Notes on System of Linear Equations

Deepesh K P

October 2023

1 Introduction

The concept of a linear equation is motivated by the equation of a line. The general form of a straight line in the XY-plane is

$$ax + by = c$$

where a, b, c are constants. In the 3-dimensional space, the similar structure is that of the plane, which is

$$ax + by + cz = d.$$

Such type of equations are called linear equations. The general form of a linear equation in n variables $x_1, x_2, \ldots, x_n$ is

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

where a_i are constants, called the coefficients of the variables. By a solution to this equation, we mean a set of values for $x_1, x_2, \ldots, x_n$ which, on substitution, satisfy the equation. When there are more than one equation to be solved, we consider the collection of such equations as a "system of equations". A system of m linear equations in n variables is

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

which is known as the **general form of a linear system**. By a **solution** to this system, we mean a set of values for $x_1, x_2, \ldots, x_n$ which on substitution, satisfy each and every equation of the system.

Such systems arise naturally in solving problems. For example, while solving a chemical equation, we need to tally the number of atoms, which results in a collection of linear equations. A simple example of such a system is:

$$\begin{aligned} 2x + 3y &= 7 \\ x - 2y &= 5 \\ 3x + y &= 12 \end{aligned}$$

There are 3 equations and 2 variables in the above system.

2 Consistency of a system

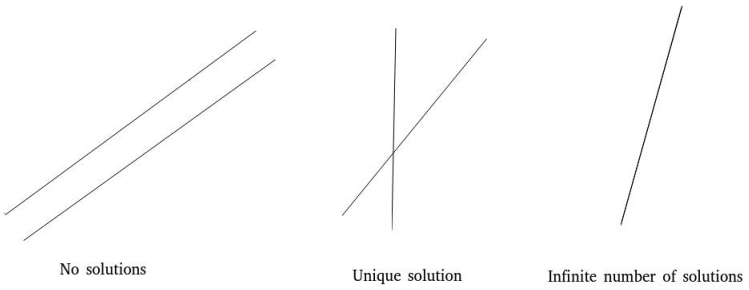
Some systems may have solutions and some others do not. We say that a linear system is consistent if it has a solution. Observe that a slight change in the coefficients may change the consistency of a system. For example, consider the three systems:

$$\begin{array}{rcl} 2x+3y &=& 7 \\ 6x+9y &=& 21, \end{array} \qquad \begin{array}{l} 2x+3y=7 \\ 6x+9.1y=21, \end{array} \qquad \begin{array}{l} 2x+3y=7 \\ 6x+9y=21.1 \end{array}$$

It can be seen that the first system has a lot of solutions like $x = 2, y = 1$ or $x = -1, y = 3$ etc., the second one has one and only one solution given by $x = \frac{-623}{2}, y = 210$; but the third one has no solutions.

Remark 2.1. For the first one, we can see that by choosing any value $x = a$, we can obtain $y = \dfrac{7-2a}{3}$, we get infinitely many solutions $(a, \dfrac{7-2a}{3})$ to the system.

Observe that geometrically these equations represent two straight lines in the two-dimensional plane. Hence three kinds of situations can happen here:



1. the lines may not intersect (parallel lines with a distance in between)
2. the lines may intersect at a unique point
3. the lines may lay one on the other (all points are same)

Thus for such a system of two equations in two variables, there can be

- no solution
- unique solution

- infinitely many solutions

So for a bigger linear system (more equations and more variables) we are interested in analyzing the following important questions:

1. Is there any solution to the system?
2. If so, is there a unique solution? or finitely many? or infinitely many?
3. How to find the unique solution?
4. How to find all solutions, if more than one?

