Matrices and Linear Algebra

Contents

1 Linear Algebra

We must start our study of linear algebra with a relatively abstract set of definitions of vectors. When we see how they are implemented with the use of matrices, their usefulness will become apparent and we will see the standard idea of a vector as having 'magnitude and direction' emerge.

1.1 Vectors and Vector Space

We will start with the definition of a vector and a vector space. These ideas will be quite abstract at the start, then we will relate them to the common idea of a vector we understand from geometry.

1.1.1 Definition of a Vector Space

A vector space V is a collection of mathematical objects, vectors, which can be subject to certain arithmetical operations in conjunction with scalars, which for our purposes will be the set of real numbers or occasionally, complex numbers. These arithmetical operations are

- Vector addition, a mapping which takes two vectors and returns one unique vector.
- Multiplication (or rescaling) by a scalar, which takes a vector and a scalar and returns a unique vector.

These are the formal ideas behind the concepts of adding two vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ in V:

$$\boldsymbol{v}_1, \boldsymbol{v}_2 \rightarrow \boldsymbol{v}_1 + \boldsymbol{v}_2,$$

and multiplication by a scalar:

$$\boldsymbol{v}_1, a \rightarrow a\boldsymbol{v}_1, \text{ where } a \text{ is a real number.}$$

The two operations are combined in a *linear combination*, where a list of n vectors and n scalars produce a vector:

$$\{a_1, a_2, \ldots\ a_n\}, \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\ \boldsymbol{v}_n\} \rightarrow \textstyle\sum_j a_j\boldsymbol{v}_j.$$

To define a vector space, these operations will follow the axioms listed here.

1. The addition must be associative:

$$(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$$

2. The addition must be commutative:

$$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}.$$

3. There is an additive identity **0**:

$$\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}.$$

4. There is an additive inverse $-\boldsymbol{v}$:

$$\boldsymbol{v} + -\boldsymbol{v} = \mathbf{0}.$$

5. There is a multiplicative identity for the scalars 1:

$$1\boldsymbol{v} = \boldsymbol{v}.$$

6. Associatively with the scalar multiplication:

$$(ab)\boldsymbol{v} = a(b\boldsymbol{v}).$$

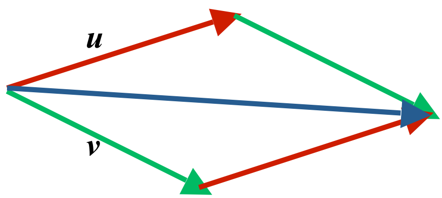
7. Distributivity with the scalar:

$$(a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v},$$

$$a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}.$$

1.1.2 Example – ‘Arrows’ and the Parallelogram Law

The first concept most students of mathematics encounter to represent vectors is the idea of ‘arrows’, in other words a representation of an object with a direction and a magnitude or length. The addition of two vectors, in other words the construction of the resultant, is done with the Parallelogram law as shown.



The red arrows are all equivalent to vector $\boldsymbol{u}$ and the green arrows are all equivalent to vector $\boldsymbol{v}$, the parallelogram law gives the resultant $\boldsymbol{u} + \boldsymbol{v}$, shown in blue.

This algorithm or set of steps gives a well-defined sum of two vectors which is also a vector. We can also define the multiplication of a vector by a scalar:

- $\boldsymbol{v} + \boldsymbol{v} = 2\boldsymbol{v}$, so $\boldsymbol{v} + 2\boldsymbol{v} = 3\boldsymbol{v}$, now any positive integer multiple of $\boldsymbol{v}$ is defined,
- We know that the additive inverse vector $-\boldsymbol{v}$ must exist, from the definition of a vector space. We can therefore set up negative multiples of $\boldsymbol{v}$ by defining $(-1)\boldsymbol{v} = -\boldsymbol{v}$, so any negative integer multiple of $\boldsymbol{v}$ is a sum of these terms.

- To find $\boldsymbol{v}$ times the rational number p/q, then construct $p\boldsymbol{v}$ as before then find the vector $q\boldsymbol{w}$ such that $q\boldsymbol{w} = p\boldsymbol{v}$. Then $\boldsymbol{w}$ is the required vector.
- A further, more abstract, set of mathematical steps will produce the general real number a multiplying a vector $\boldsymbol{v}$ to give the rescaled vector $a\boldsymbol{v}$.

1.1.3 Example – Paired Numbers, R^2

Consider the set R^2, the set of all ordered pairs of real numbers. This means every possible pair of real numbers (x, y) where

$$(x, y) \neq (y, x).$$

The following are the required operations:

1. Addition of vectors: $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$.
2. Multiplication of scalars: $a(x, y) = (ax, ay)$.

These operations satisfy the axioms listed above due to the commutative and associative properties of addition and multiplication of real numbers and therefore the set R^2 is a vector space. Clearly, it is the well-understood idea of two-dimensional space with Cartesian coordinates (x, y).

1.1.4 Definition of Linear Independence

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\ \boldsymbol{v}_n$ be a set of vectors in a vector space V. They are said to be linearly independent if there are no list of scalars $a_1, a_2, \ldots\ a_n$ such that

$$\sum_j a_j\boldsymbol{v}_j = \mathbf{0}.$$

This may be interpreted to mean that the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\ \boldsymbol{v}_n$ are 'pointing in different directions' in the vector space.

The immediate corollary to this is that if a list of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\ \boldsymbol{v}_n$ are linearly independent and there is a list of scalars $a_1, a_2, \ldots\ a_n$ such that

$$\sum_j a_j\boldsymbol{v}_j = \mathbf{0},$$

then $a_i = 0$ for every i.

The paired numbers provide a clear example; the pair of vectors (1, 0) and (0, 1) are linearly independent since

$$a(1, 0) + b(0, 1) = (a, b).$$

If this is to be (0, 0) then the only way this can happen is if $a = 0$ and $b = 0$.

1.1.5 Basis Vectors

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\ \boldsymbol{v}_n$ be a set of vectors in a vector space V. Let W be the subset of V made up of every possible linear combination of these vectors, so

$$W = \{w \mid w = \textstyle\sum_j a_j\boldsymbol{v}_j, \text{ for any list of scalars } a_1, a_2, \ldots\ a_n \}$$

The vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\ \boldsymbol{v}_n$ are said to generate W.

If it is the case for a particular list of vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots\ \boldsymbol{e}_n$ that they generate all of V and the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots\ \boldsymbol{e}_n$, are linearly independent, then it can be shown that for a given vector $\boldsymbol{v}$ in V there is a unique set of scalars $a_1, a_2, \ldots\ a_n$ such that

$$\boldsymbol{v} = \sum_j a_j \boldsymbol{e}_j,$$

then the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots \boldsymbol{e}_n$, are said to be a basis for V. The scalars $a_1, a_2, \ldots a_n$ for the vector $\boldsymbol{v}$ are said to be its coordinates or components.

We saw already that for the set of paired numbers R^2, the pair of vectors (1, 0) and (0, 1) are linearly independent. These now form a (canonical) basis for R^2 since, for any paired number (a, b),

$$(a, b) = a(1, 0) + b(0, 1).$$

The numbers a and b are the coordinates of (a, b).

1.1.6 Dimension of a Vector Space

It can be shown that every vector space has a basis and furthermore every vector space has a basis with the minimum number of elements/vectors. This is called the dimension of the vector space. These results can be readily proved if the dimension is finite.

A vector space generated by a set of vectors will have the dimension of the smallest number of linearly independent vectors found in this set. If the set is finite, this result can be readily proved with the ideas discussed so far.

Once we have established the ideas of dimensions, bases and a vector as the sum of a set of basis vectors, we have the traditional notion of the vector as an object with a magnitude and direction. For example, in the two dimensional case they are traditionally represented as **i** and **j**. Thus any two-dimensional vector $\boldsymbol{v}$ can be represented as

$$\boldsymbol{v} = a\mathbf{i} + b\mathbf{j}.$$

As we saw before, the numbers (a, b) are the coordinates of the vector $\boldsymbol{v}$ with respect to the basis **i** and **j**.

1.2 Euclidean Geometry

We still retain the idea of a vector being a geometric or even Euclidean idea, a magnitude and a particular direction. We will now introduce some definitions which bring in angles and distances into our concept of a vector. These are the familiar ideas of Euclidean geometry. We will start with the idea of a length.

1.2.1 Definition – Length

For a vector space V, the length of a vector $\boldsymbol{v}$, denoted $||\boldsymbol{v}||$, is a function that satisfies the following properties.

1. The function is real-valued, so $||\boldsymbol{v}||$ is always a real number.
2. The function is always positive, so $||\boldsymbol{v}|| > 0$.
3. The zero vector has length 0 and it is the only vector with this property; $||\boldsymbol{v}|| = 0$ if and only if $\boldsymbol{v} = \mathbf{0}$.
4. Multiplying a vector by a real scalar a re-scales the length of the vector by the magnitude of a: $||a\boldsymbol{v}|| = |a|\ ||\boldsymbol{v}||$.
5. The triangle inequality; $||\boldsymbol{v} + \boldsymbol{w}|| \leq ||\boldsymbol{v}|| + ||\boldsymbol{w}||$. We will shortly see that this means that the distance from point A through B

to C is never shorter than going directly from A to C, or the shortest distance between any two points is a straight line.

Mathematically, a function with these properties is called a *norm*, but for our purposes the word ‘length’ will suffice, as this is what we will be using it for.

1.2.2 Example – Dimension, Basis and Length in R^n

For the real-valued ordered pairs of n numbers, we define the following objects.

1. The typical element of R^n is
$$(x_1, x_2, \ldots x_n)$$
2. The basis of the set R^n is the set of vectors
$$\mathbf{e}_1 = (1, 0, \ldots 0),\ \mathbf{e}_2 = (0, 1, 0, \ldots 0),\ \ldots\ \mathbf{e}_n = (0, \ldots 0, 1).$$
This means that $(x_1, x_2, \ldots x_n) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n$.
3. The length of the vector $(x_1, x_2, \ldots x_n)$ is the Pythagorean length, defined as the positive square root of the sum of the squares of the individual values:
$$\| (x_1, x_2, \ldots x_n) \|^2 = x_1^2 + x_2^2 + \ldots + x_n^2.$$
This is the standard idea of distance in Euclidean geometry.

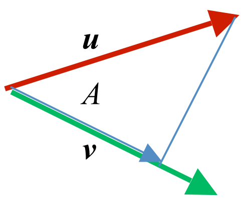
This definition of a length has all the properties outlined above.

