ is an eigenvalue of T and $\mathbf{u}$ is an eigenvector associated with the eigenvector λ.

Let T be a linear operator on a finite dimensional vector space V, where $\dim(V) \geq 1$. The determinant of T, denoted by $\det(T)$, is defined to be the determinant of the matrix $[T]_B$, where B is an ordered basis for V. Also, if B and C are two ordered bases for V. Hence, $[T]_B$ and $[T]_C$ are similar, i.e. $[T]_B = \mathbf{P}^{-1}[T]_C\mathbf{P}$ for some invertible matrix $\mathbf{P}$. Moreover, it is clear that

$$\det([T]_B) = \det([T]_C).$$

This implies that the definition of $\det(T)$ is independent of the choice of B.

We use $c_T(x)$ to denote the characteristic polynomial of T. That is,

$$c_T(x) = \det(xI_V - T).$$

For an eigenvalue λ of T, we use $E_\lambda(T)$ to denote the eigenspace of T associated with λ. That is,

$$E_\lambda(T) = \ker(T - \lambda I_V).$$

Note that for any basis for V,

$$c_T(x) = c_{[T]_B}(x)$$

and T is diagonalisable if there exists an ordered basis B for V such that $[T]_B$ is a diagonal matrix.

11.2 Triangular Canonical Forms

Even though not all square matrices are diagonalisable, we can still reduce them into a simpler form, provided that the field is big enough. That is, a field that contains $\mathbb{F}$ and all polynomials over this field can be factorised into linear factors. Note that $\mathbb{C}$ contains $\mathbb{R}$ and all polynomials over $\mathbb{C}$ can be factorised into linear factors. In this section, the characteristic polynomial of a square matrix can be factorised into linear factors, then we say that the matrix is similar to an upper triangular matrix. In particular, every complex square matrix is similar to an upper triangular matrix.

For a field $\mathbb{F}$, let $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$. If the characteristic polynomial of $\mathbf{A}$, denoted by $c_{\mathbf{A}}(x)$, can be factorised into linear factors over $\mathbb{F}$, there exists an invertible matrix $\mathbf{P} \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is an upper triangular matrix.

In addition, let T be the linear operator on a finite dimensional vector space V over $\mathbb{F}$ such that $\dim(V) \geq 1$. If $c_T(x)$ can be factorised into linear factors over $\mathbb{F}$, there exists an ordered basis B for V such that $[T]_B$ is an upper triangular matrix. This method will also work if we want to make $[T]_B$ a lower triangular matrix, for which we *create* the new $(n-1)\times(n-1)$ matrix starting from the (a_{nn})-entry.

Example: Let $\mathbf{A} = \begin{pmatrix} 8 & -6 & -2 & -2 \\ 4 & 3 & 1 & 1 \\ 8 & 6 & 10 & 2 \\ -4 & 1 & 3 & 11 \end{pmatrix}$ be a real matrix. We attempt to find an invertible matrix $\mathbf{P} \in \mathcal{M}_{4\times 4}(\mathbb{F})$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is an upper triangular matrix.

Solution: Note that 8 is an eigenvalue of $\mathbf{A}$, with a multiplicity of 4. That is, $c_{\mathbf{A}}(x) = (x-8)^4$. The corresponding eigenvector is $\begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$. We extend this vector to a basis of $\mathbb{R}^4$, for which the basis is

$$\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Write $\mathbf{S} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$. Note that

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \begin{pmatrix} 8 & 8 & 6 & 2 \\ 0 & 8 & -6 & -2 \\ 0 & 4 & 3 & 1 \\ 0 & 4 & 7 & 13 \end{pmatrix}.$$

Consider the bottom right 3×3 matrix. Call it $\mathbf{B}$, where $\mathbf{B} = \begin{pmatrix} 8 & -6 & -2 \\ 4 & 3 & 1 \\ 4 & 7 & 13 \end{pmatrix}$. It has an eigenvector $\begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$.

Similarly, we extend $\begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$ to a basis of $\mathbb{R}^3$, that is

$$\left\{ \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Let

$$\mathbf{R} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}.$$

Then,

$$\mathbf{R}^{-1}\mathbf{B}\mathbf{R} = \begin{pmatrix} 8 & -3 & -1 \\ 0 & 6 & 2 \\ 0 & -2 & 10 \end{pmatrix}.$$

The matrix $\mathbf{C} = \begin{pmatrix} 6 & 2 \\ -2 & 10 \end{pmatrix}$ has an eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Extend this to a basis of $\mathbb{R}^2$, that is,

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Let $\mathbf{Q} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Then,

$$\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q} = \begin{pmatrix} 8 & 2 \\ 0 & 8 \end{pmatrix}.$$

Then, let

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & -3 & 1 & 1 \end{pmatrix}.$$

Hence,

$$\mathbf{P}^{-1}\mathbf{AP} = \begin{pmatrix} 8 & 16 & 8 & 2 \\ 0 & 8 & -4 & -1 \\ 0 & 0 & 8 & 2 \\ 0 & 0 & 0 & 8 \end{pmatrix},$$

which is an upper triangular matrix. □

11.3 Invariant Subspaces

Let V be a vector space and $T : V \to V$ be a linear operator. A subspace W of V is T-invariant if $T(\mathbf{u})$ is contained in W for all $\mathbf{u} \in W$. That is,

$$T[W] = \{T(\mathbf{u})|\mathbf{u} \in W\} \subseteq W.$$

Moreover, the linear operator $T|_W : W \to W$ defined by

$$T|_W(\mathbf{u}) = T(\mathbf{u})$$

for $\mathbf{u} \in W$ is known as the restriction of T on W.