3 Matrices and linear systems

We can beautifully use matrix theory to solve a linear system. In general, we can consider every linear system of m equations in n variables in the matrix form:

$$A\mathrm{x} = \mathrm{b},$$

where A is the $m \times n$ **matrix of coefficients**, x is the $n \times 1$ column **matrix of variables** and b is the $m \times 1$ column **matrix of constants**.

$$A = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix} \qquad \mathrm{x} = \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} \qquad \mathrm{b} = \begin{bmatrix} b_1 \\ b_2 \\ ... \\ b_m \end{bmatrix}$$

It is easy to see that the product on the left side of the matrix form corresponds to the left side of the equations of the general form. In the system $A\mathrm{x} = \mathrm{b}$, if $\mathrm{b} = 0$, the zero matrix, then the system is called a **homogeneous system**.

To solve linear systems using linear algebra techniques, we use the following concept of an **Augmented matrix** corresponding to a linear system:

$$[A \mid b] = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} & b_1 \\ a_{21} & a_{22} & ... & a_{2n} & b_2 \\ ... & ... & ... & ... & ... \\ a_{m1} & a_{m2} & ... & a_{mn} & b_m \end{bmatrix}$$

That is, the augmented matrix is obtained by joining the b matrix to the coefficient matrix A.

Example 3.1. The matrix form of the system of equations

$$\begin{aligned} 2x + 3y &= 7 \\ 6x + 9y &= 21 \end{aligned}$$

is

$$A = \begin{bmatrix} 2 & 3 \\ 6 & 9 \end{bmatrix} \qquad \text{x} = \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{b} = \begin{bmatrix} 7 \\ 21 \end{bmatrix}$$

and the augmented matrix corresponding to the system is given by

$$\begin{bmatrix} 2 & 3 & 7 \\ 6 & 9 & 21 \end{bmatrix}$$

Note that the rows of the augmented matrix correspond to the equations in the linear system of equations.

Now, suppose that the system which we need to solve is in a nice form like

$$\begin{aligned} 2x + 3y - z &= 7 \\ 11y - 3z &= 1 \\ 2z &= 14. \end{aligned}$$

Then the last equation is a very simple equation which you can solve easily to get $z = 7$. Now since the second equation has only two variables, substituting $z = 7$ into it, we can solve it and $y = 2$. Now the remaining variable x can be fixed using the first equation as $x = 4$. This process is called **back substitution method**. So if the system is in this kind of a simpler form, back substitution gives the solution very fast.

Can we reduce every matrix to such a simpler form?

The answer is yes. But for this, we need to transform the equations in such a way that the solutions should not change.

4 Elementary row operations

Let us see some of the simple operations which do not change the solutions of a system and reduce the system into a 'simpler form' (we will describe what is meant by the 'simpler form' afterwards).

• First, let us observe that even if you interchange the positions of the system, nothing happens to the system! Its augmented matrix will have an interchange of its rows. For example,

$$\begin{aligned} 2x + 3y &= 7 \Rightarrow \qquad 6x + 9y = 2 \\ 6x + 9y &= 2, \qquad\quad\ \ 2x + 3y = 7 \end{aligned}$$

will have the same solution. Note that the corresponding augmented matrix will have an interchange of rows:

$$A = \begin{bmatrix} 2 & 3 & 7 \\ 6 & 9 & 2 \end{bmatrix} \qquad \Rightarrow \tilde{A} = \begin{bmatrix} 6 & 9 & 2 \\ 2 & 3 & 7 \end{bmatrix}$$

• Also one can see that if both sides of an equation is multiplied by the same non zero scalar, the system (also the solution) will not be affected. For example,

$$\begin{aligned} 2x+3y &= 7 \Rightarrow \qquad 4x+6y=14 \\ 6x+9y &= 2, \qquad\quad 6x+9y=2 \end{aligned}$$

will have the same solution. Here we multiplied the first equation by 2. Note that the corresponding augmented matrix will have an interchange of rows:

$$A = \begin{bmatrix} 2 & 3 & 7 \\ 6 & 9 & 2 \end{bmatrix} \qquad \Rightarrow \tilde{A} = \begin{bmatrix} 4 & 6 & 14 \\ 6 & 9 & 2 \end{bmatrix}$$