1.2.3 Definition – Inner Product

For a vector space V, the inner or scalar product of two vectors, denoted $\langle \boldsymbol{v},\boldsymbol{w} \rangle$, is a function of two vectors where:

1. The function is real-valued, so $<\boldsymbol{v},\boldsymbol{w}>$ is always a real number.
2. The function is symmetric, so $<\boldsymbol{v},\boldsymbol{w}> = <\boldsymbol{w},\boldsymbol{v}>$. [Sometimes an inner product may be defined as complex valued function, in which case $<\boldsymbol{w},\boldsymbol{v}>$ will be the complex conjugate of $<\boldsymbol{v},\boldsymbol{w}>$.]
3. The function is positive definite, so $<\boldsymbol{v},\boldsymbol{v}> \geq 0$.
4. The zero vector has inner product 0 with itself and it is the only vector with this property; $<\boldsymbol{v},\boldsymbol{v}> = 0$ if and only if $\boldsymbol{v} = \mathbf{0}$.
5. The inner product is linear in its vector arguments; for a real scalar a, $<a\boldsymbol{v},\boldsymbol{w}> = a<\boldsymbol{v},\boldsymbol{w}>$ and $<\boldsymbol{v} + \boldsymbol{u},\boldsymbol{w}> = <\boldsymbol{v},\boldsymbol{w}> + <\boldsymbol{u},\boldsymbol{w}>$.

With these properties, an associated length can be defined using the inner product: $||\boldsymbol{v}||^2 = <\boldsymbol{v},\boldsymbol{v}>$, as it satisfies all the definitions of a lengths as set out in the previous section. These two ideas now lead to the concept of an angle. Once we have defined the inner product, we can regard it as the projection of one vector onto the other. Consider two vectors ***u*** and ***v*** as shown in the diagram here.



The light blue line joins the end of ***u*** to ***v*** at the closest point and therefore makes a right angle with ***v***. The light blue vector is the component of ***u*** in the direction of ***v***, that is, it is in the same direction as ***v*** but only to the point closest distance to ***u***.

The quantity A measures the 'gap' between the two vectors; it is the angle.

This clearly is a symmetric construction. Since this quantity is 0 if the angle is a right angle ($\pi/2$ radians) and is simply the product of the lengths if the angle is 0, then we may regard it as a cosine of an angle. Therefore

$$<\boldsymbol{v},\boldsymbol{w}> = ||\boldsymbol{v}||.||\boldsymbol{w}|| \cos(A).$$

We may view this identity as defining the cosine:

$$\cos(A) = <\boldsymbol{v},\boldsymbol{w}>/||\boldsymbol{v}||.||\boldsymbol{w}|| .$$

Since the length here is defined by the inner product, then

$$\cos(A) = <\boldsymbol{v},\boldsymbol{w}>/(\sqrt{<\boldsymbol{v},\boldsymbol{v}>}\sqrt{<\boldsymbol{w},\boldsymbol{w}>}).$$

Thus with the inner product, both length and angle follow. The next definition follows form this idea.

1.2.4 Definition – Orthogonal Vectors

Let $\boldsymbol{v}$, $\boldsymbol{w}$ be two vectors in a vector space with a defined scalar product and associated definitions of length and angle. If $<\boldsymbol{v},\boldsymbol{w}> = 0$, then the two vectors are said to be orthogonal. For Euclidean geometry and vector spaces, this means that the angle between them is $\pi/2$ radians, a right angle.

1.2.5 Orthonormal Bases

Let $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, … $\boldsymbol{e}_n$ be a basis set of vectors for a vector space V. This basis set is called Orthonormal if it has the property that $<\boldsymbol{e}_i, \boldsymbol{e}_j> = 1$ if $i = j$ and $<\boldsymbol{e}_i, \boldsymbol{e}_j> = 0$ if $i \neq j$. The first part of this statement means that $||\boldsymbol{e}_i|| = \sqrt{<\boldsymbol{e}_i, \boldsymbol{e}_i>} = 1$ for every $\boldsymbol{e}_i$.

For an orthonormal basis, this means the coordinates of a given vector with respect to the basis are readily found using the scalar product. Recall that for a given vector $\boldsymbol{v}$ in V there is a unique set of scalars $a_1, a_2, \ldots\ a_n$ such that

$$\boldsymbol{v} = \sum_j a_j \boldsymbol{e}_j.$$

Now take the scalar product of $\boldsymbol{v}$ with one of the basis vectors $\boldsymbol{e}_k$ and see how this works out using its decomposition in terms of the basis vectors and taking advantage of the linearity of the scalar product:

$$<\boldsymbol{e}_k, \boldsymbol{v}> = <\boldsymbol{e}_k, \sum_j a_j \boldsymbol{e}_j> = \sum_j a_j <\boldsymbol{e}_k, \boldsymbol{e}_j>.$$

Now $<\boldsymbol{e}_k, \boldsymbol{e}_j> = 0$ for all values of j except the case $j = k$, when it is 1, so this means that

$$<\boldsymbol{e}_k, \boldsymbol{v}> = a_k.$$

Therefore a vector may be decomposed in terms of an orthonormal basis as

$$\boldsymbol{v} = \sum_j <\boldsymbol{e}_j, \boldsymbol{v}>\, \boldsymbol{e}_j = \sum_j <\boldsymbol{v}, \boldsymbol{e}_k>\, \boldsymbol{e}_j.$$

The classic example of this type of basis is the unit vector basis for the set of vectors as ‘arrows’ with direction and magnitude in the plane, represented as **i** and **j**. Thus any two-dimensional vector $\boldsymbol{v}$ can be represented as

$$\boldsymbol{v} = a\mathbf{i} + b\mathbf{j}.$$

As we saw before, the numbers (a, b) are the coordinates of the vector $\boldsymbol{v}$ with respect to the basis **i** and **j**. Then

$$<\mathbf{i}, \mathbf{i}> = 1, <\mathbf{j}, \mathbf{j}> = 1, <\mathbf{i}, \mathbf{j}> = 0.$$

Then for vector $\boldsymbol{v} = a\mathbf{i} + b\mathbf{j}$, the values of a and b can be found by using the relation established above:

$$a = \langle \boldsymbol{v}, \mathbf{i} \rangle,\ b = \langle \boldsymbol{v}, \mathbf{j} \rangle.$$

1.3 Linear Maps

A linear map is a function from one vector space to another which has the following properties. Let f be the linear map, from vector space V to vector space W. then:

1. For $f: V \rightarrow W$, $\boldsymbol{v} \in V$, then $f(a\boldsymbol{v}) = af(\boldsymbol{v})$
2. For $f: V \rightarrow W$, $\boldsymbol{v}, \boldsymbol{w} \in V$, then $f(\boldsymbol{v} + \boldsymbol{w}) = f(\boldsymbol{v}) + f(\boldsymbol{w})$

An immediate consequence of these properties is that any such function f is a bijection, that is, $f(\boldsymbol{v}) = f(\boldsymbol{w})$ if and only if $\boldsymbol{v} = \boldsymbol{w}$. The linearity property of a function f will also preserve many properties of the functions of vectors we have encountered or will change them in straightforward ways. In particular this includes the inner product and the length, which will be rescaled in a very straightforward fashion. This concept now leads to the definition of an isomorphism.

1.3.1 Definition - Isomorphism

Let f be the linear map

$$f: V \rightarrow W.$$

If there is another map g such that

1. $g(f(\boldsymbol{v})) = \boldsymbol{v}$, for any $\boldsymbol{v} \in V$,
2. $f(g(\boldsymbol{w})) = \boldsymbol{w}$, for any $\boldsymbol{w} \in W$,

then f and g are said to be inverses of each other and the maps f and g are said to be isomorphisms. Essentially this means that the two vector spaces V and W are identical. In particular, if one vector space has an inner product and associated length, then the other will have one yielding the same values for vectors related by the isomorphism. Here is an example that is almost trivial but is very important.

1.3.2 Example – R^3 and 3D Arrows

Consider the set R^3, the set of all ordered triplets of real numbers. This means every possible triplet of real numbers (x, y, z) where

$$(x, y, z) \neq (y, x, z) \text{ or any other reordering of the triplet.}$$

The following are the required operations:

3. Addition of vectors: $(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$.
4. Multiplication of scalars: $a(x, y, z) = (ax, ay, az)$.

These operations satisfy the axioms listed above due to the commutative and associative properties of addition and multiplication of real numbers and therefore the set R^3 is a vector space. Clearly, it is the well-understood idea of three-dimensional space with Cartesian coordinates (x, y, z).

Now consider the set V_3 of arrows in three dimensions, with orthonormal basis traditionally represented as **i, j** and **k**. Thus any three-dimensional vector $\boldsymbol{v}$ can be represented as

$$\boldsymbol{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

As we saw before, the numbers (a, b, c) are the coordinates of the vector $\boldsymbol{v}$ with respect to the basis **i, j** and **k**. This immediately sets up a canonical isomorphism

$$f: V \rightarrow \mathrm{R}^3, \text{ where } f(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) = (a, b, c).$$

thus the act of identifying the three coordinates of the elements of V_3 sets up an isomorphism with R^3. Note the link in this example between the inner products. The inner product of the vectors shown here are:

$$\boldsymbol{v}_1 = a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k},\ \boldsymbol{v}_2 = a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}.$$

The inner product for this vector space is defined as:

$$||\boldsymbol{v}_1||\ ||\boldsymbol{v}_2|| \cos(A),$$

where A is the angle between the vectors. Since the basis vectors are orthonormal and using the linearity of the inner product, it follows that

$$<\boldsymbol{v}_1, \boldsymbol{v}_2> = a_1a_2 + b_1b_2 + c_1c_2.$$

This leads to the concept of the inner product in R^3 defined as

$$<(a_1, b_1, c_1), (a_2, b_2, c_2)> = a_1a_2 + b_1b_2 + c_1c_2.$$

1.3.3 Representing a Linear Map

Consider now a linear function mapping vectors from one vector space V to another W. This means a function T such that for $\boldsymbol{v}, \boldsymbol{w} \in V$ and $a, b \in \mathrm{R}$, then

$$T(a\boldsymbol{v} + b\boldsymbol{w}) = aT(\boldsymbol{v}) + bT(\boldsymbol{w}).$$

These vectors are all in W.

This is a version for vectors of the general idea of a linear function. The action of a linear functions is usually represented as $T\boldsymbol{v}$ to better illustrate this property. Thus the linearity is represented as

$$T(a\boldsymbol{v} + b\boldsymbol{w}) = aT\boldsymbol{v} + bT\boldsymbol{w}.$$

This type of function is important in applied mathematics since very many simple geometric transformations such as rotations, some translations and rescaling operations have this property. We will see how it may be represented numerically in terms of the coordinates of a vector.

Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots \boldsymbol{e}_n$ be an orthonormal basis set of vectors for the vector space V and let $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots \boldsymbol{f}_n$ be the same for W. Let $\boldsymbol{v} \in V$, with coordinates $a_1, a_2, \ldots a_n$ so that

$$\boldsymbol{v} = \sum_j a_j\boldsymbol{e}_j.$$

We will look at the coordinates of $T\boldsymbol{v}$. Let us call these coordinates $b_1, b_2, \ldots b_n$, so that

$$T\boldsymbol{v} = \sum_j b_j\boldsymbol{f}_j.$$

The i-th coordinate will be given by

$$b_i = <\boldsymbol{f}_i, T\boldsymbol{v}> = <\boldsymbol{f}_i, T\sum_j a_j\boldsymbol{e}_j> = \sum_j a_j<\boldsymbol{f}_i, T\boldsymbol{e}_j>.$$

Therefore if we set

$$T_{ij} = <\boldsymbol{f}_i, T\boldsymbol{e}_j>, \text{ then } b_i = \sum_j T_{ij}a_j.$$

Some observations:

1. The number T_{ij} is the i-th component of the vector $T\boldsymbol{e}_j$ in the orthonormal basis $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots \boldsymbol{f}_n$. Therefore the coordinates b_i

are entirely specified by the effect of T on the orthonormal basis.

