Solving systems of linear equations using matrices

Introduction and prerequisite knowledge

Matrices are a concept taught in introductory *Linear algebra*, a branch of mathematics that has existed for thousands of years. Dating back to ancient Babylonians and eventually being rediscovered around the 17^{th} century and expanded on until the 19^{th} century with matrix theory being developed by Hamilton and Cayley in particular.

Matrices, and their applications to solve linear systems are shown in textbooks following the Syllabus for Further Mathematics AS level/year 1 in the UK. However, this essay focusses on other methods to solve systems of linear equations using matrices. I am assuming that the reader of this essay is familiar with the basic operations involving matrices as expressed as a part of the syllabus for Further maths a-level in the UK, e.i multiplication, addition, etc.

This essay hopes to cover some knowledge unknown to the usual Further Mathematics student, and provide a more holistic understanding of Matrices, their relation to linear systems and most importantly, what each method represents in logical and geometric models.

Row view

The row view considers matrices in a linear system with regards to the rows of the matrix and their interpretation as equations.

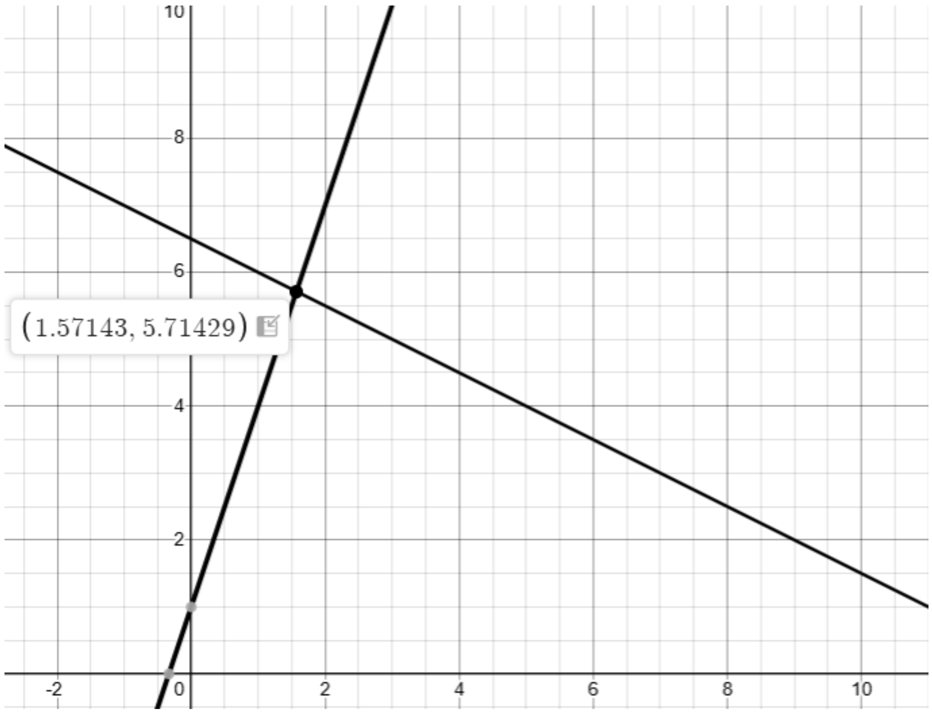
Using matrices, and vectors, a system of linear combinations of the variables x and y is given:

$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 13 \\ -1 \end{bmatrix}$$

The system of equations given can be written as follows:

$$\begin{cases} x + 2y = 13 & l_1 \\ 3x - y = -1 & l_2 \end{cases}$$

These equations can be solved using a variety of means, such as by using a graphical approach in a cartesian grid. This is because, the rows of the matrix, containing the coefficients of the first and second equations, contains only 2 elements, corresponding to a linear combination of x and y for each row. Which pairs nicely with the cartesian grid. Ultimately giving the values for x and y that satisfy the set of equations.



This method remains consistent for higher dimensions as well, although approaching them is quite tricky since it becomes harder to visualise higher dimensions.