• Further, if we add two equations (1) and (2) and form a new equation (3), the solutions of the equations (1) and (2) will be the same as the solutions of (1) and (3) (and the solutions of (2) and (3) also). Hence equation (3) can be used in place of one of the first two equations to have the solution of the system intact. For example,

$$\begin{aligned} 2x+3y &= 7 \Rightarrow \qquad 8x+12y=9 \\ 6x+9y &= 2, \qquad\quad 6x+9y=2 \end{aligned}$$

will have the same solution. Here we added the two equations and replaced the first equation with the sum. Note that the corresponding augmented matrix will have an interchange of rows:

$$A = \begin{bmatrix} 2 & 3 & 7 \\ 6 & 9 & 2 \end{bmatrix} \qquad \Rightarrow \tilde{A} = \begin{bmatrix} 8 & 12 & 9 \\ 6 & 9 & 2 \end{bmatrix}$$

Observe that the three processes do not change the solutions of the system. It is important to observe that the changes happening in the system are completely given by the corresponding changes in the row vectors of the augmented matrix (the coefficient matrix does not take care of the right side of the equations). Thus an augmented matrix will suffice to represent the system completely and the operations listed above are 'row operations' done to the rows of the augmented matrix.

Here we formally define the **elementary row operations** applicable to any matrix. They are

1 Interchange two rows ($R_1 \Leftrightarrow R_3$)

2 Multiply an entire row with a scalar (real number $\alpha \neq 0$) ($R_1 \Rightarrow \alpha \times R_1$)

3 Adding two rows and replacing one of these rows with it ($R_1 \Rightarrow R_1 + R_3$)

Remark 4.1. Note that combinations of the three operations are also allowed ($R_1 \Rightarrow \alpha \times R_1 + R_3$) and they can be repeatedly applied. These operations do not change the solutions of the linear system represented by an augmented matrix.

On applying the elementary row operations to an augmented matrix, we want to transform it into a simpler form (recall the example given after Example 3.1), which is known as the **row echelon form** of a matrix. Let us define it formally here for any matrix.

Definition 4.2. *A matrix is said to be in the Row Echelon form if*

i. Zero rows are all kept at the bottom

*ii. There are only zeros below (in the column) the first non-zero entry (called the **pivot** of the row) of each row.*

iii. The pivot of the lower row should lie to the right side of the pivot of the row above.

For example, the matrix A given below is a matrix in the row echelon form.

$$A = \begin{bmatrix} 3 & 0 & -2 & 1 \\ 0 & -4 & 7 & 0 \\ 0 & 0 & 5 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Zero rows are kept at the bottom as they represent a null equation (which do not contribute any data) and are unimportant.

Note that in order to solve the system, it is enough to solve any of its row-equivalent systems (since all will have the same solutions) and the system is very simple to solve once it reaches the row-echelon form.

Definition 4.3. *If we get a matrix Q by applying a combination of elementary row operations to a given matrix P, then we say that P is **row equivalent** to the matrix Q (and vice-versa).*

Note that two row-equivalent augmented matrices have the same solutions (as row operations do not change the solutions). Hence instead of solving the first system, one can solve the second system and get the solution of the first system. However, we will benefit from this technique only if the second system (matrix) is easier to use. Here by 'easiest form' we mean the **row echelon form**.

We can reduce any given matrix to an Echelon form by using elementary row operations. We can see that the row echelon matrix is simple as it contains many zeros. It looks almost like a triangular matrix (not exactly).

Remark 4.4. If the pivots are made 1 and the entries above the pivot also are eliminated (we can always do this using elementary row operations) then such a matrix is said to be in the **row reduced echelon form**. Here is an example for row reduced echelon matrix. An example is given here

$$[A] = \begin{bmatrix} 1 & \frac{3}{5} & 0 & 0 & \frac{21}{25} \\ 0 & 0 & 1 & 0 & \frac{56}{20} \\ 0 & 0 & 0 & 1 & \frac{8}{5} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Let us now see an example where the row echelon form helps to solve a linear system.