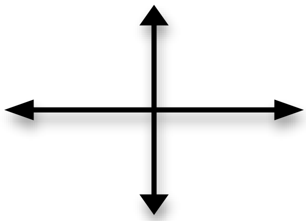
2. The equation for the coordinates of $T\boldsymbol{v}$ in terms of those of $\boldsymbol{v}$, $b_i = \sum_j T_{ij}a_j$, is a linear combination and sets the scene for matrix algebra. The numbers T_{ij} will be the matrix representation of transformation T.
3. If the mapping T goes from V to V, the numbers T_{ij} will be defined as $T_{ij} = \langle \boldsymbol{e}_i, T\boldsymbol{e}_j \rangle$. It may be possible that the space generated by the vectors $T\boldsymbol{e}_j$ may be of lower dimension than V.

1.3.4 Rotation in 3D space

Recall the set V_3 of arrows in three dimensions, with orthonormal basis **i, j** and **k**. Consider a rotation in V_3 where any vector $\boldsymbol{v}$ is turned through an angle A with the vertical axis, that is, the direction **k**, as the axis of rotation. Let R be the symbol for this rotation. We will identify the values R_{ij} and so we will be able to calculate the coordinates of any vector transformed by this rotation.

Before embarking on this exercise, care must be taken on the orientation of the axes in three-dimensional space. There are two possibilities, distinguished by our view of the x-y plane shown in the figure; the positive z-axis could be pointing into the page or out of the page.

Positive y-axis



Negative y-axis

The positive z-axis may be pointing out of the page or into the page.

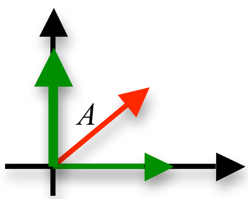
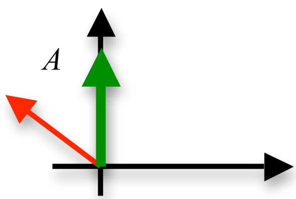
Positive x-axis

The convention is that it points out of the page.

This is known as the right-hand rule; the x-axis is the first finger, the y-axis is the second and the z-axis is the thumb.

With this convention, if a vector representing a position is rotated around the z-axis by an angle A, as measured from the positive x-axis, then the following happens to the orthonormal basis **i, j** and **k**:

1. **k** is mapped to **k**.
2. **i** is mapped to the vector **i**cos(A) + **j**sin(A).
3. **j** is mapped to –**i**sin(A) + **j**cos(A).

	
Here, the vector **i** (green) is rotated through angle A to give	Here, the vector **j** (green) is rotated through angle A to give

the unit vector in red.	the unit vector in red.

This means we have all nine values of the numbers R_{ij}. Let us represent the orthonormal basis as vectors in R^3:

$$(1, 0, 0), (0, 1, 0), (0, 0, 1).$$

Then the image of these three vectors was

$$(\cos(A), \sin(A), 0), (\sin(A), \cos(A), 0), (0, 0, 1).$$

2 Matrices

The vector space R^n is canonically used to represent the coordinates of a vector in an n-dimensional space with respect to an orthonormal basis. This leads to a natural representation of the numbers T_{ij} which determine the action of the transformation T. These numbers will be the elements in the matrix representation of T. The properties of matrices will now be outlined; all properties seen will derive directly from what is required to be consistent with the idea of a linear transformation. Once this representation is established, it will allow us to carry out actual uses of linear algebra, which have wide applications where a linear system is being studied, or where large numbers of linear equations arise. Following the introduction of these ideas, this section covers the following topics.

- The definition of a matrix
- Matrix algebra: addition, multiplication by scalars, matrix multiplication.

- The transpose of a matrix and symmetry.
- The idea of a determinant for 2 by 2 matrices
- The idea of an inverse for 2 by 2 matrices

The first few topics are naturally the most important for studying matrices.

2.1 Definition – a Matrix

A matrix consists of a set of numerical values arranged in a rectangular grid, surrounded by one set of brackets. The numbers are not separated by commas or individual brackets. If a matrix has *r* rows and *c* columns, then it is said to be an *r* by *c* matrix.

The numbers *n* by *m* describe the size of the matrix. The matrix *C* here is a 4 by 5 matrix:

$$C = \begin{pmatrix} 1 & 2 & 9 & -3 & 0 \\ 0 & -3 & 6 & 10 & 2 \\ -1 & 4 & 5 & 3 & -1 \\ 7 & 5 & -2 & 1 & 4 \end{pmatrix}.$$

The individual numbers within a matrix are called the elements of the matrix. For example, among the elements of *C* are:

$$C_{22} = -3,\ C_{32} = 6,\ C_{35} = -1.$$

We speak of rows and columns in the natural way based on this way to present the information C_{ij}. If there is just one column, we have a vector.

2.1.1 Definition – Column Matrix, Vector

A column matrix is a matrix of size n by 1, where n is a positive integer. This is also known as a vector.

The reason for the use of this term is that an n by 1 column matrix can be taken as a point in n-dimensional space, so for example a 3 by 1 column matrix can be taken as a point in the three dimensional $X - Y - Z$ space. The concept of the coordinates $(x_1, x_2, \ldots x_n,)$ of a vector will now be represented as a column matrix as shown:

$$\begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}.$$

2.2 The Algebra of Matrices

Matrices have a well-defined algebra, that is, definitions for addition, multiplications and so on, which we will explain here. These ideas are based on the idea that they represent a transformation:

- They can always be multiplied by scalars.
- They can be added, subtracted and multiplied by other matrices, given certain conditions on their size.
- Given stricter conditions, they can be inverted, and so one matrix can be 'divided' by another.

2.2.1 Addition of matrices

As might be expected, two matrices are added by adding each of the corresponding elements. This means that two matrices can only be added if they are of identical size. We will establish addition by defining it as follows. Let A and B be two matrices of size n by m. Let a_{ij} be the elements of A and similarly b_{ij} for B. Define the numbers c_{ij} such that

$$c_{ij} = a_{ij} + b_{ij}, \text{ for } 1 \le i \le n \text{ and } 1 \le j \le m.$$

Let C be the matrix of size n by m with elements c_{ij}. Then we say that

$$C \text{ is of size } n \text{ by } m \text{ and } C = A + B.$$

For example, the following two matrices can be added, since both are 2 by 2 matrices:

$$A = \begin{pmatrix} 1 & 3 \\ -5 & 3 \end{pmatrix}, \ B = \begin{pmatrix} -1 & -2 \\ 4 & -3 \end{pmatrix}.$$

The result is

$$A + B = \begin{pmatrix} 1+(-1) & 3+(-2) \\ -5+4 & 3+(-3) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In the equation above, it is shown explicitly that each of the corresponding elements has been added. Subtraction is done in an analogous way.

$$A - B = \begin{pmatrix} 1-(-1) & 3-(-2) \\ -5-4 & 3-(-3) \end{pmatrix} = \begin{pmatrix} 2 & 5 \\ -9 & 6 \end{pmatrix}.$$

2.2.2 Multiplication by a scalar

A matrix is multiplied by a scalar in the obvious way; by multiplying each element of the matrix by the scalar. More formally: let k be a scalar and let A be a matrix of size n by m with elements a_{ij}. Define the numbers b_{ij} such that

$$b_{ij} = ka_{ij}, \text{ for } 1 \leq i \leq n \text{ and } 1 \leq j \leq m.$$

Let B be the matrix of size n by m with elements b_{ij}. Then we say that

$$B \text{ is of size } n \text{ by } m \text{ and } B = kA.$$

For example, set matrix A to be

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

Then multiplying A by the number 5 gives:

$$5A = \begin{pmatrix} 5 & 15 \\ -25 & 15 \end{pmatrix}.$$

2.3 Matrix Multiplication

We have seen already that a linear mapping of a vector leads to a particular form of calculation for the coordinates of the image of that vector. We restate it here.

Let $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots\ \boldsymbol{e}_n$ be an orthonormal basis set of vectors for the vector space V and let $\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots\ \boldsymbol{f}_m$ be the same for W. Take note that vector spaces V and W have dimensions n and m respectively. Let $\boldsymbol{a} \in V$, with coordinates $a_1, a_2, \ldots\ a_n$ so that

$$\boldsymbol{a} = \textstyle\sum_j a_j\boldsymbol{e}_j.$$

Let vector $\boldsymbol{b} = T\boldsymbol{a}$ have coordinates $b_1, b_2, \ldots\ b_m$, so that

$$\boldsymbol{b} = \textstyle\sum_j b_j \boldsymbol{f}_j,$$

then the coordinates of $\boldsymbol{b}$ are calculated from those of $\boldsymbol{a}$ as:

$$b_i = \textstyle\sum_j T_{ij} a_j.$$

Consider now representing the vector $\boldsymbol{a} = \sum_j a_j \boldsymbol{e}_j$ as an n by 1 column matrix and $\boldsymbol{b}$ as an m by 1 column matrix. If we define an m by n matrix T with elements T_{ij}, then the calculation of each element b_j may be viewed as each element of a row of T multiplied by the corresponding element in the column matrix $\boldsymbol{w}$ counted down. Therefore this matrix multiplication may be viewed as the described action on these matrices:

$$\begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} & \cdots & T_{1m} \\ T_{21} & T_{22} & \cdots & T_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n1} & T_{n2} & \cdots & T_{nm} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix}.$$

This concept is generalised into the following definition of matrix multiplication, which is also consistent with the calculations done when one linear mapping follows another.

2.3.1 Definition of Matrix Multiplication

Let A be an m by n matrix with elements a_{ij} and let B be n by l matrix with elements b_{ij}. Define the matrix C, of size m by l, as the matrix with elements given by

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

Then we say that C is the matrix product of A and B:

$$C = AB.$$

The practical steps in multiplying two matrices, say A and B, are summarised here. Let A be an m by n matrix; the number of columns in A must be the same as the number of rows in B, so if the multiplication is to happen B must be an n by k matrix.

The number of rows in the product matrix will be the same as those of the first matrix, and the number of columns the same as those of the second matrix. Thus the result will be an m by k matrix. So in summary the rule for multiplication is

$$(m \text{ by } n) \text{ times } (n \text{ by } k) \text{ gives } (m \text{ by } k).$$

It should be noted that if the product AB exists, the product BA may not. For BA to exist, B must be of size n by m. Thus AB and BA may not even be the same size and are therefore not necessarily equal. Thus the normal commutative nature of multiplication will not apply to matrices.

2.3.2 An Example of Matrix Multiplication

Here are two matrices:

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

We will calculate AB and then BA. Both of these calculations are possible:

AB is a 2 by 3 times a 3 by 2, giving a 2 by 2.