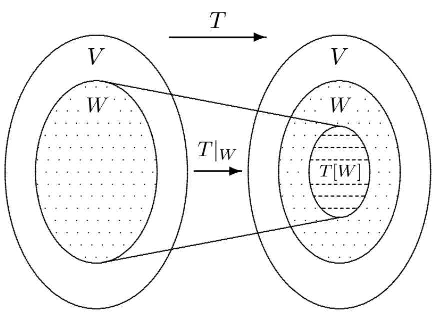

Figure 29: W is a T-invariant subspace of V

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear operator defined by $T((x, y, z)) = (y, -x, z)$ for $(x, y, z) \in \mathbb{R}^3$. Note that T is a rotation about the z-axis. Let W_1 be the xy-plane in $\mathbb{R}^3$. That is, $W_1 = \{(x, y, 0)|x, y \in \mathbb{R}\}$. Since

$$T((x, y, 0)) = (y, -x, 0) \in W_1$$

for all $(x, y, 0) \in W_1$, then W_1 is T-invariant. $T|_W(\mathbf{u})$ is simply a rotation defined on the xy-plane.

However, if W_2 is the yz-plane in $\mathbb{R}^3$, that is $W_2 = \{(0, y, z)|y, z \in \mathbb{R}\}$, then W_2 is not T-invariant since

$$(0, 1, 1) \in W_2 \text{ but } T((0, 1, 1)) = (1, 0, 1) \notin W_2.$$

Let S and T be linear operators on V, and W is a subspace of V which is both S-invariant and T-invariant. Then,

(1): W is $(S \circ T)$-invariant and $(S \circ T)|_W = S|_W \circ T|_W$
(2): W is $(S + T)$-invariant and $(S + T)|_W = S|_W + T|_W$
(3): for any scalar c, W is (cT)-invariant and $(cT)|W = c(T|_W)$
(4): Let T be a linear operator on a finite dimensional vector space V. Suppose W is a T-invariant subspace of V with $\dim(W) \geq 1$. Then, $c_T(x)$ is divisible by $c_{T|_W}(x)$.

Suppose T is a linear operator on a finite dimensional vector space V over a field $\mathbb{F}$ and W is a T-invariant subspace of V with $\dim(W) \geq 1$. Take an ordered basis $C = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ for W. For each $1 \leq j \leq m$, since $T(\mathbf{v}_j) \in W$,

$$T|_W(\mathbf{v}_j) = T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{v}_i$$

for some $a_{ij} \in \mathbb{F}$, where $1 \le i \le m$. Note that

$$[T|_W]_C = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{pmatrix}.$$

We can extend C to an ordered basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m, \mathbf{v}_{m+1}, \dots, \mathbf{v}_n\}$ for V. For each $m+1 \le j \le n$, we have

$$T(\mathbf{v}_j) = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i$$

for some $a_{ij} \in \mathbb{F}$, where $1 \le i \le n$. Then, it is clear that

$$[T]_B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} & a_{1,m+1} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2m} & a_{2,m+1} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mm} & a_{m,m+1} & \dots & a_{mn} \\ 0 & 0 & \dots & 0 & a_{m+1,m+1} & \dots & a_{m+1,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & a_{n,m+1} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{0} & \mathbf{A}_3 \end{pmatrix},$$

where $\mathbf{A}_1 = [T|_W]_C$, $\mathbf{A}_2$ is an $m \times (n-m)$ matrix and $\mathbf{A}_3$ is an $(n-m) \times (n-m)$ matrix.

Example: Consider the linear operator T on $\mathbb{R}^4$, where

$$T((a,b,c,d) = (4a+2b+2c+2d, a+2b-c, -a-b, -3a-2b-2c-d)$$

for $(a,b,c,d) \in \mathbb{R}^4$. Let $W = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1 = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 0 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 0 \\ -3 \\ 1 \\ 1 \end{pmatrix}$. It is clear that these two vectors is a basis for W. Note that

$$T(\mathbf{v}_1) = \mathbf{v}_2 \text{ and } T(\mathbf{v}_2) = -2\mathbf{v}_1 + 3\mathbf{v}_2.$$

Hence,

$$[T|_W]_C = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}.$$

We extend this to a basis for $\mathbb{R}^4$, where $\mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$ and $\mathbf{v}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$. Observe that

$$T(\mathbf{v}_3) = 2\mathbf{v}_1 - \frac{5}{3}\mathbf{v}_2 + \frac{5}{3}\mathbf{v}_3 - \frac{1}{3}\mathbf{v}_4 \text{ and } T(\mathbf{v}_4) = 2\mathbf{v}_1 - \frac{4}{3}\mathbf{v}_2 + \frac{4}{3}\mathbf{v}_3 + \frac{1}{3}\mathbf{v}_4.$$

As such, we can set up the following block matrix, $[T]_B$:

$$[T]_B = \begin{pmatrix} 0 & -2 & 2 & 2 \\ 1 & 3 & -\frac{5}{3} & -\frac{4}{3} \\ 0 & 0 & \frac{5}{3} & \frac{4}{3} \\ 0 & 0 & -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Setting $\mathbf{A} = \begin{pmatrix} 4 & 2 & 2 & 2 \\ 1 & 2 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -3 & -2 & -2 & -1 \end{pmatrix}$ and $\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & -3 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$, we have

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 0 & -2 & 2 & 2 \\ 1 & 3 & -\frac{5}{3} & -\frac{4}{3} \\ 0 & 0 & \frac{5}{3} & \frac{4}{3} \\ 0 & 0 & -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

and we are done. □

11.3.1 Cyclic Subspace

Let T be a linear operator on a vector space V over a field $\mathbb{F}$. For any $\mathbf{u} \in V$, define

$$W = \operatorname{span}\left\{\mathbf{u}, T(\mathbf{u}), T^2(\mathbf{u}), \ldots\right\}.$$

For any $\mathbf{w} \in W$, $\mathbf{w} = a_0\mathbf{u} + a_1T(\mathbf{u}) + \ldots + a_mT^m(\mathbf{u})$ for some $m \in \mathbb{N}$ and $a_0, a_1, \ldots, a_m \in \mathbb{F}$. Then,

$$T(\mathbf{w}) = a_0T(\mathbf{u}) + a_1T^2(\mathbf{u}) + \ldots + a_mT^{m+1}(\mathbf{u}) \in W,$$

implying that W is T-invariant. We call W the T-cyclic subspace of V generated by $\mathbf{u}$.