$$\begin{aligned} x + y + z &= 6 \\ 2x - y - z &= 3 \\ x + 2y - z &= 4 \\ 4x + 2y - z &= 13 \end{aligned}$$

The augmented matrix of the system is

$$[Ab] = \begin{bmatrix} 1 & 1 & 1 & 6 \\ 2 & -1 & -1 & 3 \\ 1 & 2 & -1 & 4 \\ 4 & 2 & -1 & 13 \end{bmatrix}$$

One of the echelon form that you reach by doing elementary row operations ($R_2 \to R_2 - 2R_1$, $R_3 \to R_3 - R_1$, $R_4 \to R_4 - 4R_1$; $R_2 \longleftrightarrow R_3$, $R_3 \to R_3 + 3R_2$, $R_4 \to R_4 + 2R_2$; $R_4 \to R_4 - R_3$) is

$$[Ab] \sim \begin{bmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & -2 & -2 \\ 0 & 0 & -9 & -15 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Now the system corresponding to the echelon matrix is

$$\begin{aligned} x + y + z &= 6 \\ y - 2z &= -2 \\ -9z &= -15 \\ 0 &= 0 \end{aligned}$$

We can observe that the equations get simpler and simpler as we go down and the last row has vanished. Solve back from the third equation and solving for the variables one by one, we get the solution

$$x = 3,\ y = \frac{4}{3},\ z = \frac{5}{3},$$

which is the only possible solution for the given system.

Since the elementary row operations we applied will not change the solutions of the system, we can say that every equivalent matrices that you get have the same solution as that of the original one. But since the one given by the echelon form is the simplest to solve among these, we can solve the system after making the augmented matrix into the echelon form.

5 Linearly independent row vectors and the rank of a matrix

We have seen that some rows has got vanished (by applying the row transformations using the other rows) when the matrix reaches in the echelon form. This means that these rows can be created also (by applying the opposite row operations that were applied for eliminating these rows) using the other rows by doing row operations. Such rows are said to be **dependent rows** as they are dependent on the other rows.

We say that a vector is dependent on other vectors if it can be created (or eliminated) using other vectors by means of row operations.

Definition 5.1. *We say that the row vectors $\{v_1, v_2, \ldots, v_n\}$ of an $m \times n$ matrix are **linearly dependent** if one among these vectors can be written as a linear combination of the other vectors. This is equivalent to saying that a non-zero combination of the vectors if equal to zero vector, then the vectors are linearly independent. This means*

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = 0 \text{ with at least one } \alpha_i \neq 0,$$

as in this case,

$$v_i = -\frac{\alpha_1}{\alpha_i} v_1 - \frac{\alpha_2}{\alpha_i} v_2 - \ldots - \frac{\alpha_{i-1}}{\alpha_i} v_{i-1} - \frac{\alpha_{i+1}}{a_i} v_{i+1} - \frac{\alpha_n}{\alpha_i} v_n,$$

which means that v_i can be written in terms of others.

For example, consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 5 & 7 & 5 & 0 \end{bmatrix}$$

Here $(1, 2, 1, 0), (1, 1, 1, 0), (5, 7, 5, 0)$ are dependent vectors since $v_3 = 2v_2 + 3v_3$ or in other words, $2v_2 + 3v_3 + -1v_3 = (0, 0, 0)$. Note that we may also write $v_2 = \frac{1}{2} v_3 - \frac{3}{2} v_1$. But the matrix

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

has its rows are not independent since none of the rows can be written in terms of the others.

Definition 5.2. *We say that the row vectors $\{v_1, v_2, \ldots, v_n\}$ of an $m \times n$ matrix are **linearly independent** if*

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = 0 \Rightarrow a_1 = \alpha_2 = \ldots = \alpha_n = 0,$$

as in this case, there is no nonzero relation between the vectors.

There is a close connection between the number of independent rows of the coefficient (as well as augmented matrix) of a system and the number of solutions of the system. We first define the rank of a matrix in this regard.

