Matrix Algebra

A course of lectures given to second year physics students as part of the Maths II module.

Lecturer: Professor Peter Main

This course builds upon the short introduction to matrix algebra you were given in the first year. This is a very powerful branch of mathematics much used by physicists. The notes start by describing the matrix algebra you should already know and so can be used for revision. They include the essentials of matrix algebra such as addition, multiplication and algebraic manipulation of matrices and matrix inversion.

Matrices are then used to represent a set of linear simultaneous equations and their solution. This is a powerful piece of mathematics as huge sets of equations can be represented by very few symbols. We also find out how to deal with the situation where there is insufficient information in the equations to give a unique solution.

The matrix is also used as an operator to scale, rotate and reflect two-dimensional objects. Combinations of operations are also considered. This forms the basis of computer graphics and, in other courses, you will find more uses of the matrix as an operator.

Next, we consider the eigenvectors and eigenvalues of symmetric and Hermitian matrices. This is the most powerful of all the mathematical techniques in the course, which allows you to analyse complicated systems in terms of much simpler components. A mathematical example of this is presented as a case study.

Finally, there is a brief introduction to tensors. Examples of their use in physics are presented and some of their mathematical properties are described.

http://www-users.york.ac.uk/~pm1/PMweb/teaching.htm

Matrices

A matrix is a rectangular array of **elements**, which are typically numbers (they may be complex) or algebraic expressions. The whole array is enclosed in brackets.

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

A 2×3 matrix

A matrix which has m rows and n columns is of order $m\times n$.

Notation

A matrix can be represented by a single symbol, e.g. $\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 0 & -1 \\ 2 & 4 \end{pmatrix}$ or

$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} = \left(a_{ij}\right)$ where a symbol is given to each element.

The element a_{37} is the element in row 3 and column 7.

The fact that a large array of data can be represented by a single symbol makes it very convenient for the algebraic manipulation of vast amounts of information.

Matrix equality

Two matrices **A** and **B** are equal only if they are of the same order and corresponding elements are equal, i.e. $a_{\mathrm{ij}} = b_{\mathrm{ij}}$ for all i and j.

Addition and subtraction

Matrices can be added or subtracted only if they have the same order:

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

The sums (or differences) of corresponding elements are performed.

Multiplication by a scalar

$$3\begin{pmatrix} -1 & 4 \\ 2 & -2 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} -3 & 12 \\ 6 & -6 \\ 9 & 3 \end{pmatrix} \quad \text{also written as} \quad k\left(a_{ij}\right) = \left(k\,a_{ij}\right)$$

i.e. multiply every element by the scalar.

The transpose of a matrix

If $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 3 & -2 \\ 4 & 3 \end{pmatrix}$ then the transpose is $\mathbf{A}^{\mathrm{T}} = \begin{pmatrix} 2 & 3 & 4 \\ 1 & -2 & 3 \end{pmatrix}$, i.e. rows become columns and columns become rows. An $m\times n$ matrix becomes $n\times m$ by transposing it.

Special matrices

There are many matrices with special names. Here are some you may meet in this course:

Square matrix – this has the same number of columns as rows.

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

Symmetric matrix – this is equal to its own transpose, i.e. $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$

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

Diagonal matrix – only the diagonal elements are non-zero

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

Identity matrix – a diagonal matrix with all diagonal elements = 1

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

Hermitian matrix – the transpose is the complex conjugate

$$\begin{pmatrix} a_{11} & a_{12} + i\,b_{12} \\ a_{12} - i\,b_{12} & a_{22} \end{pmatrix}$$

Matrix multiplication

Two matrices can be multiplied together only if the number of columns in the first is equal to the number of rows in the second.

If **A** is of order $m\times n$ then **B** must be $n\times p$ to form the product **AB**.

Look at the pattern in the result of the following multiplication:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \qquad \mathbf{AB} = \begin{pmatrix} a_{11}b_1 + a_{12}b_2 + a_{13}b_3 \\ a_{21}b_1 + a_{22}b_2 + a_{23}b_3 \end{pmatrix}$$

Each element in the result is the dot product of a row in **A** with a column of **B**. The same process is carried out in the more general case:

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

It can be seen that

$$\begin{array}{ccccc} \mathbf{A} & \mathbf{B} & = & \mathbf{C} \\ m\times n & n\times p & & m\times p \end{array}$$

and c_{ij} is the scalar product of the ith row of **A** with the jth column of **B**:

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

It should also be clear that even if **AB** exists, **BA** may not (unless $m = p$). Even if **BA** does exist, it is in general not equal to **AB**.

$$\left.\begin{array}{cccc} \mathbf{A} & \mathbf{B} & = & \mathbf{C} \\ m\times n & n\times m & & m\times m \\ \mathbf{B} & \mathbf{A} & = & \mathbf{D} \\ n\times m & m\times n & & n\times n \end{array}\right\} \text{ } \mathbf{AB} \text{ and } \mathbf{BA} \text{ are of different orders.}$$

See also: $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ -1 & 3 \end{pmatrix}$ $\quad \mathbf{B} = \begin{pmatrix} -1 & 3 \\ 1 & 2 \end{pmatrix}$ $\quad \mathbf{AB} = \begin{pmatrix} -1 & 8 \\ 4 & 3 \end{pmatrix}$ $\quad \mathbf{BA} = \begin{pmatrix} -5 & 8 \\ 0 & 7 \end{pmatrix}$

The order of the matrices in a product is clearly important – matrices do not commute.