Suppose $\dim(V) \geq 1$. Take a non-zero vector $\mathbf{u} \in V$. Suppose the T-cyclic subspace $W = \operatorname{span}\left\{\mathbf{u}, T(\mathbf{u}), T^2(\mathbf{u}), \ldots\right\}$ generated by $\mathbf{u}$ is finite dimensional. Then, $\dim(W)$ is the smallest $k \in \mathbb{N}$ such that $T^k(\mathbf{u})$ is a linear combination of $\mathbf{u}, T(\mathbf{u}), \ldots, T^{k-1}(\mathbf{u})$.

If $\dim(W) = k$, then
(1): $\left\{\mathbf{u}, T(\mathbf{u}), \ldots, T^{k-1}(\mathbf{u})\right\}$ is a basis for W
(2): If $T^k(\mathbf{u}) = a_0\mathbf{u} + a_1T(\mathbf{u}) + \ldots + a_{k-1}T^{k-1}(\mathbf{u})$, where $a_0, a_1, \ldots, a_{k-1} \in \mathbb{F}$, then

$$c_{T|_W}(x) = -a_0 - a_1x - \ldots - a_{k-1}x^{k-1} + x^k.$$

11.4 The Cayley-Hamilton Theorem

Let $\mathbb{F}$ be a field and $p(x)$ be a polynomial of degree m. That is, $p(x) = a_0 + a_1x + \ldots + a_mx^m$, where $a_0, a_1, \ldots, a_m \in \mathbb{F}$ and $a_m \neq 0$. For a linear operator T on a vector space V over $\mathbb{F}$, we use $p(T)$ to denote the linear operator $a_0I_V + a_1T + \ldots + a_mT^m$ on V. For an $n \times n$ matrix $\mathbf{A}$ over $\mathbb{F}$, we use $p(\mathbf{A})$ to denote the square matrix $a_0\mathbf{I}_n + a_1\mathbf{A} + \ldots + a_m\mathbf{A}^m$.

(1): Suppose V is finite dimensional, where $\dim(V) = n \geq 1$. For any ordered basis B for V, the matrix for $p(T)$ relative to B is $[p(T)]_B = p([T]_B)$.
(2): If W is a T-invariant subspace of V, then W is also a $p(T)$-invariant subspace of V, and $p(T)|_W = p(T|_W)$
(3): Suppose $q(x)$ and $r(x)$ are polynomials over $\mathbb{F}$ such that $p(x) = q(x)r(x)$. Then,

$$p(T) = q(T) \circ r(T) = r(T) \circ q(T) \text{ and } p(\mathbf{A}) = q(\mathbf{A})r(\mathbf{A}) = r(\mathbf{A})q(\mathbf{A}).$$

We are now ready to state the Cayley-Hamilton Theorem.

Cayley-Hamilton Theorem

Let T be a linear operator on a finite dimensional vector space V, where $\dim(V) \geq 1$. Then, $c_T(T) = O_V$, where O_V is the zero operator on V. Let $\mathbf{A}$ be a square matrix. Then, $c_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$.

Example: Consider the matrix $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}$. We can show that $\mathbf{A}^2 - 2\mathbf{A} - 5\mathbf{I} = \mathbf{0}$.

Solution: The characteristic polynomial is $c_{\mathbf{A}}(x) = (1-x)^2 - 6 = -x^2 - 2x + 5$. Thus, $c_{\mathbf{A}}(\mathbf{A}) = -\mathbf{A}^2 - 2\mathbf{A} + 5\mathbf{I}$. The result follows by the Cayley-Hamilton Theorem. □

11.5 Minimal Polynomials

Let $p(x)$ be a polynomial of degree m over $\mathbb{F}$. That is, $p(x) = a_0 + a_1x + \ldots + a_mx^m$, where $a_0, a_1, \ldots, a_m \in \mathbb{F}$ and $a_m \neq 0$. If $a_m = 1$, $p(x)$ is monic. Let T be a linear operator on a finite dimensional vector space V over $\mathbb{F}$, where $\dim(V) \geq 1$. The minimal polynomial of T is the polynomial $m_T(x)$ over $\mathbb{F}$ such that
(1): $m_T(x)$ is monic
(2): $m_T(T) = O_V$
(3): if $p(x)$ is a non-zero polynomial over $\mathbb{F}$ such that $p(T) = O_V$, then $\deg(p(x)) \geq \deg(m_T(x))$

That is, the minimal polynomial $m_T(T)$ of T is the monic polynomial $p(x)$ of least degree for which $p(T) = O_V$. In other words, for a square matrix $\mathbf{A}$, the minimal polynomial $m_{\mathbf{A}}(x)$ of $\mathbf{A}$ is the monic polynomial $p(x)$ of least degree for which $p(\mathbf{A}) = \mathbf{0}$.

The bad news is that it is generally not easy to find the minimal polynomial of a linear operator. We can only check the results by trial and error. However, we will provide a method to guess what the minimal polynomial looks like.

Guesing $m_{\mathbf{A}}(x)$

Let T be a linear operator on a finite dimensional vector space V, where $\dim(V) \geq 1$.
(1): Let $p(x)$ be a polynomial over $\mathbb{F}$. Then, $p(T) = O_V$ if and only if $p(x)$ is divisible by $m_T(x)$
(2): If W is a T-invariant subspace of V with $\dim(W) \geq 1$, then $m_T(x)$ is divisible by $m_{T|_W}(x)$
(3): Suppose λ is an eigenvalue of T such that $(x-\lambda)^r$ strictly divides $c_T(x)$. That is, $c_T(x) = (x-\lambda)^r q(x)$, where $q(x)$ is a polynomial over $\mathbb{F}$ which is not divisible by $x - \lambda$. Then,

$$m_T(x) = (x-\lambda)^s q_1(x),$$

where $1 \leq s \leq r$, $q_1(x)$ is a polynomial over $\mathbb{F}$ and $q_1(x)$ divides $q(x)$.