- Given system of equations:

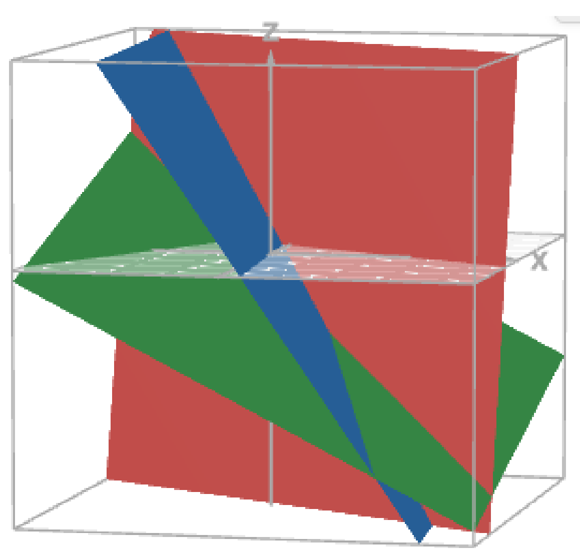
$$\begin{bmatrix} 4 & 7 & -2 \\ -10 & -1 & -7 \\ -2 & 1 & -4 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 21 \\ 0 \\ 9 \end{bmatrix}$$

- Written as:

$$\begin{cases} 4x + 7y - 2z = 21 & l_1 \\ -10x - y - 7z = 0 & l_2 \\ -2x + y - 4z = 9 & l_3 \end{cases}$$

Note that the defined equations need to be considered in three dimensions due to each equation containing a linear combination of 3 variables. Also, each equation represents a plane and not a line. For the aforementioned set of equations, the solution here is, x=2 y=1 z=-3, which are the coordinates for the point of intersection of all planes.

On a 3D graph with axis’: x, y and z



The row view functions due to the means by which multiplication of matrices and vectors are carried out. Allowing for the distribution of elements within the vector (variables of x, y and z in the example), to the coefficients of the variables of the linear equations expressed by the matrix, where each row acts as each equation's variable's coefficients.

Column view

The column view considers matrices in a linear system with regards to the columns of the matrix and their interpretation as vectors.

Using matrices, and vectors, a system of linear combinations of the variables x and y is given:

$$\begin{bmatrix} 8 & -3 \\ 5 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

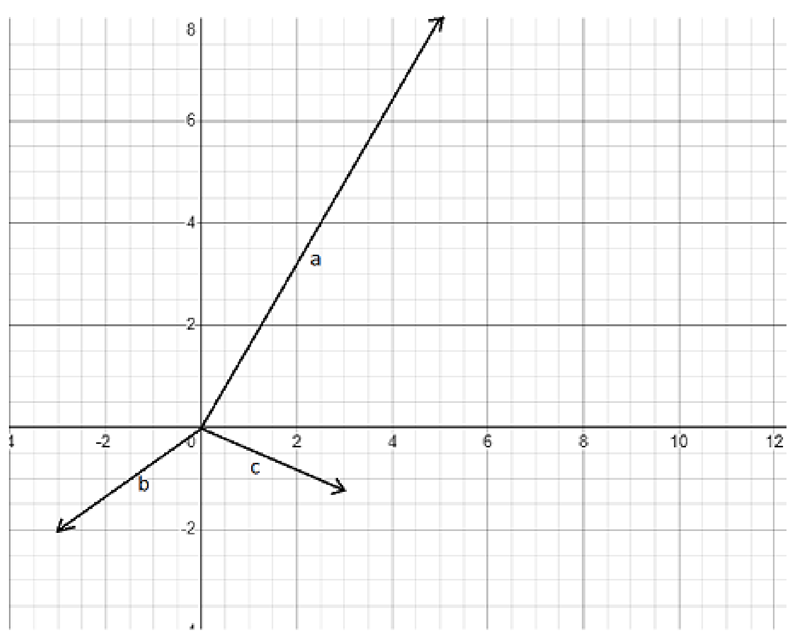
Can be written as:

$$x\begin{bmatrix} 8 \\ 5 \end{bmatrix} + y\begin{bmatrix} -3 \\ -2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

This is because, by means of multiplying the matrix and vector, the elements in the first column of the matrix are multiplied by the top-most element in the vector, while the elements in the second column are multiplied by the bottom most element (in this example).