BA is a 3 by 2 times a 2 by 3, giving a 3 by 3.

Here are the two results. Product AB:

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

The details are:

First row from A, first column from B:

$$1x2 + 0x3 + (-1)x(-1) = 2 + 0 + 1 = 3$$

Put this in row 1, column 1 of the answer.

First row from A, second column from B:

$$1x1 + 0x(-4) + (-1)x0 = 1 + 0 + 0 = 1$$

Put this in row 1, column 2 of the answer.

Finished with the first row, move on to the second:

Second row with first column:

$$(-1)x2 + 2x3 + (-2)x(-1) = -2 + 6 + 2 = 6$$

Put this in row 2, column 1 of answer matrix.

Second row with second column:

$$(-1)x1 + 2x(-4) + (-2)x0 = -1 - 8 + 0 = -9$$

Put this in row 2, column 2 of the answer.

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

For BA, the calculations are shown as well:

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

First row with first column:

2x1 + 1x(–1) = 1, put in row 1 column 1.

First row with second column:

2x0 + 1x2 = 2, put in row 1 column 2.

First row with third column:

2x(–1) + 1x(–2) = –2 – 2 = –4, put in row 1 column 3.

Second row with first column:

3x1 + (–4)x(–1) = 7, drop in row 2 column 1.

Second row with second column:

3x0 + (–4)x2 = –8, drop in row 2 column 2.

Second row with third column:

3x(–1) + (–4)x(–2) = –3 + 8 = 5, drop in row 2 column 3.

Third row with first column:

(–1)x1 + 0x(–1) = –1 + 0 = –1, drop in row 3 column 1.

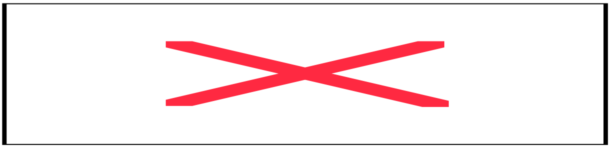
Third row with second column:

(–1)x0 + 0x2 = 0, drop in row 3 column 2.

Third row with third column:

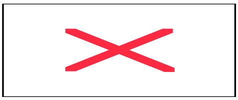
(–1)x(–1) + 0x(–2) = 1 + 0 = 1, drop in row 3 column 3.

Our result is therefore:

 .

2.3.3 The Transpose of a Matrix

When working with a matrix A with elements a_{ij}, it will be useful in many contexts to refer to the matrix elements a_{ji}. Thus the element in position row i and column j is now in row j column i in matrix A^{T}. For an example, consider the following 2 by 3 matrix:

 .

The element in row 1, column 2 (which has the value 0), now sits in row 2, column 1. The rows of this matrix become columns, the columns become rows. For A, the matrix found by doing this is

$$A^T = \begin{pmatrix} 1 & -1 \\ 0 & 2 \\ -1 & -2 \end{pmatrix}.$$

The two matrices here, A and A^{T}, have the converse or opposite size; A is a 2 by 3 matrix and A^{T} is 3 by 2. This means that the products

A by A^{T} and A^{T} by A

are well defined, and will be square matrices; this is true of any matrix.

2.3.4 Definition – Symmetry

A matrix X is said to be symmetric if $X^{\mathrm{T}} = X$.

Here is an example of a symmetric matrix:

$$F = \begin{pmatrix} 1 & -2 & 4 \\ -2 & 0 & 5 \\ 4 & 5 & 9 \end{pmatrix}.$$

Recall the following two matrices:

$$AA^{\mathrm{T}} \text{ and } A^{\mathrm{T}}A$$

It can be readily seen and proven without too much difficulty that both are symmetric for any matrix A.

2.4 The Algebra of Matrices

The arithmetic operations on matrices will have algebraic properties derived directly from their definitions in terms of their elements. We will recap on them here, before moving on to define the inverse, and note how this is similar and different to algebra with scalars.

2.4.1 The Commutative Law for Addition

Clearly when adding two matrices A and B, where it is well defined,

$$A + B = B + A.$$

2.4.2 No Commutative Law for Multiplication

When multiplying two matrices A and B, where this is possible, and where the two products can even be compared, generally

$$AB \neq BA.$$

Therefore if we have a product AB, we specify that A pre-multiplies B or alternatively the matrix B post-multiplies A.

2.4.3 The Distributive Law

For three matrices A, B and C, where the terms in this expression are valid, we can see that

$$(A + B)C = AC + BC.$$

Again, where it is possible to carry out the operations:

$$A(B + C) = AB + AC.$$

This can be proved with the full definition of matrix multiplication supplied above; it essentially relies on the scalar version of the commutative law.

2.4.4 The Associative Law for Multiplication

For three matrices A, B and C, where the terms in this expression are valid,

$$(AB)C = A(BC).$$

This means that when three (or more) matrices are multiplied in sequence, we can unambiguously write

$$ABC,$$

so there is no doubt about the result of the multiplication, irrespective of the order in which the task is done. This result is also proved by using the scalar associative law.

2.4.5 The Transpose and Addition and Multiplication

For matrices A and B, where the terms in this expression are valid, it is clear that the mathematical definition means we can say that

$$(A + B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}.$$

It is less obvious but true nonetheless that

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}.$$

Thus far, we have matrix analogues of addition and multiplication. These came with restrictions on the size of the matrices, increasing as we used more arithmetical operations.

2.5 Some further Definitions

There are certain matrices that come up often in calculations because of their useful properties. The most important example is the identity matrix, so-called because it is the multiplicative identity, the equivalent for matrices of the number 1 in scalars.

2.5.1 Definition – Kronecker's Delta

Let δ_{ij} be a function of two integers with values 0 or 1, given by

$$\delta_{ij} = 1 \text{ if } i = j \text{ and } \delta_{ij} = 0 \text{ otherwise.}$$

This is a very useful item of notation which will arise frequently. It has a very important application in representing the multiplicative identity for matrices, as follows.

2.5.2 Definition – Identity Matrix

The identity matrix is an n by n matrix with elements given by the Kronecker delta. Therefore it is a square matrix with all zero elements off the diagonal and 1's on the diagonal. Since there will be an identity matrix for various values of n, we will distinguish them, where necessary, with a subscript. The 2 by 2 identity matrix is:

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

The 3 by 3 identity matrix is

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

Let X and T be the matrices

$$X = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 2 & -2 \end{pmatrix}, \; T = \begin{pmatrix} 2 & 1 \\ 3 & -4 \\ -1 & 0 \end{pmatrix}.$$

Carrying out the permitted calculations we see that $I_2X = X$:

$$I_2X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \\ -1 & 2 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 2 & -2 \end{pmatrix}.$$

and $TI_2 = T$:

$$TI_2 = \begin{pmatrix} 2 & 1 \\ 3 & -4 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 3 & -4 \\ -1 & 0 \end{pmatrix}.$$

We can multiply X times I_3 to show that $XI_3 = X$:

$$XI_3 = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 2 & -2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 2 & -2 \end{pmatrix}.$$

2.5.3 The Inner Product

Let X and Y be column matrices given by

$$X = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}, \ Y = \begin{pmatrix} 1 \\ -4 \\ 0 \end{pmatrix}.$$

Recall that the inner product of these two vectors is the sum of the products of their elements. Therefore $<X, Y>$ may be written using matrix algebra and notation as $X^T Y$. Then

$$X^T Y = \begin{pmatrix} 2 & 3 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ -4 \\ 0 \end{pmatrix} = -10.$$

2.6 The Inverse of a Matrix

The matrices with the most interesting algebra are square matrices, as any of the previous arithmetical operations gives a matrix of the same size. In particular it is possible to consider the inverse of a matrix.

For a scalar number, such as 4, the inverse is ¼ because the product of 4 times ¼ is 1. Then dividing by 4 is defined as multiplying by ¼. We will look for something similar with matrices, where instead of the number 1 we will be dealing with the identity matrix.

2.6.1 Definition – The Inverse of a Matrix

Let A be an n by n square matrix. Let B be a second matrix of the same size so that the products BA and AB are well defined. If B has the property that

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

then it is said to be the inverse of A.

The two elements of the definition are required because of the non-commutativity of matrix multiplication as shown here.

2.6.2 The Uniqueness of the Inverse

It can be readily shown that the two conditions $AB = I$ and $BA = I$ are needed to ensure that there is one and only one inverse for the matrix A. For the n by n matrix A, let B_1 and B_2 be two candidates for the post-inverse. Let B_2 be a pre-inverse. This means that:

$$AB_1 = I,\ AB_2 = I,\ B_3A = I.$$

If this is the case, then subtract the first two equations to give:

$$AB_1 - AB_2 = \underline{0},$$

Where the symbol $\underline{0}$ represents the matrix with all zeros as its elements. We know that we can apply the distributive law to matrices, so this means that

$$A(B_1 - B_2) = \underline{0}.$$

Now since we have a left-inverse B_3 so that $B_3A = I$, we can multiply across this equation by B_3 to get

$$B_3A(B_1 - B_2) = B_3\underline{0}.$$

This then means that

$$B_1 - B_2 = \underline{0}, \text{ so } B_1 = B_2.$$

We have therefore shown that if there is a pre-inverse, there is only one unique post-inverse. The converse follows in the same way. Finally, with the post-inverse B_1 and pre-inverse B_3 so that

$$AB_1 = I,\ B_3A = I,$$

then multiplying the first equation across by B_3 gives

$$B_3AB_1 = B_3,$$

Substituting in the second equation gives

$$IB_1 = B_3,$$

So that $B_1 = B_3$. The converse would also apply. Therefore there is only one matrix B serving as the inverse.

As in ordinary arithmetic, this can be denoted using negative powers: A^{-1} is the inverse of A.

2.6.3 The Inverse of a Product

For matrices A and B, where the terms in this expression are valid, the non-commutative multiplication means that

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

This can be proved algebraically as follows. Let X be the inverse of AB. Then by definition

$$ABX = I.$$

Since the inverses of A and B exist, it follows that

$$A^{-1}ABX = A^{-1}I, \text{ so } BX = A^{-1}.$$

Now multiply across by B^{-1}:

$$B^{-1}BX = B^{-1}A^{-1}, \text{ so that } X = B^{-1}A^{-1}.$$

Therefore the inverse of AB is $B^{-1}A^{-1}$ as required.

2.6.4 Finding the Inverse: 2 by 2

For matrices in general, we will produce algorithms or equations that deliver the inverse of the matrix, defined in terms of the matrix in question. We will start with the simple case of a 2 by 2 matrix. Let A be the most general 2 by 2 matrix.

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

Consider the second 2 by 2 matrix with related elements as shown:

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

Multiply these two matrices in both possible orders:

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix}, \text{ and}$$

$$BA = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix}.$$

Both of these results are simply the 2 by 2 identity matrix I_2,

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

multiplied by the factor $ad - bc$. Then

$$BA = AB = (ad - bc)I_2.$$

Therefore the matrix B divided by the factor $ad - bc$ will satisfy exactly the requirements of an inverse. This proves that the inverse of the matrix

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

is

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

This equation provides a straightforward way of writing down the unique inverse of a 2 by 2 matrix.