Example: Suppose $\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ and we wish to find $m_{\mathbf{A}}(x)$.

Solution: Note that $c_{\mathbf{A}}(x) = (x-2)^3$. We must have $m_{\mathbf{A}}(x) = (x-2)^s$, where $s = 1, 2$ or 3 and $(\mathbf{A} - 2\mathbf{I})^s = \mathbf{0}$. It is easy to verify that $s_{\min} = 2$, so $m_{\mathbf{A}}(x) = (x-2)^2$. □

Let T be a linear operator on a vector space V. Suppose W_1 and W_2 are T-invariant subspaces of V. Then, $W_1 + W_2$ is also T-invariant and if W_1 and W_2 are finite dimensional with $\dim(W_1), \dim(W_2) \geq 1$, $m_{T|_{W_1+W_2}}(x)$ is equal to the lcm of $m_{T|_{W_1}}(x)$ and $m_{T|_{W_2}}(x)$.

Let T be a linear operator on a finite dimensional vector space V such that

$$c_T(x) = \prod_{i=1}^{k} (x-\lambda_i)^{r_i},$$

where λ_i are distinct eigenvalues of T, where $1 \leq i \leq k$. Note that $\sum_{i=1}^{k} r_i = \dim(V)$. We have the following equivalent statements:
(1): T is diagonalisable
(2): $m_T(x) = \prod_{i=1}^{k} (x-\lambda_i)$
(3): $\dim(E_{\lambda_i}(T)) = r_i$ for $1 \leq i \leq k$
(4): $V = \bigoplus_{i=1}^{k} E_{\lambda_i}(T)$

11.6 Jordan Canonical Forms

Let λ be a scalar. The $t \times t$ matrix

$$\mathbf{J}_t(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$$

is a Jordan Block of order t associated with λ. For example,

$$\mathbf{J}_1(\lambda) = \begin{pmatrix} \lambda \end{pmatrix}, \ \mathbf{J}_2(\lambda) = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \text{ and } \mathbf{J}_3(\lambda) = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}.$$

In relation to the minimal polynomial, given a Jordan Block $\mathbf{J}_t(\lambda)$, we have

$$c_{\mathbf{J}}(x) = m_{\mathbf{J}}(x) = (x-\lambda)^t.$$

Let T be the linear operator on a finite dimensional vector space V over a field $\mathbb{F}$, where $\dim(v) \geq 1$. Suppose $c_T(x)$ can be factorised into linear factors over $\mathbb{F}$. Then, there exists an ordered basis B for V such that $[T]_B = \mathbf{J}$, where

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{t_1}(\lambda_1) & & & & \mathbf{0} \\ & \mathbf{J}_{t_2}(\lambda_2) & & & \\ & & \ddots & & \\ & & & \ddots & \\ \mathbf{0} & & & & \mathbf{J}_{t_m}(\lambda_m) \end{pmatrix},$$

and the λ_i's are the eigenvalues of T, where $1 \leq i \leq m$. We call $\mathbf{J}$ the Jordan Canonical Form for T.

Hence, if $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{F})$ and $c_{\mathbf{A}}(x)$ can be factorised into linear factors over $\mathbb{F}$, we can find an invertible matrix $\mathbf{P} \in \mathcal{M}_{n\times n}(\mathbb{F})$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J}$. We say that $\mathbf{J}$ is also the Jordan Canonical Form for $\mathbf{A}$.

Note that Jordan Canonical Forms are not unique. However, two Jordan Canonical Forms for a linear operator have the same collection of Jordan Blocks in different orders.

We have three properties given that the Jordan Matrix is that of the one above. They are namely
(1):

$$c_T(x) = \prod_{i=1}^{m} (x-\lambda_i)^{t_i}$$

(2): $m_T(x)$ is the least common multiple of all the $(x-\lambda_i)^{t_i}$'s, where $1 \leq i \leq m$
(3): For every eigenvalue λ of T, $\dim(E_\lambda(T))$ is the total number of Jordan Blocks associated with λ in the matrix $\mathbf{J}$

Example: Consider the matrix

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

We wish to obtain the Jordan Canonical Form, $\mathbf{J}$, of $\mathbf{A}$.

Solution: Note that $c_{\mathbf{A}}(\lambda) = (\lambda-1)(\lambda-2)(\lambda-4)^2$, so the eigenvalues are 1,2 and 4. The eigenvectors associated with $\lambda = 1, 2, 4$ are

$$\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \ \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

respectively. Hence,

$$\mathbf{J} = \mathbf{J}_1(1) \oplus \mathbf{J}_1(2) \oplus \mathbf{J}_2(4) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

Consider $(\mathbf{A} - 4\mathbf{I})^s$, where $s = 1$ or 2. Observe that

$$\ker((\mathbf{A} - 4\mathbf{I})^2) = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \end{pmatrix} \right\}.$$

Thus, we set

$$\mathbf{P} = \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

and it is easy to verify that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{J}$. $\square$

Example: Let $\mathbf{A}$ be a real square matrix such that

$$c_{\mathbf{A}}(x) = (x-1)^3(x-2)^2 \text{ and } m_{\mathbf{A}}(x) = (x-1)^2(x-2).$$

We wish to find a Jordan Canonical Form for $\mathbf{A}$.

Solution: Let $\mathbf{J}$ be the required matrix. Using $c_{\mathbf{A}}(x)$, along the diagonal of $\mathbf{J}$, there are three entries with 1 and two entries with 2. Using $m_{\mathbf{A}}(x)$, the largest Jordan Block associated with 1 has order 2 and the largest Jordan Block associated with 2 has order 1. Hence, $\mathbf{J}$ is similar to

$$\begin{pmatrix} \mathbf{J}_2(1) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_1(1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_1(2) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}_1(2) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

$\square$

12 Inner Product Spaces

Recall in MA2001, we used the dot product to define lengths, distances and angles in $\mathbb{R}^n$. However, there is a *flaw* in the setup. How do we define the angle between two vectors in $\mathbb{R}^n$? We require an abstract version of the dot product, which is why we now draw our attention to inner product spaces. Note that we only study vector spaces over $\mathbb{R}$ and $\mathbb{C}$ in this chapter.