This means that, the matrix in this example can be interpreted as a set of linear combinations of Vectors that sums to a final vector.

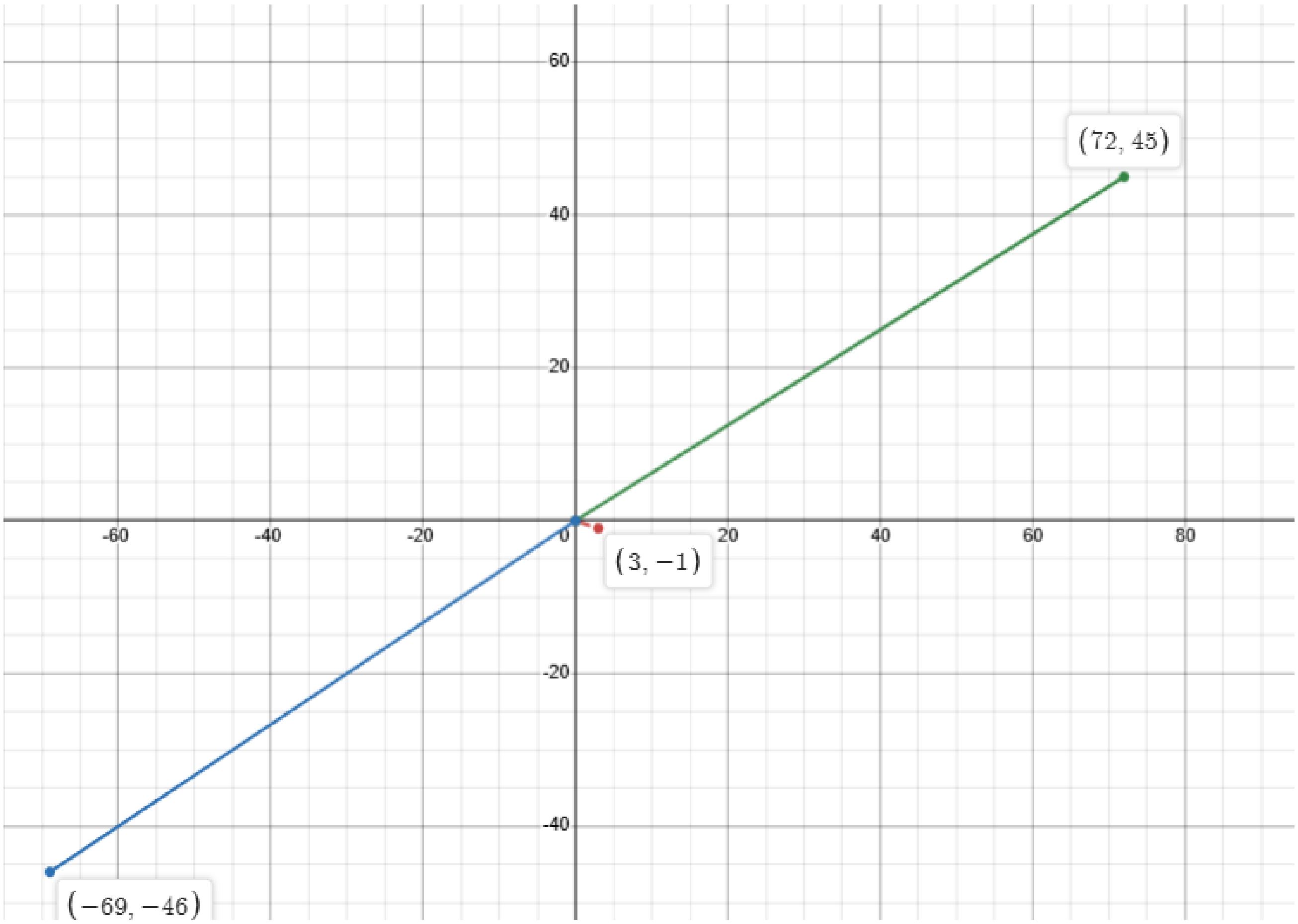
On a graph, this is represented as:



Where ***c*** is the result of a combination of vectors ***a*** and ***b*** as specified by the equation. ***a*** is the column vector with the parameter of x, and ***b*** is the column vector with the parameter of y, where x and y are scalar quantities.