Definition 5.3. *The number of independent row vectors of a matrix is called the* ***row rank*** *(or* ***rank****) of the matrix. It is denoted as* $rank(A)$ *or* $\rho(A)$ *for a matrix* A*.*

For example, the rank of the matrix B given above is 3 since all rows are independent. But the rank of the matrix A is 2 since the first two rows are independent and the third row is dependent on the first two.

As we said earlier, if one vector can be created using others, it can also be deleted using other vectors. Hence, if some row is getting deleted while making the matrix into row-echelon form, then that row must be dependent on others. Hence, it can be observed that:

Proposition 5.4. *The rank of a matrix is given by the number of non-vanishing rows of the echelon form of the matrix.*

If we reduce the matrices A and B above to echelon forms, it can be seen that they contain 2 and 3 non-vanishing rows, respectively.

Remark 5.5. Note that there is an analogous concept of **column rank** of a matrix. It is a famous result in matrix theory that row rank = column rank always. However, we will not be using column operations or the concept of column rank here.

6 Rank and the consistency of a system

As we mentioned, there is a close connection between the number of solutions of a system and the ranks of the coefficient and augmented matrices of the system. Suppose that the rank of the coefficient matrix is less than that of the augmented matrix. Then there will be a zero row in the echelon matrix of A corresponding to which the column of b will contain a non-zero entry. For example, consider

$$[Ab] = \begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ 0 & b_2 & c_2 & d_2 \\ 0 & 0 & 0 & d_3 \end{bmatrix}$$

Here $\rho(A) = 2$ and $\rho(Ab) = 3$ (assuming $d_3 \neq 0$) and so they are not equal.

In such a case, there will be an equation which gives $0 = d_3$, which can not have a solution as $d_3 \neq 0$! Thus whenever $\rho(A) \neq \rho(Ab)$, the system will not have a solution.

Remark 6.1. A more rigorous argument can be given using linear algebra techniques. It uses the idea that the vector b is not in the range or column space (span of column vectors) of A.

Now suppose $r = rank(A) = rank(Ab) < n$, where n is the number of variables. For example, suppose the augmented matrix of a system gets into the echelon form

$$[Ab] = \begin{bmatrix} a_1 & b_1 & c_1 & d_1 \\ 0 & b_2 & c_2 & d_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Here $rank(A) = 2 = rank(Ab)$ and number of variables $n = 3$. Then going back to the system of equations, we can see that

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ b_2y + c_2z &= d_2 \\ 0 &= 0 \end{aligned}$$

In such a case, we can see that effectively there are only r (here, 2) equations to control the 3 variables. As a result, $n - r$ (here $3 - 2 = 1$) variables have to be let free (called the free variable or excess variable) as ideally to control n variable, we need n independent equations. The free variable can take any value, resulting in an infinite number of solutions to the system.

For example, here since z has no pivot in its column, it is not controlled by any equation. Hence we need to assign $z = a$, where a can be arbitrary constant. As a result, the system gets into the form

$$\begin{aligned} a_1x + b_1y + c_1a &= d_1 \\ b_2y + c_2a &= d_2 \end{aligned}$$

giving the solutions $z = a$, $y = \frac{d_2 - c_2a}{b_2}$, $x = \frac{d_1 - b_1\frac{d_2 - c_2a}{b_2}}{a_1}$ on back substitution. Thus, we get an infinite number of solutions by assigning different values to the free variable $z = a$.

Thus we can say that if $\rho(A) = \rho(Ab) \neq n$, the number of variables, then the system will have infinitely many solutions.

But if $\rho(A) = \rho(Ab) = n$, then there will be n equations effectively to control the n variables, assigning a unique value to each variable, resulting in a unique solution to the system.

Putting all together, we state the *fundamental theorem on linear system of equations* in the following.