For a complex number $c = a + bi$, where $a, b \in \mathbb{R}$, we use $\overline{c}$ to denote the complex conjugate of c. That is, $\overline{c} = a - bi$. Hence, it is clear that if c is real, then $c = \overline{c}$.

Let $\mathbf{A} = (a_{ij})_{m\times n}$ be a complex matrix. We use $\overline{\mathbf{A}}$ to denote the conjugate of $\mathbf{A}$. That is, $\overline{\mathbf{A}} = (\overline{a}_{ij})_{m\times n}$. Next, we define the conjugate transpose of $\mathbf{A}$ by $\mathbf{A}^*$, where

$$\mathbf{A}^* = \overline{A}^{\mathrm{T}} = (\overline{a}_{ji})_{n\times m}.$$

Hence, if $\mathbf{A}$ is a real matrix, then $\overline{\mathbf{A}} = \mathbf{A}$ and $\mathbf{A}^* = \mathbf{A}^{\mathrm{T}}$. The following three properties are easy to derive:
(1): $(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*$
(2): $(\mathbf{AC})^* = \mathbf{C}^*\mathbf{A}^*$
(3): $(c\mathbf{A})^* = \overline{c}\mathbf{A}^*$

12.1 Inner Products

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and let V be a vector space over $\mathbb{F}$. An inner product on V is a mapping which assigns to each ordered pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a scalar $\langle \mathbf{u}, \mathbf{v}\rangle \in \mathbb{F}$ such that it satisfies the following four axioms:
(IP1) Conjugate symmetry: For all $\mathbf{u}, \mathbf{v} \in V$, $\langle \mathbf{u}, \mathbf{v}\rangle = \overline{\langle \mathbf{v}, \mathbf{u}\rangle}$
(IP2) Linearity: For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$
(IP3) Linearity: For all $c \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in V$, $\langle c\mathbf{u}, \mathbf{v}\rangle = c\,\langle \mathbf{u}, \mathbf{v}\rangle$
(IP4) Positive definiteness: $\langle \mathbf{0}, \mathbf{0}\rangle = 0$ and for all non-zero $\mathbf{u} \in V$, $\langle \mathbf{u}, \mathbf{u}\rangle > 0$

A vector space V over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ equipped with an inner product is an inner product space. If $\mathbb{F} = \mathbb{R}$, then V is a real inner product space. If $\mathbb{F} = \mathbb{C}$, then V is a complex inner product space.

Recall that in MA2001, we introduced the idea of transpose being similar to the dot product. Here in inner product spaces, we have for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we have $\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}\mathbf{v}^{\mathrm{T}}$.

Example: For $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$,

$$\langle \mathbf{u}, \mathbf{v}\rangle = u_1^2 + u_2^2 + v_1^2 + v_2^2$$

is not an inner product on $\mathbb{R}^2$.

Solution: We verify that (IP2) and (IP3) are not satisfied. Suppose $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^2$. Then,

$$\begin{aligned}\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle &= (u_1+v_1)^2 + (u_2+v_2)^2 + w_1^2 + w_2^2\\ &= u_1^2 + u_2^2 + v_1^2 + v_2^2 + 2w_1^2 + 2w_2^2 - w_1^2 - w_2^2 + 2u_1v_1 + 2u_2v_2\\ &= \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle - w_1^2 - w_2^2 + 2u_1v_1 + 2u_2v_2\end{aligned}$$

In general, $w_1^2 + w_2^2 \neq 2u_1v_1 + 2u_2v_2$, so (IP2) is not satisfied.

Next, we verify that (IP3) is not satisfied. Let $c \in \mathbb{F}$. Then,

$$\begin{aligned}\langle c\mathbf{u}, \mathbf{v}\rangle &= (cu_1)^2 + (cu_2)^2 + v_1^2 + v_2^2\\ &= c^2u_1^2 + c^2u_2^2 + v_1^2 + v_2^2\end{aligned}$$

This, in general, is not equal to $c\,\langle \mathbf{u}, \mathbf{v}\rangle = cu_1^2 + cu_2^2 + cv_1^2 + cv_2^2$. $\square$

Example: For $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$,

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + 2u_2 v_2$$

is an inner product on $\mathbb{R}^2$.

Solution: Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ and $c \in \mathbb{F}$.

We verify that (IP1) is satisfied.

$$\overline{\langle \mathbf{v}, \mathbf{u} \rangle} = \overline{v_1 u_1 + 2v_2 u_2} = \overline{v_1 u_1} + 2\overline{v_2 u_2} = u_1 v_1 + 2u_2 v_2 = \langle \mathbf{u}, \mathbf{v} \rangle$$

For (IP2),

$$\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = (u_1 + v_1)w_1 + 2(u_2 + v_2)w_2 = u_1 w_1 + 2u_2 w_2 + v_1 w_1 + 2v_2 w_2 = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle.$$

For (IP3),

$$\langle c\mathbf{u}, \mathbf{v} \rangle = (cu_1)v_1 + 2(cu_2)v_2 = c(u_1 v_1 + 2u_2 v_2) = c \langle \mathbf{u}, \mathbf{v} \rangle.$$

For (IP4), it is clear that $\langle \mathbf{0}, \mathbf{0} \rangle = 0$ and for all non-zero $\mathbf{u} \in \mathbb{R}^2$,

$$\langle \mathbf{u}, \mathbf{u} \rangle = u_1^2 + 2u_2^2 > 0.$$

Since all the four axioms are satisfied, $\langle \mathbf{u}, \mathbf{v} \rangle$ is an inner product on $\mathbb{R}^2$. □