The final asnwer is x=9 and y=23

Shown below:



The green line is 9***a***, the blue line is 23***b*** and the red line is the sum of this linear combination of vectors ***a*** and ***b***, clearly showing the vector defined in the equation, which we dubbed ***c***.

The column veiw is particularly useful in multiplying matrices, it allows the application of a deconstructed approach of the typical method of multiplying matrices found in the Further maths syllabus.

Consider you are calculating the results of the product of the matrices:

$$\begin{bmatrix} 5 & -1 & 2 \\ 8 & 3 & -4 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ 9 & -3 \\ 7 & 4 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

This can be re-written as shown below to get the results of the column vectors, of the resulting matrix

$$2\begin{bmatrix} 5 \\ 8 \end{bmatrix} + 9\begin{bmatrix} -1 \\ 3 \end{bmatrix} + 7\begin{bmatrix} 2 \\ -4 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix}$$

$$2\begin{bmatrix} 5 \\ 8 \end{bmatrix} - 3\begin{bmatrix} -1 \\ 3 \end{bmatrix} + 4\begin{bmatrix} 2 \\ -4 \end{bmatrix} = \begin{bmatrix} b \\ d \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 15 & 21 \\ 15 & -9 \end{bmatrix}$$

Gaussian elimination

Perhaps the quickest way to solve a system of linear equations is using Gaussian elimination also known as row reduction. It is a sequence of steps involving the rows of the matrix to find the linear combination of elements that satisfies every equation.

The steps used are called **Elementary row operations**. They consist of:

- Row swapping.
- Row multiplication, with a non 0 constant.
- Row addition.

Consider the system: also written as:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & -1 \\ -3 & 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 8 \\ 1 \\ -1 \end{bmatrix}$$

$$\begin{cases} x + 2y + z = 8 \\ 2x + y - z = 1 \\ -3x + y + 2z = -1 \end{cases}$$

Using an augmented matrix, another column is added containing the value of each row respectively.

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

The overall goal of this section, using row operations, is to diminish the matrix into its **row echelon form**. There will be pivots in this form, which are specific non zero elements used to eliminate other entries in the column that it occupies. In this case, the first pivot is the number 1 on the first column and row, top left most element. This will be used to alter the second row, such that the elements under the first pivot is 0.

If you multiply the first row by 2, and subtract it from the second row, the matrix turns into:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \\ 0 & -3 & -3 & -15 \\ -3 & 1 & 2 & -1 \end{array}\right]$$

If you multiply the first row by 3, and add it to the third row, the matrix turns into:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \\ 0 & -3 & -3 & -15 \\ 0 & 7 & 5 & 23 \end{array}\right]$$

Following along with these steps, if we divide row 2 by -3 (which is equivalent to multiplying by -1/3), we get this matrix:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \\ 0 & 1 & 1 & 5 \\ 0 & 7 & 5 & 23 \end{array}\right]$$

From here the 1 in the second row and second column becomes the second pivot, and we use this pivot to eliminate the element below, in the third row. If we multiply row 2 by 7 and subtract that from row three, we are left with:

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 8 \\ 0 & 1 & 1 & 5 \\ 0 & 0 & -2 & -12 \end{array}\right]$$

Finally, if we divide the bottom row by -2 we are left with:

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

This can be interpretted as:

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ 6 \end{bmatrix}$$

also written as:

$$\begin{cases} x + 2y + z = 8 \\ y + z = 5 \\ z = 6 \end{cases}$$

The matrix is now in **row echelon form(upper triangluar),** by using **back substitution** you can easily work out the values of x, y and z(even though z has been worked out in the example).

x=4

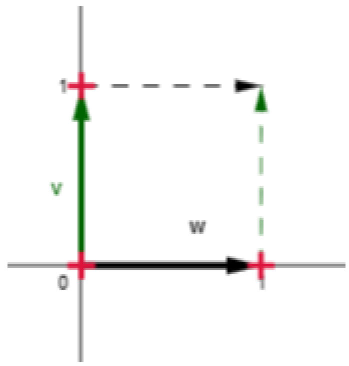
y=-1

z=6

Determinants

The determinant is a scalar value calculated from a **square matrix** that describes the scaling factor of a matrix, under the interpretation of matrices as linear transformations. The determinant determines how the area changes for the product of the base vectors in a matrix, first in the identity matrix, where the area is always 1, and secondly in the matrix carrying out the transformation. Note that the use of “area” is a bit inaccurate, as it suggests a limitation to 2 dimensions, but the determinant is not limited to only 2 by 2 matrices, hence the determinant in a 3 by 3 matrix is in reference to volume.