2.6.5 The Determinant

Clearly the quantity $ad - bc$ is of paramount importance in this definition. If it is zero then the inverse is not defined and so does not exist for A. This quantity is known as the determinant. So the determinant of the 2 by 2 matrix

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

is given by $ad - bc$. It can be denoted by

$$\det(A),\ |A| \text{ or by a delta symbol: } \Delta(A).$$

Thus

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \Delta\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

We will be using the first notation, det(A). The equation for the 2 by 2 inverse can now be written as follows; for matrix A given by

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

define det(A) = $ad - bc$, then if det(A) $\neq 0$,

$$A^{-1} = \frac{1}{\det(A)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \text{ where .}$$

For a 2 by 2 matrix, the condition $ad - bc = 0$ can be rearranged to show that the top row is equal to the second row multiplied by a scalar value. The same applies to the columns.

2.6.6 Example – Rotations in the Plane

Let P be a point in R^2 with coordinates given by the column matrix

$$P = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Let $R(A)$ be the operation of rotating a point vector through an angle A. From geometry, we know that the point (x, y) will be mapped to the point

$$(x\cos(A) + y\sin(A), -x\sin(A) + y\cos(A))$$

This can be rewritten as matrix arithmetic; set

$$R(A) = \begin{pmatrix} \cos(A) & \sin(A) \\ -\sin(A) & \cos(A) \end{pmatrix}.$$

The point P is then mapped to the point $R(A)P$:

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} \cos(A) & \sin(A) \\ -\sin(A) & \cos(A) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

This result may also be established by noting the results of mapping the canonical orthonormal basis vectors for R^2,

$$(a, b) = a(1, 0) + b(0, 1).$$

Thus the matrix $R(A)$ represents the rotation $R(A)$. Now look at the inverse of this matrix. Firstly, the determinant is

$$\det[\ R(A)\ ] = \cos^2(A) + \sin^2(A) = 1.$$

Following the algorithm for producing the inverse, it is

$$R(-A)^{-1} = \begin{pmatrix} \cos(A) & -\sin(A) \\ \sin(A) & \cos(A) \end{pmatrix}.$$

The inverse must, however, be equivalent to a rotation through the 'opposite' angle, that is, $-A$. This then has the matrix representation

$$R(-A) = \begin{pmatrix} \cos(-A) & \sin(-A) \\ -\sin(-A) & \cos(-A) \end{pmatrix}.$$

Using the properties of these trigonometric functions, this is:

$$R(-A) = \begin{pmatrix} \cos(A) & -\sin(A) \\ \sin(A) & \cos(A) \end{pmatrix}.$$

Therefore we see that

$$R(A)^{-1} = R(-A),$$

which is intuitively correct.

2.6.7 Matrices and Linear Equations

One of the most important applications of matrix algebra comes when it is applied to solving systems of linear equations. We will divert to look at this application. Let x and y be two variables, and B a 2 by 1 column matrix formed from them:

$$B = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Consider what happens if we multiply this matrix by the square matrix A given by

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

The result is:

$$AB = \begin{pmatrix} 1 & 3 \\ -5 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + 3y \\ -5x + 3y \end{pmatrix}.$$

This is a 2 by 1 column matrix whose elements are linear expressions in x and y. Now we look at the case when this is in turn set equal to a constant, that is, known, column matrix; we get the matrix equation

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix}.$$

This means that the matrix equation is the same as:

$$\begin{pmatrix} x + 3y \\ -5x + 3y \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

Which in turn is the same as the system of equations

$$x + 3y = 2, \; -5x + 3y = 8.$$

Looking at this in reverse, the system of equations

$$x + 3y = 2, \; -5x + 3y = 8,$$

can be written as

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

Since the equation was written with all the x and y on the left hand side and in the same order, writing down the matrix was a matter of reading off the coefficients in the order they appear. Now the matrix equation will be solved with the inverse:

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

so multiply across by the inverse of A:

$$A^{-1}A\begin{pmatrix} x \\ y \end{pmatrix} = A^{-1}\begin{pmatrix} 2 \\ 8 \end{pmatrix},$$

Now using the definition of the inverse for the 2 by 2 case, $A^{-1}A = I_2$, the equation is now

$$\begin{pmatrix} x \\ y \end{pmatrix} = A^{-1}\begin{pmatrix} 2 \\ 8 \end{pmatrix}.$$

If the inverse of A exists, in other words $\det(A)$ is not 0, the equation above can be solved immediately for the matrix of unknowns. So solving this system of equations is equivalent to finding the inverse of A. If the determinant of the matrix A were 0, this would mean that one equation was equivalent to the other.

2.6.8 Another Example

It is required to solve the system of equations using a matrix inverse.

$$4x - 3y = 1, y - 2x + 1 = 0.$$

The first step is to write the equations in the correct form, with x and y terms in the same order, and all constant terms on the right-hand side. This is

$$4x - 3y = 1,$$

$$-2x + y = -1.$$

This is now written as a matrix equation:

$$\begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The solution will be:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Finding the inverse of the matrix, firstly calculate the determinant:

$$4\text{x}1 - (-2)(-3) = -2.$$

Then the inverse is

$$\begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}^{-1} = \frac{1}{-2}\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$$

The overall solution is then:

$$\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} -2 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

So the result is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

In other words, $x = 1$ and $y = 1$.

2.6.9 An Interpretation of the Determinant

Let A be a 2 by 2 square matrix

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

where $\det(A) = ad - bc$. Consider now the case of two vectors in two-dimensional space,

$$\begin{pmatrix} a \\ c \end{pmatrix} \text{ and } \begin{pmatrix} b \\ d \end{pmatrix}.$$

The resultant of the addition of the two vectors is

$$\begin{pmatrix} a+b \\ c+d \end{pmatrix}.$$

Consider now the parallelogram with vertices given by the origin and the three vectors listed:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \text{ and } \begin{pmatrix} a+b \\ c+d \end{pmatrix}.$$

From the study of vectors, it is the parallelogram formed during the eponymous 'parallelogram law' of the addition of vectors. It can be shown, by elementary geometry, that the area of this parallelogram is

$$ad - bc.$$

If one column is a multiple of the other, we saw that the determinant was immediately 0. This can be interpreted geometrically; if one vector is a multiple of the other, then the parallelogram is flat and the area is 0. It is also the case that if the rows rather than the columns are used as the vectors, the same interpretation applies. We will extend this idea when we define and study the determinant of larger matrices.

2.6.10 Some Properties of the Determinant

For 2 by 2 matrices, the simple definition of the determinant for this class can be used to establish the following properties, though with considerable algebra. We will see that these results are also true for any square matrices, when the determinant has been defined for these cases.

1. $\det(I_n) = 1$, for any n.
2. $\det(A^T) = \det(A)$
3. $\det(kA^T) = k^n \det(A)$ for an n by n matrix.
4. $\det(AB) = \det(A)\det(B)$.
5. $\det(A^{-1}) = 1/\det(A)$

2.7 Higher Order Determinants

In this section, we will define determinants for higher order matrices. This will lead to a method for calculating inverses for higher order matrices, though it will not be of practical use. We will then establish a method called Cramer's Rule for solving linear systems of equations.

2.7.1 Definition

We saw that the determinant of the general 2 by 2 matrix

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

was $\det(A) = ad - bc$. The determinants of n by n matrices are defined recursively in terms of the 2 by 2 case.

2.7.2 Definition – Cofactors of a Matrix

For a square matrix A, of size n by n, let M_{ij}, be the determinant of the matrix formed by leaving out row i and column j. Then the cofactor for i, j is given by

$$C_{ij} = (-1)^{i+j} M_{ij}.$$

This equation means that if the value of $i + j$ is even, then the cofactor is

$$C_{ij} = M_{ij},$$

and if the value of $i + j$ is odd, then the cofactor is

$$C_{ij} = -M_{ij}.$$

The cofactor may be viewed as the determinant of the sub matrix formed from continuing the matrix elements of A on beyond the value n at the end of each row and column and repeating the elements. The cofactor C_{ij} is then the determinant of the n–1 by n–1 matrix which starts immediately below and to the right of element a_{ij}. This defines a cofactor for each value of i and j, in other words, for every element a_{ij} of matrix A. We can now define the determinant for matrix A.

2.7.3 Definition – the Determinant of a Matrix

Let A be a square matrix. Let a_{ij} be the elements of the matrix, where i and j are integers between 1 and n. The determinant of matrix A is given by

$$\det A = \sum_j a_{ij}C_{ij}$$

This is the elements times the corresponding cofactors summed across row i. Alternatively, using column j:

$$\det A = \sum_i a_{ij}C_{ij}$$

The determinant of an n by n matrix is defined now in terms of the determinants of an n–1 by n–1 matrix, in other words, recursively. This means that were we to use this definition for manually calculating a determinant for anything above a 4 by 4 matrix, it would be very tedious if used as is. Alternative methods will be required.

2.7.4 An Example of Calculating a 3 by 3 determinant

Here is an example of a 3 by 3 matrix A:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 1 & 5 \\ 6 & 0 & 2 \end{pmatrix}.$$

For $a_{11} = 1$, omit the 1st row and 1st column, giving sub-matrix:

$$\begin{pmatrix} 1 & 5 \\ 0 & 2 \end{pmatrix}.$$

The determinant of this matrix is 1x2 – 0x5 = 2. Thus

$$M_{11} = 2.$$

The value of $i + j$ is 2, which is even, so the cofactor is

$$C_{11} = 2.$$

Using the alternative way of constructing these sub-matrices we arrive at the same sub-matrix and no sign check is required. For the 1st row and 2nd column, the sub-matrix is

$$\begin{pmatrix} 4 & 5 \\ 6 & 2 \end{pmatrix}.$$

The determinant of this sub-matrix is –22, so $M_{12} = -22$. The value of $i + j$ is 3, which is odd, so the cofactor is

$$C_{12} = 22.$$

Alternatively, we may look at the matrix starting below and to the right of element a_{12} and assume matrix A is continued or repeated. This means sub-matrix

$$\begin{pmatrix} 5 & 4 \\ 2 & 6 \end{pmatrix}.$$

This sub-matrix has determinant 22. Finally, for the 3 in the 1st row and 3rd column, the sub-matrix is

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

The determinant here is –6, so $M_{13} = -6$. The value of $i + j$ is 4, which is even, so the cofactor is

$$C_{13} = -6.$$

These values can now all be put in the equation

$$\det A = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13},$$

giving

$$\det A = 1\text{x}2 + 2\text{x}22 + 3\text{x}(-6) = 28.$$

2.7.5 An Interpretation of the Determinant

Let A be a square matrix. Let a_{ij} be the elements of the matrix, where i and j are integers between 1 and n. Let v_1, v_2 to v_n be the columns of the matrix A, so that

$$A = (v_1, v_2 \dots v_n).$$

The columns of A will now be regarded as vectors in n-dimensional space. Consider what happens if we take every sum of a combination of k of these vectors, where $k \leq n$. This will be

$$\underline{0},\ \{v_i + v_j \mid \text{every pair } i, j \leq n\},$$

$$\{v_i + v_j + v_k \mid \text{every triplet } i, j, k \leq n\},$$

$$\{\text{ every selection of } k \text{ vectors, where } k < n\},$$

$$\{\textstyle\sum_j v_j\}.$$

The last term, since it selects all n of the column vectors, is simply the sum of those vectors. Now consider the number of points in n-dimensional space this represents, it is

$$1 + {}^{n}\mathrm{C}_2 + {}^{n}\mathrm{C}_3 + {}^{n}\mathrm{C}_4 + \dots + 1.$$

This is

$${}^{n}\mathrm{C}_0 + {}^{n}\mathrm{C}_2 + {}^{n}\mathrm{C}_3 + {}^{n}\mathrm{C}_4 + \dots + {}^{n}\mathrm{C}_n = (1+1)^n = 2^n.$$

Therefore this selection of combinations of the sums of the vectors is a set of 2^n points. In fact, these are the vertices of an n-dimensional parallelogram. The determinant of matrix A can be interpreted as the signed volume of this shape. This has the further implication that if any two of the vertices formed from sums of the vectors are in the same line, then the resulting shape is n–1 dimensional and so has volume 0. Therefore we can say that if $\det(A) = 0$, then one of the columns is a linear combination of the others. Equivalently, it means that there is a set of numbers $a_1, a_2, \ldots a_n$ such that

$$\sum_i a_i v_i = 0.$$

This means that the columns of A are not linearly independent. Conversely if $\det(A) \neq 0$, the columns are linearly independent.