Theorem 6.2. *For a system* $A\mathrm{x} = \mathrm{b}$ *of* m *equations in* n *variables*

1. *no solution exists if and only if* $\rho(A) \neq \rho(Ab)$
2. *solutions exists if and only if* $\rho(A) = \rho(Ab)$
    - i. *unique solution exists if and only if* $\rho(A) = \rho(Ab) = n$
    - ii. *infinite number of solutions exist if and only if* $\rho(A) = \rho(Ab) \neq n$

7 homogeneous systems

A linear system $Ax = b$ is homogeneous if $b = 0$. Such a system always has a so-called "trivial solution", which is $\mathrm{x} = 0$. It is easy to see that

$$x_1 = 0,\, x_2 = 0,\, \ldots, x_n = 0$$

forms a solution to any homogeneous system. Hence, it is always consistent. This can be seen from $rank(A) = rank(Ab)$ also, as the last column b is fully zero and it makes no impact on the rank of Ab. For example,

$$\begin{aligned} 2x + 3y &= 0 \\ 6x + 9y &= 0 \end{aligned}$$

is a homogeneous system and $x = 0$, $y = 0$ is clearly a solution to it.

So for a homogeneous system, if $rank(A) = n$, we do not have to solve it as the only solution is $x_1 = x_2 = \ldots = x_n = 0$. Thus, we need to solve it only in the case of non-trivial solutions, which happens if and only if $rank(A) \neq n$.

The solutions of homogeneous systems have many special properties.

Theorem 7.1. *If* x_1 *and* x_2 *are two solutions of a homogeneous system, then*

i. $\mathrm{x}_1 + \mathrm{x}_2$ *is also a solution.*

ii. $\alpha\mathrm{x}_1 + \beta\mathrm{x}_2$ *is also a solution for the system for any scalars* α *and* β*.*

Proof. Let the system be $\mathrm{Ax} = 0$. Since x_1 and x_2 are two solutions,

$$\mathrm{Ax}_1 + \mathrm{Ax}_2 = 0 \Rightarrow \mathrm{A}(\mathrm{x}_1 + \mathrm{x}_2) = 0.$$

Similarly for any scalar α, we get

$$\alpha\mathrm{Ax}_1 = 0 \Rightarrow \mathrm{A}(\alpha\mathrm{x}_1) = 0.$$

Combining the above, we get (ii). □

Hence, the collection of all solutions contains all sums and scalar products of solutions. This collection of all the solutions is known as the solution space of the system or the null space of the coefficient matrix.

Definition 7.2. *The solution space of a homogeneous system* $\mathrm{Ax} = 0$ *is*

$$\{\mathrm{x} :\, \mathrm{Ax} = 0\}.$$

The *null space* of a matrix is one of the four important spaces associated with the matrix.

Definition 7.3. *For a matrix A, the null space* $N(A)$ *is the solution space of the associated homogeneous system* $\mathrm{Ax} = 0$*. That is*

$$N(A) = \{\mathrm{x} :\, \mathrm{Ax} = 0\}.$$

Note that the solution space and the null space are effectively the same. For a system, it is called the solution space, and for the matrix, it is the null space. These concepts are very important in **linear algebra**, as they provide **subspaces** of the **vector space** $\mathbb{R}^n$.

The maximum number of independent vectors in the null space of a matrix has got a special name.

Definition 7.4. *The nullity $n(A)$ of a matrix A is the number of independent vectors in $N(A)$. Equivalently, it is the number of independent solutions of the homogeneous system $\mathrm{Ax} = 0$. Note that if there is only the zero vector in the null space, the nullity is considered as 0.*

For finding the nullity of a matrix, we need to solve the corresponding homogeneous system, find all the solutions of it and count how many of them are independent. The following is an example.