The identity matrix: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ Can be represented as:

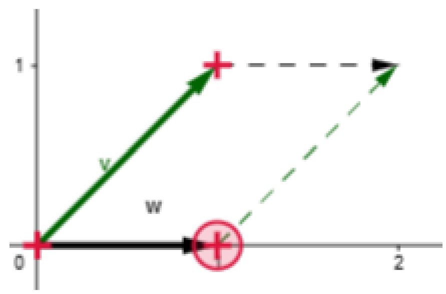


due to the column view of matrices.

The determinant of this matrix is the area enclosed by the vectors in the arrangement above =1

Consider another matrix: $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

Can be geometrically represented as:

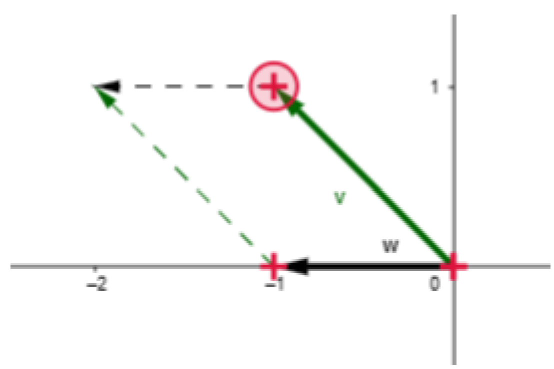


Since the area, as a result of multiplying the base vectors is still 1, the determinant of this matrix is also 1, however if we consider this matrix: $\begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix}$

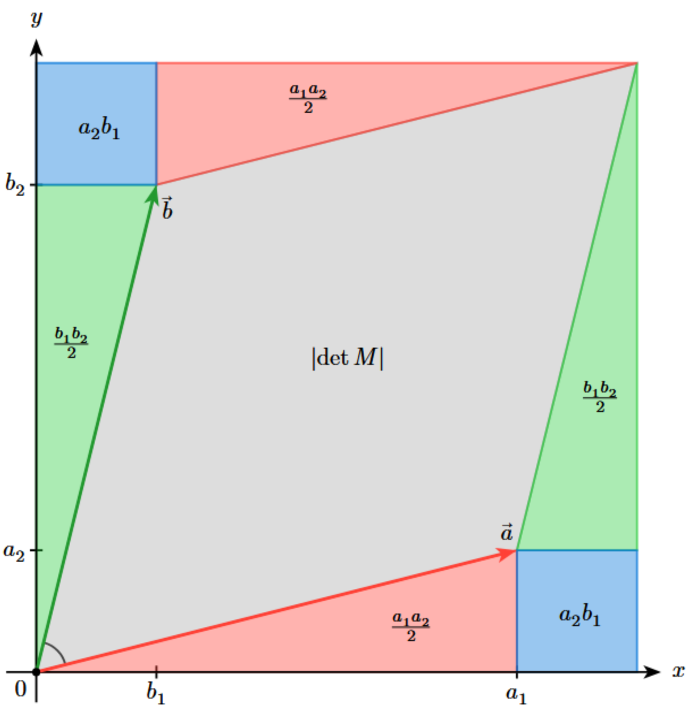
The determinant of this matrix is -1

Where the sign indicates the orientation of the transformation/ arrangement of the column vectors.

This allows the formulation of the idea that the determinant of a matrix is based on the products of the elements along the leading diagonal, although there are more considerations to be made with this idea.



Consulting the diagram below allows you to see how the determinant can be calculated in 2 dimensions/ 2 by 2 matrices.