12.2 Norms and Distances

Let v be an inner product space.
(1): For $\mathbf{u} \in V$, the norm, or length, of $\mathbf{u}$ is defined to be $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$. Vectors of norm 1 are unit vectors.
(2): For $\mathbf{u}, \mathbf{v} \in V$, the distance between $\mathbf{u}$ and $\mathbf{v}$ is $d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$

As such, for n-tuple vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n), \mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$, we have

$$\|\mathbf{u}\| = \sqrt{\sum_{i=1}^{n} u_i^2}$$

and

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}.$$

For an inner product space V over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, we have the following properties:
(1): $\|\mathbf{0}\| = 0$ and for any non-zero $\mathbf{u} \in V$, $\|\mathbf{u}\| > 0$
(2): For any $c \in \mathbb{F}$ and $\mathbf{u} \in V$, $\|c\mathbf{u}\| = |c| \|\mathbf{u}\|$

We also have the Cauchy-Schwarz and Triangle Inequalities which we will mention in a moment.

12.2.1 Cauchy-Schwarz Inequality

As an inner product, the Cauchy-Schwarz Inequality states that for any $\mathbf{u}, \mathbf{v} \in V$, $| \langle \mathbf{u}, \mathbf{v} \rangle | \le \|\mathbf{u}\| \|\mathbf{v}\|$. Equality is attained if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent. That is, $\mathbf{u} = \alpha\mathbf{v}$ or $\mathbf{v} = \beta\mathbf{u}$ for some $\alpha, \beta \in \mathbb{F}$.

We also have the Cauchy-Schwarz Inequality for real numbers and continuous functions.

For any real numbers $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n$, we have

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) \ge \left(\sum_{i=1}^{n} x_i y_i\right)^2,$$

and equality holds if and only if $x_i = y_i$ for $1 \leq i \leq n$.

As for continuous functions, for any f and $g \in C([a,b])$, where $[a,b]$, with $a < b$, is a closed interval on the real line, we have

$$\left(\int_a^b (f(t))^2\ dt\right)\left(\int_a^b (g(t))^2\ dt\right) \geq \left(\int_a^b f(t)g(t)\ dt\right)^2.$$

12.2.2 Triangle Inequality

The Triangle States that for any $\mathbf{u}, \mathbf{v} \in V$, $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

12.3 Orthogonal and Orthonormal Bases

We state some definitions.

(1): Two vectors $\mathbf{u}, \mathbf{v} \in V$ are orthogonal to each other $\langle \mathbf{u}, \mathbf{v} \rangle = 0$
(2): Let W be a subspace of V. A vector $\mathbf{u}$ is orthogonal to W if $\mathbf{u}$ is orthogonal to all vectors in W
(3): A subset B of V is orthogonal if the vectors in B are pairwise orthogonal. If B is an orthogonal set and it is a basis for V, then B is an orthogonal basis for V.
(4): A subset B of V is orthonormal if B is orthogonal and all vectors in B are unit vectors. Hence, if B is an orthonormal set and it is a basis for V, then B is an orthonormal basis for V.

Suppose V is a finite dimensional inner product space. To determine whether a set B of non-zero vectors from V is an orthogonal basis for V, it suffices to prove that B is orthogonal and $|B| = \dim(V)$.

Moreover, if $B = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ is an orthonormal basis for V, then for any vector $\mathbf{u} \in V$,

$$\mathbf{u} = \sum_{i=1}^{n} \langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i.$$

12.4 Orthogonal Complements and Orthogonal Projections

Let V be an inner product space and W a subspace of V. The orthogonal complement of W is defined to be the set

$$\begin{aligned} W^\perp &= \{\mathbf{v} \in V | \mathbf{v} \text{ is orthogonal to } W\} \\ &= \{\mathbf{v} \in V | \langle \mathbf{v}, \mathbf{u} \rangle = 0 \text{ for all } \mathbf{u} \in W\} \subseteq V \end{aligned}$$

Let V be an inner product space and W a subspace of V. Then,
(1): $W^\perp$ is a subspace of V
(2): $W \cap W^\perp = \{\mathbf{0}\}$, i.e. $W + W^\perp$ is a direct sum
(3): if W is finite dimensional, $V = W \oplus W^\perp$
(4): if W is finite dimensional, $\dim(V) = \dim(W) + \dim(W^\perp)$

Proof: To prove the first property, it is clear that $\mathbf{0} \in W^\perp$, so $W^\perp$ is non-empty. Then, take any $\mathbf{v}, \mathbf{w} \in W^\perp$. For all $\mathbf{u} \in W$ and any scalar c,

$$\langle c\mathbf{v} + \mathbf{w}, \mathbf{u} \rangle = \langle c\mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle = c \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{u} \rangle$$

and by the definition of orthogonal complement, the above is simply 0. □

Proof: For the second property, let $\mathbf{v} \in W \cap W^\perp$. Then, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ and so $\mathbf{v} = 0$. The result follows. □

Proof: For the third property, the result is clear if W is the zero space. If W is not a zero space, let $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ be an orthonormal basis for W. For any $\mathbf{u} \in V$, we write it as

$$\mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{u}, \mathbf{w}_i \rangle \mathbf{w}_i$$

and $\mathbf{v}' = \mathbf{u} - \mathbf{v}$, where $\mathbf{v} \in W$. For $1 \le i \le k$,

$$\begin{aligned}\langle \mathbf{v}', \mathbf{w}_i \rangle &= \left\langle \mathbf{u} - \sum_{j=1}^{k} \langle \mathbf{u}, \mathbf{w}_j \rangle \mathbf{w}_j, \mathbf{w}_i \right\rangle \\ &= \langle \mathbf{u}, \mathbf{w}_i \rangle - \sum_{j=1}^{k} \langle \mathbf{u}, \mathbf{w}_j \rangle \langle \mathbf{w}_j, \mathbf{w}_i \rangle \\ &= \langle \mathbf{u}, \mathbf{w}_i \rangle - \langle \mathbf{u}, \mathbf{w}_i \rangle = 0\end{aligned}$$

As $\mathbf{v}' \in W^\perp$, then $V = W + W^\perp$ and the result follows. □

The fourth property will not be proven since it is obvious.