We now outline some properties of determinants which become extremely useful and powerful when carrying out calculations for determinants.

2.7.6 Property 1

If two rows or columns of a matrix are interchanged, the determinant changes in sign. In the example here, we switch column one and two:

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

2.7.7 Property 2

If the elements of one row or column are all multiplied by the same factor, the determinant is multiplied by that factor. This arises immediately from the recursive definition. Then for a matrix A of order n, if every element is multiplied by a scalar k, then:

$$\det(kA) = k^n \det A.$$

This is because the factor of k must be taken out for each row or column.

2.7.8 Property 3

If the elements of one row or column are added to another, the determinant remains unchanged. Here the elements of row 1 are added onto the elements of row 2:

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

In fact any linear combination of rows or columns may be added. As an immediate consequence, if any two rows or columns are the same, the determinant is 0. For example:

$$\det\begin{pmatrix} 2 & 2 & -1 \\ -2 & -2 & -2 \\ 6 & 6 & 1 \end{pmatrix} = 0.$$

Indeed if any row or column of a matrix is a multiple of any other, the determinant is also 0. In the following example matrix, one of the rows or columns here is a multiple of another:

$$\det\begin{pmatrix} 2 & 1 & -2 \\ 3 & -4 & 8 \\ 0 & -1 & 2 \end{pmatrix} = 0.$$

These rules are useful, especially the last one, since they can be used to bring in zeros or reduce the size of calculations.

2.8 Calculating General Inverses

Let A be a matrix with elements a_{ij}. We have established that there is a cofactor C_{ij} associated with each element a_{ij}.

We saw that the determinant was given by the recursive equation

$$\det(A) = \sum_j a_{ij}C_{ij} \text{ or } \det(A) = \sum_i a_{ij}C_{ij}$$

We can use any row, or column, as long as we find the appropriate cofactors. It will always give us the same value for the determinant. However, if we used the elements of one row and the cofactors of another row, the result is 0, in other words

$$\sum_j a_{ij}C_{kj} = 0 \text{ or } \sum_i a_{kj}C_{ij} = 0.$$

This suggests a way of forming an inverse.

Consider the matrix C, formed with elements given by the cofactors C_{ij}, placed in row i column j. The two equations

$$\det(A) = \sum_j a_{ij}C_{ij} \text{ and } \sum_j a_{ij}C_{kj} = 0$$

may be viewed as the calculations for the elements of a matrix which is the product of A and C transposed. It means that

$$\sum_j a_{ij}C_{kj} = \det(A)\delta_{ij}.$$

In other words,

$$AC^{\mathrm{T}} = \det(A)I_n.$$

The second equations,

$$\det(A) = \textstyle\sum_i a_{ij}C_{ij} \text{ and } \sum_i a_{kj}C_{ij} = 0$$

mean that

$$C^{\mathrm{T}}A = \det(A)I_n.$$

It immediately follows that

$$A^{-1} = \frac{1}{\det A}C^T.$$

The transposed matrix of cofactors, the matrix C^T, is known as the *adjoint* of A. The inverse of a matrix is its transposed matrix of cofactors, divided by its determinant.

2.8.1 Example of an Inverse

Find the inverse of the 3 by 3 matrix A:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 1 & 5 \\ 6 & 0 & 2 \end{pmatrix}.$$

To find this inverse, first work out the matrix of cofactors. We had already calculated cofactors for the top row:

$$C_{11} = 2,\ C_{12} = 22,\ C_{13} = -6.$$

These values were used to get the determinant:

$$\det A = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = 1\text{x}2 + 2\text{x}22 + 3\text{x}(-6) = 28.$$

The matrix of cofactors is

$$C = \begin{pmatrix} 2 & 22 & -6 \\ -4 & -16 & 12 \\ 7 & 7 & -7 \end{pmatrix},$$

and the adjoint, the transposed matrix of cofactors, is

$$C^T = \begin{pmatrix} 2 & -4 & 7 \\ 22 & -16 & 7 \\ -6 & 12 & -7 \end{pmatrix}.$$

The inverse is then this matrix divided by the determinant:

$$A^{-1} = \frac{1}{\det A} C^T = \frac{1}{28} \begin{pmatrix} 2 & -4 & 7 \\ 22 & -16 & 7 \\ -6 & 12 & -7 \end{pmatrix}.$$

We will now use this result to solve a system of equations.

2.8.2 Another Set of Equations

Solve the following system of equations:

$$7x = 11 + 2y,$$

$$3y + 4z = 4x + 2,$$

$$3x = y + 5.$$

The equations are written with the unknowns in the same order:

$$7x - 2y = 11,$$

$$-4x + 3y + 4z = 2,$$

$$3x - y = 5.$$

Now the system of equations can be easily written as a matrix equation:

$$\begin{pmatrix} 7 & -2 & 0 \\ -4 & 3 & 4 \\ 3 & -1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 11 \\ 2 \\ 5 \end{pmatrix}.$$

To find the inverse of the 3 by 3 matrix, the first step is calculating the matrix of cofactors. It is:

$$C = \begin{pmatrix} 4 & 12 & -5 \\ 0 & 0 & 1 \\ -8 & -28 & 13 \end{pmatrix}.$$

Finding the determinant of the matrix is made simple by using the third column – the two zero's mean the determinant is

$$0x(-5) + 4x1 + 0x13 = 4.$$

The inverse is then the transposed matrix of cofactors, divided by the determinant.

$$A^{-1} = \frac{1}{\det A} C^T.$$

The result is:

$$A^{-1} = \frac{1}{4}\begin{pmatrix} 4 & 0 & -8 \\ 12 & 0 & -28 \\ -5 & 1 & 13 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -2 \\ 3 & 0 & -7 \\ -1.25 & 0.25 & 3.25 \end{pmatrix}.$$

Using this matrix inverse, the solution is then:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & -2 \\ 3 & 0 & -7 \\ -1.25 & 0.25 & 3.25 \end{pmatrix}\begin{pmatrix} 11 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}.$$

Thus $x = 1$, $y = -2$ and $z = 3$.

2.9 Cramer's Rule

Consider a 3 by 3 general system of equations and the elements involved in solving it using matrix inverses:

$$A\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

If the a_{ij} are the elements of the matrix A, then this is equivalent to the system shown here:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2,$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3.$$

Denoting the cofactor from row i, column j by c_{ij}, the solution to the matrix equation is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \frac{1}{\det A}\begin{pmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

The matrix has been transposed, so the cofactors are in different positions than the indices would indicate. When these matrices are multiplied out to give the solutions, the solution for x_1 is:

$$x_1 = (c_{11}b_1 + c_{21}b_2 + c_{31}b_3)/\det A.$$

The expression in brackets is the same as a determinant calculated for the matrix A, using column 1, but with the elements a_{11}, a_{21} and a_{31} replaced by b_1, b_2, b_3.

In other words, x_1 is given by

$$x_1 = \frac{1}{\det A}\det\begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix},$$

and in a similar way,

$$x_2 = \frac{1}{\det A}\det\begin{pmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{pmatrix},\ x_3 = \frac{1}{\det A}\det\begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{pmatrix}.$$

We have come to the following procedure for solving a linear system, known as Cramer's Rule.

2.9.1 Cramer's Rule

Let A be an n by n square matrix, $\boldsymbol{x}$ a column matrix of unknowns, and $\boldsymbol{b}$ a column matrix of known values. The linear system

$$A\mathbf{x} = \boldsymbol{b},$$

is solved as follows: Let A_i be the matrix formed by replacing the i-th column of A with the column matrix $\boldsymbol{b}$**.** Then the value of the unknown x_i is:

$$x_i = {\det A_i}/{\det A}\cdot$$

2.9.2 An Application of Cramer's Rule

We saw the following system of equations:

$$2x - y - z = 1,$$

$$3x + y + 2z = 2,$$

$$x - 2y - 3z = 1.$$

written in matrix form as:

$$A\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \text{ where } A = \begin{pmatrix} 2 & -1 & -1 \\ 3 & 1 & 2 \\ 1 & -2 & -3 \end{pmatrix}.$$

The first step in solving this matrix equation should always be to find the determinant of the matrix A, since if it is zero then there is no solution to the system of equations. Here, the determinant of A is –2. To solve for the variable x, replace the first column of A with the column matrix of known values; this is the matrix A_1:

$$A_1 = \begin{pmatrix} 1 & -1 & -1 \\ 2 & 1 & 2 \\ 1 & -2 & -3 \end{pmatrix}.$$

The solution for x is then

$$x = {\det A_1}\big/{\det A}.$$

The determinant of A_1 is:

$$\det A_1 = \det\begin{pmatrix} 1 & -1 & -1 \\ 2 & 1 & 2 \\ 1 & -2 & -3 \end{pmatrix} = \det\begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 4 \\ 1 & -1 & -2 \end{pmatrix}.$$

The Determinant of this matrix can be quickly found using the top row, the element in row 1, column 1, with its cofactor:

$$\det A_1 = 1 \text{x} \det\begin{pmatrix} 3 & 4 \\ -1 & -2 \end{pmatrix} = 3\text{x}(-2) - (-1)\text{x}4 = -2.$$

The value of x is then

$$x = \det A_1/\det A = -2/(-2) = 1.$$

To find y, replace the second column by the column matrix of constants:

$$A_2 = \begin{pmatrix} 2 & 1 & -1 \\ 3 & 2 & 2 \\ 1 & 1 & -3 \end{pmatrix}.$$

The determinant of this matrix can be simplified by adding the second column to the third:

$$\det A_2 = \det\begin{pmatrix} 2 & 1 & -1 \\ 3 & 2 & 2 \\ 1 & 1 & -3 \end{pmatrix} = \det\begin{pmatrix} 2 & 1 & 0 \\ 3 & 2 & 4 \\ 1 & 1 & -2 \end{pmatrix}.$$

Then add twice the third row to the second row:

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

This determinant is then the element in row 3, column 3, by its cofactor:

$$\det A_2 = (-2)\text{x}\det\begin{pmatrix} 2 & 1 \\ 5 & 4 \end{pmatrix} = (-2)\text{x}[2\text{x}4 - 5\text{x}1] = -6.$$

The value of y is then

$$y = -6/(-2) = 3.$$

We can now just use on of the original equations for z:

$$2x - y - z = 1 \text{ so } z = 2x - y - 1,$$

Then $z = -2$.