The vector ***a*** has the x coordinate of a_1 ,and the y coordinate of a_2. Similarly, the vector ***b*** has the x coordinate of b_1 and the y coordinate of b_2. If you imagine another vector that's the sum of the two vectors, spanning the diagonal, it will have an x coordinate of $a_1 + b_1$ and a y coordinate of $a_2 + b_2$. To get the determinant, you must subtract the coloured regions on the diagram from the product of the x and y coordinates of the summed vectors.

$$(a_1+b_1)(a_2+b_2) - 2((a_1a_2)/2) + ((b_1b_2)/2) + (a_2b_1))$$

$$= a_1a_2+a_1b_2+b_1a_2+b_1b_2 - a_1a_2 - b_1b_2 - 2(a_2b_1)$$

$$= a_1b_2 + b_1a_2 - 2(a_2b_1)$$

$$= \underline{a_1b_2 - a_2b_1}$$

for the matrix: $\begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix}$

therefore: $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$

This allows for the consideration that if either b or c are 0, then the determinant of a 2x2 matrix is the product of the terms along its leading diagonal. Fortunately, this is scalable to higher dimensions as well, given that they are in row echelon form/upper triangular (relating to the c term being 0)

$$\begin{vmatrix} 2 & 1 & 3 \\ 0 & -1 & 2 \\ 0 & 0 & 4 \end{vmatrix} = 2 \times -1 \times 4 = -8$$

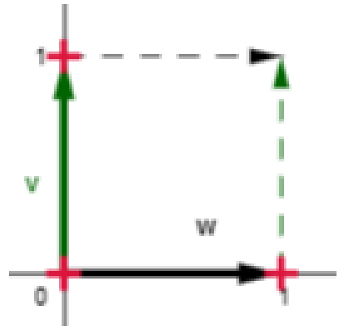
It is important to remember, the area of any shape is scaled by the determinant of the matrix that transform it, as is expressed in the further maths curriculum.

Cramer's rule

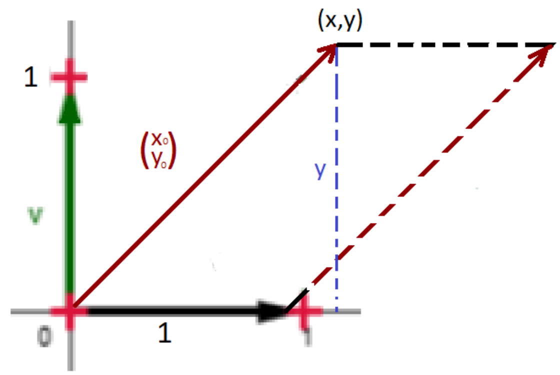
Cramer's rule utilizes determinants and scaled areas to find solutions to systems of linear equations.

Consider the transformation: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$

The matrix can be represented by:



If you imagine the vector $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ on the diagram as this:



Then you can consider the area enclosed by the unit vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and the general vector $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ to be 1 times y_0, or just y_0.

This can be seen by taking the determinant of the matrix: $\begin{bmatrix} 1 & x_0 \\ 0 & y_0 \end{bmatrix}$