Let V be an inner product space and W a subspace of V. Then,
(1): $W \subseteq (W^\perp)^\perp$
(2): if W is finite dimensional, $W = (W^\perp)^\perp$

If W is infinite dimensional, the second property may not always hold true. For example, $W^\perp = \{\mathbf{0}\}$, $(W^\perp)^\perp = V$. It satisfies the first property, but not the second.

12.5 Adjoints of Linear Operators

Let V be an inner product space and T be a linear operator on V. A linear operator T^* on V is the adjoint of T if

$$\langle T(\mathbf{u}), \mathbf{v} \rangle = \langle \mathbf{u}, T^*(\mathbf{v}) \rangle .$$

If the adjoint of T exists, it must be unique. The adjoint satisfies the property that

$$\langle \mathbf{u}, T(\mathbf{v}) \rangle = \overline{\langle T(\mathbf{v}), \mathbf{u} \rangle} = \overline{\langle \mathbf{v}, T^*(\mathbf{u}) \rangle} = \langle T^*(\mathbf{u}), \mathbf{v} \rangle.$$

Note that I_V and O_V are the adjoints of themselves. That is, $I_V^* = I_V$ and $O_V^* = O_V$.

If V is a finite dimensional inner product space, then
(1): T^* always exists
(2): if B is an ordered orthonormal basis for V, $[T^*]_B = ([T]_B)^*$
(3): $\operatorname{rank}(T) = \operatorname{rank}(T^*)$ and $\operatorname{nullity}(T) = \operatorname{nullity}(T^*)$

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and let V be an inner product space over $\mathbb{F}$. Suppose S and T are linear operators on V such that S^* and T^* exist. Then,
(1): $(S + T)^*$ exists and $(S + T)^* = S^* + T^*$
(2): for any $c \in \mathbb{F}$, $(cT)^*$ exists and $(cT)^* = \overline{c}T^*$
(3): $(S \circ T)^*$ exists and $(S \circ T)^* = T^* \circ S^*$
(4): $(T^*)^*$ exists and $(T^*)^* = T$
(5): if W is a subspace of V which is T-invariant and T^*-invariant, then $(T|_W)^*$ exists and $(T|_W)^* = T^*|_W$

12.5.1 Unitary and Orthogonal Operators

Let T be invertible and $T^{-1} = T^*$. If $\mathbb{F} = \mathbb{C}$, T is a unitary operator. If $\mathbb{F} = \mathbb{R}$, T is an orthogonal operator.

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, V a finite dimensional inner product space over $\mathbb{F}$ and T a linear operator on V. Then, the following are equivalent:
(1): T is unitary or orthogonal
(2): For all $\mathbf{u}, \mathbf{v} \in V$, $\langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$
(3): For all $\mathbf{u} \in V$, $\|T(\mathbf{u})\| = \|\mathbf{u}\|$
(4): There exists an orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ for V, where $n = \dim(V)$, such that

$$\{T(\mathbf{w}_1), T(\mathbf{w}_2), \ldots, T(\mathbf{w}_n)\}$$

is orthonormal

Let $\mathbf{A}$ be an $n \times n$ complex matrix. Suppose $\mathbb{C}^n$ is equipped with the usual inner product. The following statements are equivalent:
(1): $\mathbf{A}$ is unitary
(2): The rows of $\mathbf{A}$ form an orthonormal basis for $\mathbb{C}^n$
(3): The columns of $\mathbf{A}$ form an orthonormal basis for $\mathbb{C}^n$

Let V be a complex finite dimensional inner product space, where $\dim(V) \geq 1$. If B and C are ordered orthonormal bases for V, then the transition matrix from B to C is a unitary matrix. That is,

$$[I_V]_{B,C} = ([I_V]_{C,B})^{-1} = ([I_V]_{C,B})^*.$$

12.6 Unitary and Orthogonal Diagonalisation

Let V be an inner product space and T be a linear operator on V such that T^* exists.
(i): If $T = T^*$, T is *self-adjoint*
(ii): If $T \circ T^* = T^* \circ T$, T is *normal*

Let $\mathbf{A}$ be a complex square matrix.
(i): $\mathbf{A}$ is a *Hermitian Matrix* if $\mathbf{A} = \mathbf{A}^*$
(ii): $\mathbf{A}$ is a *normal matrix* if $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$

If $\mathbf{A}$ is real and it satisfies $\mathbf{A} = \mathbf{A}^*$, it is symmetric. All Hermitian, real symmetric, unitary and orthogonal matrices are normal.

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and V be a finite dimensional inner product space over $\mathbb{F}$, where $\dim(V) \geq 1$, and T a linear operator on V. Take an ordered orthonormal basis B for V and let $\mathbf{A} = [T]_B$. It is clear that if $\mathbb{F} = \mathbb{C}$, T is self-adjoint if and only if $\mathbf{A}$ is Hermitian, whereas if $\mathbb{F} = \mathbb{R}$, T is self-adjoint if and only if $\mathbf{A}$ is symmetric. Lastly, T is normal if and only if $\mathbf{A}$ is normal.

If T is normal, we have the following properties:
(1): For all $\mathbf{u}, \mathbf{v} \in V$, $\langle T(\mathbf{u}), T(\mathbf{v}) \rangle = \langle T^*(\mathbf{u}), T^*(\mathbf{v}) \rangle$
(2): For any $c \in \mathbb{F}$, the linear operator $T - cI_V$ is normal
(3): If $\mathbf{u}$ is an eigenvector of T associated with λ, then $\mathbf{u}$ is an eigenvector of T^* associated with $\overline{\lambda}$
(4): If $\mathbf{u}, \mathbf{v}$ are eigenvectors of T associated with λ and μ respectively, where $\lambda \neq \mu$, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal

We will mainly discuss unitary diagonalisation since orthogonal diagonalisation has been discussed in MA2001. The only thing to add on here is regarding self-adjoint operators in relation to orthogonal diagonalisation since it should be clear from now that T is orthogonally diagonalisable if and only if T is self-adjoint.