The advantage of Cramer's rule is that it allows the calculation of any one of the unknowns, without having to calculate all of them together. A further advantage of this method is that each calculation is two n by n determinants, and this can be made easier using row and column addition for determinants. This comes into its own with larger size systems.

2.9.3 Example – a Fourth Order Case

Solve the following system of equations for the second variable and the last:

$$x - y + 2z + 3w = 6,$$

$$2x - y + 5z + 4w = 13,$$

$$3x + 2y + z - 4w = 3,$$

$$3x + y + 2z - 3w = 6.$$

Only two of the four possible variables are required. Writing the system as a matrix equation gives:

$$\begin{pmatrix} 1 & -1 & 2 & 3 \\ 2 & -1 & 5 & 4 \\ 3 & 2 & 1 & -4 \\ 3 & 1 & 2 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 6 \\ 13 \\ 3 \\ 6 \end{pmatrix}.$$

Let A be the matrix above:

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

The determinant of A should be calculated first; it is simplified by adding row 1 to row 4:

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

Then subtract column 1 from column 3:

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

The 4th row now has three 0's, so use this row to calculate the determinant:

$$4\times\left((-1)\times\det\begin{pmatrix}-1 & 1 & 3\\ -1 & 3 & 4\\ 2 & -2 & -4\end{pmatrix}\right).$$

It is multiplied by a –1 since the element '4' is in row 4, column 1. The determinant can be reduced to a 2 by 2 determinant. The calculations are:

$$-4\times\det\begin{pmatrix}-1 & 1 & 3\\ -1 & 3 & 4\\ 2 & -2 & -4\end{pmatrix} = -4\times\det\begin{pmatrix}0 & 1 & 3\\ 2 & 3 & 4\\ 0 & -2 & -4\end{pmatrix}.$$

Now break down the calculation for the 3 by 3:

$$= 4 \times 2\left(\det\begin{pmatrix}1 & 3\\ 2 & 4\end{pmatrix}\right) = 8\times\det\begin{pmatrix}1 & 3\\ 2 & 4\end{pmatrix} = 16.$$

The first part of the problem is to calculate y, the second variable. Its value is given by replacing the second column with the column matrix on the right-hand side. Call this A_2:

$$A_2 = \begin{pmatrix} 1 & 6 & 2 & 3 \\ 2 & 13 & 5 & 4 \\ 3 & 3 & 1 & -4 \\ 3 & 6 & 2 & -3 \end{pmatrix}.$$

Then the value of y is given by:

$$y = \det A_2/\det A.$$

The A_2 determinant must be calculated next. Here add the 1st row to the last, and the 2nd row to the 3rd:

$$\det\begin{pmatrix} 1 & 6 & 2 & 3 \\ 2 & 13 & 5 & 4 \\ 3 & 3 & 1 & -4 \\ 3 & 6 & 2 & -3 \end{pmatrix} = \det\begin{pmatrix} 1 & 6 & 2 & 3 \\ 2 & 13 & 5 & 4 \\ 5 & 16 & 6 & 0 \\ 4 & 12 & 4 & 0 \end{pmatrix}.$$

A factor of 4 can now be taken out from the 4th row. This step is not, in itself, essential, but it does mean the figures involved in the calculation are lower in magnitude. The result is:

$$\det\begin{pmatrix} 1 & 6 & 2 & 3 \\ 2 & 13 & 5 & 4 \\ 5 & 16 & 6 & 0 \\ 4 & 12 & 4 & 0 \end{pmatrix} = 4 \times \det\begin{pmatrix} 1 & 6 & 2 & 3 \\ 2 & 13 & 5 & 4 \\ 5 & 16 & 6 & 0 \\ 1 & 3 & 1 & 0 \end{pmatrix}.$$

Subtract 3 times the 1st column from the 2nd, and following this the 1st from the 3rd. The calculations are:

$$4 \times \det\begin{pmatrix} 1 & 6 & 1 & 3 \\ 2 & 13 & 3 & 4 \\ 5 & 16 & 1 & 0 \\ 1 & 3 & 0 & 0 \end{pmatrix} = 4 \times \det\begin{pmatrix} 1 & 3 & 1 & 3 \\ 2 & 7 & 3 & 4 \\ 5 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

This determinant can now be reduced to one 3 by 3 determinant by using the 4th row:

$$4 \times (-1) \times \det\begin{pmatrix} 3 & 1 & 3 \\ 7 & 3 & 4 \\ 1 & 1 & 0 \end{pmatrix} = (-4) \times \det\begin{pmatrix} 3 & 1 & 3 \\ 7 & 3 & 4 \\ 1 & 1 & 0 \end{pmatrix}.$$

Subtracting the 2nd column from the 1st brings one more 0, and then the determinant is reduced to a 2 by 2:

$$-4 \times \det\begin{pmatrix} 2 & 1 & 3 \\ 4 & 3 & 4 \\ 0 & 1 & 0 \end{pmatrix} = (-4) \times (-1) \times \det\begin{pmatrix} 2 & 3 \\ 4 & 4 \end{pmatrix} = -16.$$

Then

$$y = -16/16 = -1.$$

The next variable to calculate is the fourth, w. The original matrix system was:

$$\begin{pmatrix} 1 & -1 & 2 & 3 \\ 2 & -1 & 5 & 4 \\ 3 & 2 & 1 & -4 \\ 3 & 1 & 2 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 6 \\ 13 \\ 3 \\ 6 \end{pmatrix}.$$

The variable w, is given by replacing the fourth column with the column matrix on the right-hand side. So then:

$$A_4 = \begin{pmatrix} 1 & -1 & 2 & 6 \\ 2 & -1 & 5 & 13 \\ 3 & 2 & 1 & 3 \\ 3 & 1 & 2 & 6 \end{pmatrix},$$

And then $w = \det(A_4)/\det(A)$. Calculating the determinant of A_4, it is simplified by subtracting row 4 from row 1. The result is:

$$\det A_4 = \det\begin{pmatrix} 1 & -1 & 2 & 6 \\ 2 & -1 & 5 & 13 \\ 3 & 2 & 1 & 3 \\ 3 & 1 & 2 & 6 \end{pmatrix} = \det\begin{pmatrix} -2 & -2 & 0 & 0 \\ 2 & -1 & 5 & 13 \\ 3 & 2 & 1 & 3 \\ 3 & 1 & 2 & 6 \end{pmatrix}.$$

Now subtracting column 1 from column 2 results in 3 0's in the top row, and the determinant can be reduced to a 3 by 3. The details are:

$$\det\begin{pmatrix} -2 & -2 & 0 & 0 \\ 2 & -1 & 5 & 13 \\ 3 & 2 & 1 & 3 \\ 3 & 1 & 2 & 6 \end{pmatrix} = \det\begin{pmatrix} -2 & 0 & 0 & 0 \\ 2 & -3 & 5 & 13 \\ 3 & -1 & 1 & 3 \\ 3 & -2 & 2 & 6 \end{pmatrix} = -2 \times \det\begin{pmatrix} -3 & 5 & 13 \\ -1 & 1 & 3 \\ -2 & 2 & 6 \end{pmatrix}.$$

Now row 3 is twice row 2, so $\det A_4 = 0$ and so $w = 0$.

3 Eigenvalues and Eigenvectors

Many applications of linear algebra involve matrix equations of the form

$$A\boldsymbol{x} = \lambda \boldsymbol{x},$$

where A is an n by n square matrix, $\boldsymbol{x}$ is an n by 1 column matrix and λ is a scalar. This equation means that for a particular square matrix A, if the column matrix $\boldsymbol{x}$ and scalar λ can be found, then the effect of multiplying column matrix $\boldsymbol{x}$ by matrix A is the same as if it has been simply multiplied by a scalar number λ. This also applies to any multiple of matrix $\boldsymbol{x}$.

The number λ is called an *eigenvalue*, and the column matrix $\boldsymbol{x}$ is called an *eigenvector*. A method must be produced for finding the solution to this equation. The first step is to find the possible values of scalar λ.

3.1.1 Finding the Eigenvalue

From the original equation

$$A\boldsymbol{x} = \lambda\boldsymbol{x},$$

start by bringing $\lambda\boldsymbol{x}$ to the left-hand side. The equation is now

$$A\boldsymbol{x} - \lambda\boldsymbol{x} = \mathbf{0}.$$

The boldface $\mathbf{0}$ indicates that the right-hand side is the n by 1 column matrix of zeros. The equation can be written as

$$A\boldsymbol{x} - \lambda I\boldsymbol{x} = \mathbf{0},$$

where I is the n by n identity matrix, so that we can take out the common factor of column matrix $\boldsymbol{x}$:

$$(A - \lambda I)\boldsymbol{x} = \mathbf{0}.$$

If the matrix $A - \lambda I$, multiplying column matrix $\boldsymbol{x}$ on the left-hand side, can be inverted, then the solution for matrix $\boldsymbol{x}$ is

$$\boldsymbol{x} = (A - \lambda I)^{-1}\,\mathbf{0}, \text{ so that } \boldsymbol{x} = \mathbf{0}.$$

To avoid this trivial solution, matrix $A - \lambda I$ is singular:

$$\det(A - \lambda I) = 0.$$

This gives an equation for the scalar λ. This is called the *characteristic equation*. If there are solutions to this equation, which will be a polynomial of order n, then the matrix

$$A - \lambda I$$

is singular, and so there are values of λ for which a non-trivial solution exists for

$$(A - \lambda I)\boldsymbol{x} = \mathbf{0}.$$

3.1.2 A Starting 2 by 2 Example

We will find the solution of the above equation for the matrix A:

$$A = \begin{pmatrix} 4 & 1 \\ 5 & 0 \end{pmatrix}.$$

The characteristic equation is

$$\det(A - \lambda I) = 0.$$

The first task:

$$A - \lambda I = \begin{pmatrix} 4 & 1 \\ 5 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 5 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} = \begin{pmatrix} 4-\lambda & 1 \\ 5 & -\lambda \end{pmatrix}.$$

The characteristic equation is now

$$\det\begin{pmatrix} 4-\lambda & 1 \\ 5 & -\lambda \end{pmatrix} = 0$$

This determinant is written in terms of λ; it is

$$(4 - \lambda)(-\lambda) - 5\text{x}1 = 0.$$

This is a quadratic:

$$\lambda^2 - 4\lambda - 5 = 0.$$

Factorise:

$$(\lambda + 1)(\lambda - 5) = 0, \text{ and so } \lambda = -1 \text{ and } \lambda = 5.$$

Thus the eigenvalues of the matrix

$$A = \begin{pmatrix} 4 & 1 \\ 5 & 0 \end{pmatrix}$$

are $\lambda = -1$ and $\lambda = 5$.