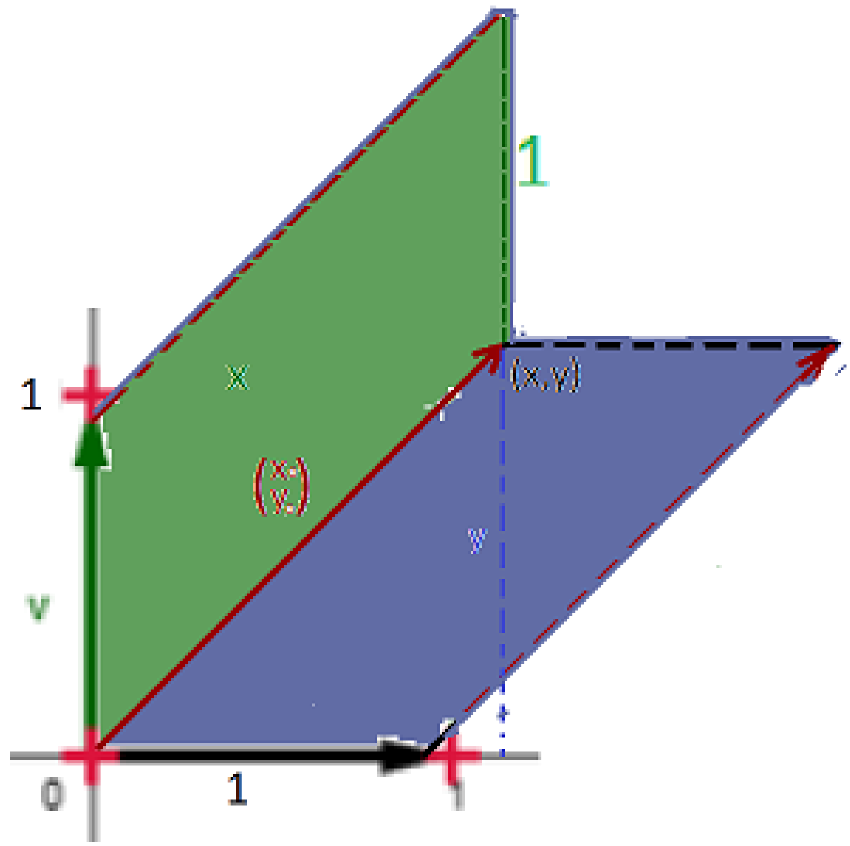
The same course of action can be taken to get an expression for x as the determinant of the matrix : $\begin{bmatrix} x_0 & 0 \\ y_0 & 1 \end{bmatrix}$.

This is useful because we know that the determinant of a matrix is the scaling factor of its transformation, as such if we take the system:

$$\begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$$

And consider the area enclosed by the general vector and the unit vectors, i and j, (giving expressions for y and x as seen in the example above), we can see that the area enclosed by the column vectors in the matrix and the output vector, will be the area enclosed by the general vector and the unit vectors scaled by the determinant of the transforming matrix. This may sound strange but this example may help you understand.

Using the system defined previously, we can represent y and x as the areas in blue and green.



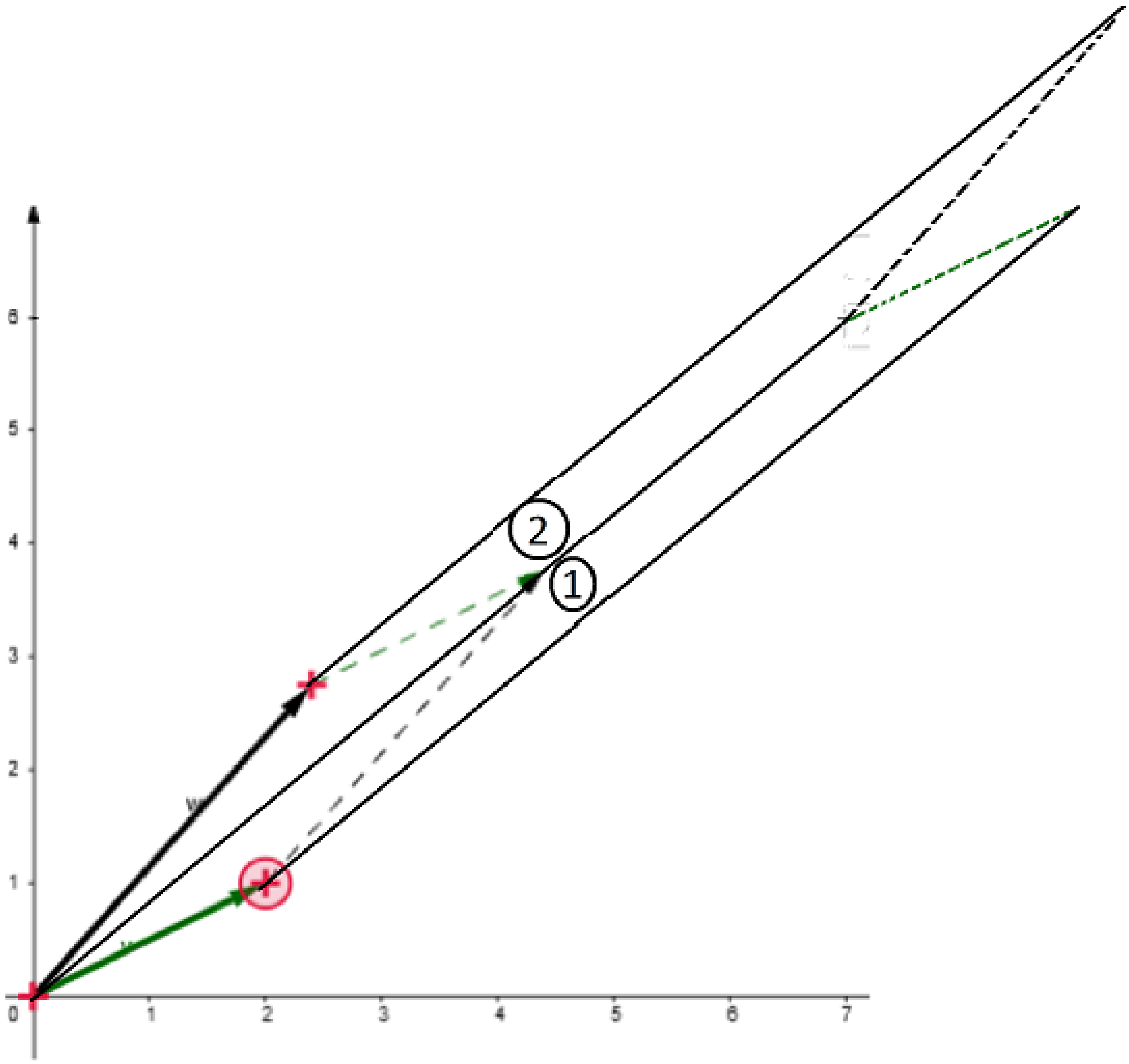
This has been attained by taking the determinants of the matrices:

And $\begin{bmatrix} 1 & x_0 \\ 0 & y_0 \end{bmatrix}$ $\begin{bmatrix} x_0 & 0 \\ y_0 & 1 \end{bmatrix}$

Where y=$\begin{vmatrix} 1 & x \\ 0 & y \end{vmatrix}$and x=$\begin{vmatrix} x & 0 \\ y & 1 \end{vmatrix}$

It is important to remember that these expressions are with respect to the identity matrix.

Once the transformation by the matrix $\begin{bmatrix} 3 & 2 \\ 4 & 1 \end{bmatrix}$ is applied, then the expressions for y and x are scaled by the determinant of this matrix, this is because the original unit vectors are mapped onto the column vectors of the matrix.



Note how the vector for the <u>second column</u> of the transforming matrix, the green one, is now "under" the first column vector, this just means we have to consider the area enclosed by the parallelogram of the green vector and the output vector to be related to the expression that was a result of the <u>second column</u> in the identity matrix, meaning that, the area enclosed and labelled 1, is going to give us the value for x and vice versa. Secondly the areas labelled 1 and 2 will be found by taking the determinant of the matrices $\begin{bmatrix} 7 & 2 \\ 6 & 1 \end{bmatrix}$ and $\begin{bmatrix} 3 & 7 \\ 4 & 6 \end{bmatrix}$

Where $\begin{bmatrix} 7 \\ 6 \end{bmatrix}$ is the output vector. We can say that the area 1, is x times the determinant of the matrix applying the transformation due to how determinants work as scaling factors. Similarly, we can say that the area 2 is y times the determinant of the matrix. So we can easily find values for x and y.

$$\begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix} x = \begin{vmatrix} 7 & 2 \\ 6 & 1 \end{vmatrix} \quad \begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix} y = \begin{vmatrix} 3 & 7 \\ 4 & 6 \end{vmatrix}$$

$$x = \frac{\begin{vmatrix} 7 & 2 \\ 6 & 1 \end{vmatrix}}{\begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix}} \quad y = \frac{\begin{vmatrix} 3 & 7 \\ 4 & 6 \end{vmatrix}}{\begin{vmatrix} 3 & 2 \\ 4 & 1 \end{vmatrix}}$$

$$x = \frac{-5}{-5} \; y = \frac{-10}{-5}$$

$$x = 1 \; y = 2$$

Conclusion

I hope this essay added some greater context to some methods of solving linear systems using matrices. Admittedly, not everything is covered in this essay and many parts are crude, do forgive me. Thank you for reading.