Example 7.5. Consider the following system of 3 equations in 4 variables:

$$\begin{aligned} x - y + 5z + w &= 0 \\ 2x - 3y + z - w &= 0 \\ 3x - 4y + 6z &= 0 \end{aligned}$$

The echelon form of the augmented matrix is

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

See that the rank of the coefficient and augmented matrices is both 2 (since there are 2 non-zero rows, or equivalently, 2 pivots in the matrices). The number of variables $n = 4$ and hence there are $4 - 2 = 2$ excess variables. Choosing the non-pivoted variables as the excess variables $z = a$ and $w = b$, we have the solutions

$$\begin{bmatrix} -14a - 4b \\ -9a - 3b \\ a \\ b \end{bmatrix}$$

for each real numbers a, b. Note that the solutions space of $\mathrm{Ax} = 0$ or the null space of A is

$$N(A) = \left\{ \begin{bmatrix} -14a - 4b \\ -9a - 3b \\ a \\ b \end{bmatrix} : a, b \in \mathbb{R} \right\}$$

This space can also be written as

$$N(A) = \left\{ a \begin{bmatrix} -14 \\ -9 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} -4 \\ -3 \\ 0 \\ 1 \end{bmatrix} : a, b \in \mathbb{R} \right\}$$

This means that all the solution vectors are combinations of the two vectors $v_1 = \begin{bmatrix} -14 \\ -9 \\ 1 \\ 0 \end{bmatrix}$ and $v_2 = \begin{bmatrix} -4 \\ -3 \\ 0 \\ 1 \end{bmatrix}$ and so all other solutions are dependent on these two only. Hence, there are only two solutions which are independent, giving the nullity 2.

One can also observe that the nullity is the number of excess variables while solving the system. Now the rank was the number of independent rows, which is effectively the number of pivoted variables (non-excess variables) of the system. So the number of pivoted variables added with the number of excess variables gives the total number of variables! This fact is known as the famous **Rank-Nullity Theorem** in linear algebra.

Theorem 7.6. *(Rank-Nullity Theorem)*
For an $m \times n$ matrix A, $rank(A) + nullity(A) = n$.

This theorem helps us to find the nullity of a matrix when the rank is known. For example, the above matrix has rank 2 and $n = 4$. So nullity must be $4 - 2 = 2$. It also helps us to conclude that if the rank of the matrix is the same as the number of columns, then $Ax = 0$ will have only the trivial solution.

When it is a square matrix, we know that it has a determinant, and there is a close connection between the determinant and the rank of a square matrix.

Theorem 7.7. *An $n \times n$ matrix A has $rank(A) = n$ if and only if $|A| \neq 0$.*

Proof. Observe that the rank of the square matrix is n if and only if there are n pivots in its echelon form. This happens if and only if the echelon form is an upper triangular matrix, with all diagonal entries non-zero.

Also, the elementary row transformations change the determinant very little. Suppose that the determinant of A is α. Then interchanging makes it $-\alpha$, scalar multiplication with a non-zero c makes it $c\alpha$ and adding of two rows and replacing one of them makes no change to the determinant. Thus, if A has $|A| = 0$, then after row transformations, the determinant is 0 and vice versa.

Thus, if $rank(A) = n$, the echelon form is an upper triangular matrix with non-zero diagonal entries, and so, its determinant is non-zero and hence $|A| \neq 0$. Conversely, if $rank(A) \neq n$, then the echelon matrix is a diagonal matrix with one diagonal entry equal to 0, which gives $|A| = 0$ in a similar way. □

The above information is quite useful while finding the characteristic equation of a square matrix.

8 Additional reading materials

We mentioned the vector spaces and subspaces, null spaces and all. Here we give some basic definitions of these concepts.

8.1 Vector spaces

Definition 8.1. *A non-empty set V together with two operations $+$ (called vector addition) and . (scalar multiplication) is called a vector space if it satisfy the following properties (here we assume that u, v, w are arbitrary elements of V and $\alpha,\ \beta$ are arbitrary real numbers):*

1. $u + v \in V$
2. $u + v = v + u$
3. $u + (v + w) = (u + v) + w$
4. *there is an element* 0 *in* V *such that* $0 + u = u$
5. *there is an element* $-u$ *in* V *such that* $u + (-u) = 0$
6. $\alpha.u \in V$
7. $(\alpha\ + \beta).u = \alpha.u + \beta.u$
8. $\alpha.(u + v) = \alpha.u + \alpha.v$
9. $(\alpha\ \beta).u = \alpha\ (\beta.u) = \beta\ (\alpha.u)$
10. $1.u = u$

The elements of a vector space V are called *vectors* and the real numbers $\alpha,\ \beta, \ldots$ are called *scalars*. A set should satisfy all the 10 properties of the operations if it is to be called a vector space. If any one of these 10 properties is not satisfied, the set can not be a vector space.