3.1.3 Finding the Eigenvector

Before we calculate the eigenvector for these eigenvectors, we will anticipate some properties it may have. The original defining equation for λ and $\boldsymbol{x}$ was

$$A\boldsymbol{x} = \lambda\boldsymbol{x},$$

and in order that there be non-trivial solutions, the characteristic equation must be satisfied:

$$\det(A - \lambda I) = 0.$$

The first matrix equation

$$(A - \lambda I)\boldsymbol{x} = \mathbf{0},$$

represents a set of equations for the n unknown values of the elements of the column matrix $\boldsymbol{x}$. But since we insist that

$$\det(A - \lambda I) = 0,$$

this means that the system cannot be fully solved. If we write them out as a set of linear equations, we will get at most $n - 1$ distinct equations. Looking again at the defining equation:

$$(A - \lambda I)\boldsymbol{x} = \mathbf{0}.$$

Multiply both sides of this equation by a fixed number, a to get

$$a(A - \lambda I)\boldsymbol{x} = \mathbf{0}.$$

Take the a through to the $\boldsymbol{x}$ gives

$$(A - \lambda I)a\boldsymbol{x} = \mathbf{0}.$$

This result means that if the column matrix $\boldsymbol{x}$ is a solution, then so is the column matrix $a\boldsymbol{x}$, for any number a. The column matrix $\boldsymbol{x}$ can therefore only be found up to a multiplicative factor, and this will be seen when solutions are found from the resulting set of equations.

We will find the solution of the above equation for the matrix A. Our eigenvalues were found to be 5 and –1.

$$A = \begin{pmatrix} 4 & 1 \\ 5 & 0 \end{pmatrix}.$$

For $\lambda = 5$, the equation is

$$(A - 5I)\boldsymbol{x} = \mathbf{0}.$$

Write up the 2 by 2 matrix:

$$A - 5I = \begin{pmatrix} 4-5 & 1 \\ 5 & -5 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 5 & -5 \end{pmatrix}.$$

Then $(A - 5I)\boldsymbol{x} = \mathbf{0}$ means

$$\begin{pmatrix} -1 & 1 \\ 5 & -5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Multiplying out gives

$$-x_1 + x_2 = 0,$$

$$-5x_1 + 5x_2 = 0.$$

These are both the same equation, $x_1 = x_2$, so no more information about x_1 and x_2 can be found. The convention we will adopt is to choose a value of 1 for x_1, if possible, and then find the remaining values from this. Thus $x_2 = 1$ and the column matrix $\boldsymbol{x}$ associated with the eigenvalue $\lambda = 5$ is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The second convention employed now is to ensure that the eigenvalue is of length 1; this is done by dividing by the length of the vector found so far. This eigenvector is called the *unit* eigenvector. The vector $(1, 1)^{\mathrm{T}}$ is of length $\sqrt{(1^2 + 1^2)} = \sqrt{2}$. The unit eigenvector for $\lambda = 5$ is then

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

For the second eigenvalue, $\lambda = -1$, the equation is

$$(A + I)\boldsymbol{x} = \mathbf{0},$$

so that

$$\begin{pmatrix} 5 & 1 \\ 5 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

In this case there is only one equation:

$$5x_1 + x_2 = 0.$$

This means that

$$x_2 = -5x_1.$$

Following the same convention and taking $x_1 = 1$, the column matrix $\boldsymbol{x}$ associated with the eigenvalue $\lambda = -1$ has then been shown to be

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -5 \end{pmatrix}.$$

The corresponding unit eigenvector is then

$$\frac{1}{\sqrt{26}}\begin{pmatrix} 1 \\ -5 \end{pmatrix}.$$

3.1.4 A 3 by 3 Example

Find the eigenvalues and eigenvectors of the matrix

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

The characteristic equation is

$$\det(A - \lambda I) = 0.$$

The matrix $A - \lambda I$ is

$$A - \lambda I = \begin{pmatrix} 2-\lambda & 0 & 1 \\ -1 & 4-\lambda & -1 \\ -1 & 2 & -\lambda \end{pmatrix}.$$

The determinant equation is now

$$\det\begin{pmatrix} 2-\lambda & 0 & 1 \\ -1 & 4-\lambda & -1 \\ -1 & 2 & -\lambda \end{pmatrix} = 0.$$

Use the cofactor method on the top row, since it has a zero. For element in row 1, column 1, the element is $2 - \lambda$. Leaving out row 1 and column 1, the sub-matrix is:

$$\begin{pmatrix} 4-\lambda & -1 \\ 2 & -\lambda \end{pmatrix}.$$

Therefore cofactor for row 1 column 1 is $\lambda^2 - 4\lambda + 2$. The next element is row 1, column 3. Leaving out row 1 and column 3, the sub-matrix is:

$$\begin{pmatrix} -1 & 4-\lambda \\ -1 & 2 \end{pmatrix}.$$

The cofactor is $2 - \lambda$. The determinant is then each element by its corresponding cofactor:

$$\det(A - \lambda I) = (2 - \lambda)(\lambda^2 - 4\lambda + 2) + 0 + 1(2 - \lambda).$$

This in turn was to be set to zero, giving the equation:

$$(2 - \lambda)(\lambda^2 - 4\lambda + 2) + 2 - \lambda = 0.$$

There is a common factor of $2 - \lambda$, so simplifying the LHS gives the cubic equation

$$(2 - \lambda)[(\lambda^2 - 4\lambda + 2) + 1] = 0.$$

$$(2 - \lambda)[\lambda^2 - 4\lambda + 3] = 0.$$

We have kept in the factor $(2 - \lambda)$, so this is our first factor to solve: this is 0 when $\lambda = 2$. The remaining roots come from the quadratic:

$$\lambda^2 - 4\lambda + 3$$

The roots of this quadratic are:

$$\lambda^2 - 4\lambda + 3 = (\lambda - 1)(\lambda - 3).$$

The other two roots are then 1 and 3. The solutions to this equation are: $\lambda = 2$, $\lambda = 1$ and $\lambda = 3$.

Now calculate the eigenvectors. Set $\boldsymbol{x} = (x_1, x_2, x_3)^T$ so that the eigenvector equation is

$$\begin{pmatrix} 2-\lambda & 0 & 1 \\ -1 & 4-\lambda & -1 \\ -1 & 2 & -\lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

For $\lambda = 1$, the matrix equation becomes

$$\begin{pmatrix} 1 & 0 & 1 \\ -1 & 3 & -1 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This is multiplied out to give the system of equations

$$x_1 + x_3 = 0,$$

$$-x_1 + 3x_2 - x_3 = 0,$$

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

Subtracting the third equation from the second gives a value

$$x_2 = 0.$$

The other equations are all equivalent to

$$x_1 + x_3 = 0,$$

in other words, $x_3 = -x_1$, so with $x_1 = 1$ then $x_3 = -1$. Thus the eigenvector for $\lambda = 1$ is

$$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \text{ or as a unit vector: } \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

For $\lambda = 2$, the matrix equation becomes

$$\begin{pmatrix} 0 & 0 & 1 \\ -1 & 2 & -1 \\ -1 & 2 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This gives the system of equations

$$x_3 = 0,$$

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

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

Putting in the value $x_3 = 0$ into the second or third equation gives

$$x_1 = 2x_2,$$

so with $x_1 = 2$ then $x_2 = 1$. Thus the eigenvector for $\lambda = 2$ is

$$\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \text{ or as a unit vector: } \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

For $\lambda = 3$, the matrix equation becomes

$$\begin{pmatrix} -1 & 0 & 1 \\ -1 & 1 & -1 \\ -1 & 2 & -3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

This gives the system of equations

$$-x_1 + x_3 = 0,$$

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

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

The first equation is $x_1 = x_3$. Using this to remove x_3 in the second or third equation gives

$$-2x_1 + x_2 = 0, \text{ so } x_2 = 2x_1.$$

With $x_1 = 1$ then $x_2 = 2$. Thus the eigenvector for $\lambda = 3$ is

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \text{ or as a unit vector: } \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

3.1.5 Rotation Matrices – Eigenvalues and Eigenvectors

Let the value A be an angle. We will find the eigenvalues and eigenvectors of the following matrix, interpreting the result:

$$R(A) = \begin{pmatrix} \cos A & -\sin A & 0 \\ \sin A & \cos A & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The characteristic equation means:

$$\text{Det}[\, R(A) - \lambda I \,] = 0.$$

For this matrix, we have:

$$\det\begin{pmatrix} \cos A - \lambda & -\sin A & 0 \\ \sin A & \cos A - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{pmatrix} = 0.$$

Apply the cofactor method, the determinant is:

$$(1-\lambda)\det\begin{pmatrix}\cos A-\lambda & -\sin A\\ \sin A & \cos A-\lambda\end{pmatrix}=0.$$

This is equivalent to

$$(1-\lambda)(\ (\cos A-\lambda)^2+\sin^2 A)=0.$$

The first factor here is $1-\lambda$, so the first root is $\lambda=1$.

The other roots are found from the quadratic:

$$(\cos A-\lambda)^2+\sin^2 A=0.$$

Bring over the sine term and take the square root of both sides, noting the two possibilities:

$$\cos A-\lambda=\pm \mathrm{j}\sin A.$$

Then $\lambda=\cos A\pm \mathrm{j}\sin A=\exp(\pm\, \mathrm{j}A)$. Thus there is one real root, 1, and two complex roots: $e^{\mathrm{j}A}$ and $e^{-\mathrm{j}A}$.

We will find the eigenvector of the real root. To do this, return to the system of equations

$$(A-\lambda I\,)\boldsymbol{x}=\mathbf{0}.$$

This is:

$$\begin{pmatrix}\cos A-1 & -\sin A & 0\\ \sin A & \cos A-1 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

Multiplying out this matrix equation would not 'pick up' the x_3 variable, and there would only be two equations in x_1 and x_2. So we can leave it as a set of equations in x_1 and x_2 :

$$\begin{pmatrix}\cos A-1 & -\sin A\\ \sin A & \cos A-1\end{pmatrix}\begin{pmatrix}x_2\\ x_3\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}.$$

See if this two by two matrix can be inverted; its determinant is

$$(\cos A - 1)^2 + \sin^2 A.$$

Multiply this out:

$$\cos^2 A - 2\cos A + 1 + \sin^2 A.$$

This is:

$$2 - 2\cos A = 2(1 - \cos A).$$

Then so as long as A is not any multiple of 2π, we can say that the equation has a solution in the normal way of solving any matrix equation; the inverse will exist. It can be solved so that

$$\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

leaving x_3 as the only non-zero value. Thus the unit eigenvector for eigenvalue $\lambda = 1$ is

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

which is the z axis. It is left to the reader to identify what happens when A is a multiple of 2π.