Let $\mathbb{F} = \mathbb{C}$ and $\mathbb{V}$ be a finite dimensional inner product space over $\mathbb{F}$, where $\dim(V) \geq 1$ and T be a linear operator on V. Suppose there exists an ordered orthonormal basis V such that $[T]_B$ is a diagonal matrix. We say that T is unitary diagonalisable. In addition, a complex square matrix $\mathbf{A}$ is unitary diagonalisable if there exists a unitary matrix $\mathbf{P}$ such that $\mathbf{P}^*\mathbf{A}\mathbf{P}$ is a diagonal matrix.

Unitary Diagonalisation Process

Let T be a normal operator on a complex finite dimensional vector space V, where $\dim(V) \geq 1$.
Step 1: Find an orthonormal basis C for V and compute $\mathbf{A} = [T]_C$
Step 2: Express the characteristic equation $c_{\mathbf{A}}(x)$ as

$$c_{\mathbf{A}}(x) = \prod_{i=1}^{k} (x - \lambda_i)^{r_i},$$

where the λ_i's are distinct eigenvalues of $\mathbf{A}$ and $\sum_{i=1}^{k} r_i = \dim(V)$
Step 3: For each eigenvalue λ_i, find a basis for the eigenspace $E_{\lambda_i}(T)$ and use the Gram-Schmidt Process to transform it into an orthonormal basis B_{λ_i}
Step 4: Let

$$B = \bigcup_{i=1}^{k} B_{\lambda_i},$$

so B is an orthonormal basis for V. Using it as an ordered basis, $\mathbf{D} = [T]_B$ is a diagonal matrix.

To conclude, $\mathbf{D} = \mathbf{P}^*\mathbf{A}\mathbf{P}$, where $\mathbf{P} = [I_V]_{C,B}$ is the transition matrix from B to C.

Example: Suppose $T : \mathcal{M}_{2\times 2}(\mathbb{C}) \to \mathcal{M}_{2\times 2}(\mathbb{C})$ is the linear operator defined by

$$T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} -b - ic + id & -a + ic - id \\ ia - ib - d & -ia + ib - c \end{pmatrix} \text{ for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_{2\times 2}(\mathbb{C}).$$

Note that

$$\mathbf{A} = [T]_C = \begin{pmatrix} 0 & -1 & -i & i \\ -1 & 0 & i & -i \\ i & -i & 0 & -1 \\ -i & i & -1 & 0 \end{pmatrix},$$

which is Hermitian, and hence normal, so T is a normal operator. We have $c_{\mathbf{A}}(x) = (x+1)^3(x+3)$, so -1 and -3 are the eigenvalues. By using the Gram-Schmidt Process,

$$B_{-1} = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} i & -i \\ 2 & 0 \end{pmatrix}, \frac{1}{\sqrt{12}} \begin{pmatrix} -i & i \\ 1 & 3 \end{pmatrix} \right\}$$

and

$$B_{-3} = \left\{ \frac{1}{2} \begin{pmatrix} i & -i \\ -1 & i \end{pmatrix} \right\},$$

which are orthonormal bases for their respective eigenspaces. Let

$$B = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \frac{1}{\sqrt{6}} \begin{pmatrix} i & -i \\ 2 & 0 \end{pmatrix}, \frac{1}{\sqrt{12}} \begin{pmatrix} -i & i \\ 1 & 3 \end{pmatrix}, \frac{1}{2} \begin{pmatrix} i & -i \\ -1 & i \end{pmatrix} \right\}.$$

Then, it is clear that

$$\mathbf{D} = [T]_B = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Finding $\mathbf{P}$ is easy so I guess that is it! □

Example: Let $\mathbf{A}$ be a Hermitian Matrix. That is, $\mathbf{A} = \mathbf{A}^*$. For all $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, define

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^* \mathbf{A} \mathbf{u}.$$

It is known that $\langle \mathbf{u}, \mathbf{v} \rangle$ satisfies (IP1), (IP2) and (IP3). Find a condition on the eigenvalues of $\mathbf{A}$ such that $\langle \mathbf{u}, \mathbf{v} \rangle$ is an inner product.

Solution: We make reference to (IP4). Note that $\langle \mathbf{u}, \mathbf{v} \rangle$ is an inner product if and only if $\langle \mathbf{0}, \mathbf{0} \rangle = 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle > 0$ for all non-zero vectors $\mathbf{u} \in \mathbb{C}^n$. The former is obvious. For the latter, we use the method of unitary diagonalisation. Let $\mathbf{P}$ be a unitary matrix.

Suppose $\mathbf{P}^*\mathbf{A}\mathbf{P} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix, where the diagonal entries are the eigenvalues of $\mathbf{A}$, namely $\lambda_1, \lambda_2, \ldots, \lambda_n$. Note that $\langle \mathbf{u}, \mathbf{u} \rangle = \mathbf{u}^*\mathbf{A}\mathbf{u}$. Set $\mathbf{w} = \mathbf{P}^*\mathbf{u}$, so $\mathbf{P}\mathbf{w} = \mathbf{u}$ since $\mathbf{P} = \mathbf{P}^*$ by property of unitary matrices.

$$\langle \mathbf{u}, \mathbf{u} \rangle = \mathbf{w}^*\mathbf{P}^*\mathbf{A}\mathbf{P}\mathbf{w} = \mathbf{w}^*\mathbf{D}\mathbf{w}$$

Setting

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \ldots \\ w_n \end{pmatrix},$$

$$\langle \mathbf{u}, \mathbf{u} \rangle = \sum_{k=1}^{n} \lambda_k |w_k|^2.$$

Hence, the required condition is that all the eigenvalues of $\mathbf{A}$ must be positive and real. □