Example 8.2.

1. The set of all $m \times n$ matrices denoted by $\mathcal{M}_{m\times n}$, for each $m, n \in \mathbb{N}$, with usual matrix addition and scalar multiplication.
2. The set $\mathbb{R}^n$ of all $n-$ tuples with usual addition of n-tuples and multiplication of an $n-$tuple by a number.

$$\begin{aligned}(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)\\ \alpha\ (x_1, x_2, \ldots, x_n) &= (\alpha\, x_1, \alpha\, x_2, \ldots, \alpha\, x_n)\end{aligned}$$

   Since we can consider $(x_1, x_2, \ldots, x_n)$ as a row matrix or column matrix, one may think of this as a special case of [1], the cases of $1 \times n$ row/$n \times 1$ column matrices)
3. The set of all real numbers $\mathbb{R}$ (special case of [2], when $n = 1$)
4. The set $\mathcal{P}_n$ of all polynomials of order less than or equal n with usual polynomial addition and multiplication of polynomial with numbers.

$$\begin{aligned}(a_0 + a_1 t + \ldots + a_n t^n) &+ (b_0 + b_1 t + \ldots + b_n t^n)\\ &= (a_0 + b_0) + (a_1 + b_1)t + \ldots + (a_n + b_n)t^n\\ \alpha\,(a_0 + a_1 t + \ldots + a_n t^n) &= \alpha\, a_0 + \alpha\, a_1 t + \ldots + \alpha\, a_n t^n\end{aligned}$$

5. The set of all polynomials $\mathcal{P}$, as an extended case of [4].

6. The set of all real-valued functions defined on $[a, b]$ with

$$\begin{aligned}(f+g)(x) &= f(x)+g(x), \quad x \in [a,b] \\ (\alpha\, f)(x) &= \alpha.\, f(x), \quad x \in [a,b].\end{aligned}$$

Now we give some examples of sets which are not vector spaces.

Example 8.3. Let V be the set of all matrices of the form

$$\left\{ \begin{bmatrix} a & b \\ c & 1 \end{bmatrix} : a, b, c \in \mathbb{R} \right\}$$

with usual addition and scalar multiplication. Here one can observe that V does not satisfy the properties (1) and (6). To see that (1) is not satisfied,

$$\begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 4 & 2 \end{bmatrix} \notin V$$

So, even though V satisfy the properties (2), (3), (7), (8), (9), and (10), it is not a vector space.

The following is the concept of a subspace of a vector space.

Definition 8.4. *Let V be a vector space. A subset W of V is said to be a subspace of V if W itself is a vector space under the induced operations.*

This means that a vector space sitting inside a bigger vector space is called a subspace. Of course, the operations in the subspace should be the same as those of the bigger space.

Example 8.5. Let W be the set of all matrices of the form

$$\left\{ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} : a, b \in \mathbb{R} \right\}$$

with addition and scalar multiplication as in $\mathcal{M}_{2\times 2}$. Obviously, this collection is a subset of $\mathcal{M}_{2\times 2}$. One can easily check that the properties (1) to (10) of a vector space are satisfied by this W. Hence W is a subspace of V.

Observe that the null space of any $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. To see this, recall that

$$N(A) = \{\mathrm{x} \subseteq \mathbb{R}^{\mathrm{n}} : \mathrm{Ax} = 0\}.$$

Then $\mathrm{x}_1 + \mathrm{x}_2$ and $\alpha \mathrm{x}$ are also a solutions of $A\mathrm{x} = 0$ for any solutions $\mathrm{x}_1, \mathrm{x}_2$ and any scalar α. Thus $N(A)$ is a subspace.