In the product **AB**, **A** premultiplies **B** and **B** postmultiplies **A**.

Multiplication by the identity matrix

$$\begin{pmatrix} 3 & 0 & -1 \\ 1 & 2 & 1 \\ -2 & 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & -1 \\ 1 & 2 & 1 \\ -2 & 0 & 2 \end{pmatrix} \quad \text{i.e. } \mathbf{AI} = \mathbf{A} \quad \text{also} \quad \mathbf{IA} = \mathbf{A}$$

Another matrix product

$$\begin{pmatrix} 2 & 1 & -3 \\ 6 & 3 & -9 \end{pmatrix}\begin{pmatrix} 1 & 9 \\ 4 & -6 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$ which is the null matrix.

Note that $\mathbf{AB} = \mathbf{0}$ does not imply either $\mathbf{A} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$.

Similarly, $\mathbf{AB} = \mathbf{AC}$ does not imply that $\mathbf{B} = \mathbf{C}$.

Matrix representation of linear simultaneous equations

Consider the equations:

$$3x_1 + 2x_2 = 1$$
$$2x_1 - x_2 = -4$$

In matrix notation: $\begin{pmatrix} 3 & 2 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -4 \end{pmatrix}$ – multiply out the lhs to get the same as above.

Using fewer symbols, this is: $\mathbf{A} \quad \mathbf{x} \quad = \quad \mathbf{b}$

lhs matrix   vector of unknowns   rhs vector

This is the most concise representation of the equations. It is extremely powerful since the symbols can represent matrices and vectors of any order – millions of equations in millions of unknowns is not unheard of.

The inverse matrix

The solution of the equations can also be represented using symbols as follows:

If we write a simple equation in terms of scalars: $ax = b$

then its solution is: $x = \dfrac{b}{a} = a^{-1}b$

On the rhs, we have multiplied b by the reciprocal of a to get the solution. Since matrix division is not defined, we must do something similar to solve the matrix equation, i.e. define an inverse matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. This allows us to perform the following manipulation:

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

Premultiply by $\mathbf{A}^{-1}$: $\mathbf{A}^{-1}\,\mathbf{A}\,\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b}$

Therefore $\mathbf{I}\,\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b}$

Therefore $\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b}$

So the solution is formally obtained by premultiplying the rhs vector by the inverse of the lhs matrix.

Useful results without proof

There is little point in proving the following, but you may find the results useful:

Associative rule: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ and $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

Combination of determinants: $\det(\mathbf{AB}) = \det(\mathbf{A}).\det(\mathbf{B})$

Transpose of a product: $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$

Inverse of a product: $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Calculation of the inverse matrix

Let us calculate the inverse of $\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ -2 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}$ such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$

Let $\mathbf{A}^{-1} = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix}$ then $\begin{pmatrix} 1 & 0 & -1 \\ -2 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix}\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Now multiply out the lhs matrices and equate elements with the rhs matrix. This gives a system of 9 equations in 9 unknowns which can be written as:

$$\left.\begin{array}{rrrcl} x_1 & & -z_1 & = & 1 \\ -2x_1 & +y_1 & & = & 0 \\ x_1 & -y_1 & +2z_1 & = & 0 \end{array}\right\} \quad \left.\begin{array}{rrrcl} x_2 & & -z_2 & = & 0 \\ -2x_2 & +y_2 & & = & 1 \\ x_2 & -y_2 & +2z_2 & = & 0 \end{array}\right\} \quad \left.\begin{array}{rrrcl} x_3 & & -z_3 & = & 0 \\ -2x_3 & +y_3 & & = & 0 \\ x_3 & -y_3 & +2z_3 & = & 1 \end{array}\right\}$$

Note that the lhs matrix is the same in all 3 cases. This means the 3 systems of equations can be written in the following compact form, where only the elements of lhs matrix and the right-hand-sides are recorded:

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

Columns 1, 2, 3, 4 will give x_1, y_1, z_1
Columns 1, 2, 3, 5 will give x_2, y_2, z_2
Columns 1, 2, 3, 6 will give x_3, y_3, z_3

The matrix represents a system of linear simultaneous equations with three right-hand-sides. They may be solved by any suitable method, but Gauss elimination is recommended. (Gauss elimination uses combinations of rows.) You are expected to know this method, so little explanation will be given.

$$\begin{array}{c} \\ (2)+2(1) \\ (3)-(1) \end{array}\begin{pmatrix} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 2 & 1 & 0 \\ 0 & -1 & 3 & -1 & 0 & 1 \end{pmatrix} \Rightarrow \begin{array}{c} \\ \\ (2)+(3) \end{array}\begin{pmatrix} 1 & 0 & -1 & 1 & 0 & 0 \\ 0 & 1 & -2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Now eliminate variables above the diagonal.

$$\begin{array}{c} (1)+(3) \\ (2)+2(3) \\ \\ \end{array}\begin{pmatrix} 1 & 0 & 0 & 2 & 1 & 1 \\ 0 & 1 & 0 & 4 & 3 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

This gives the result $\mathbf{A}^{-1} = \begin{pmatrix} 2 & 1 & 1 \\ 4 & 3 & 2 \\ 1 & 1 & 1 \end{pmatrix}$ which can be checked by calculating $\mathbf{AA}^{-1}$.

The recipe is:

1. Write down the matrix **A**
2. To the right of **A** write down the identity matrix of the same order.
3. Use Gauss elimination to reduce **A** to the identity matrix.
4. The original identity matrix has been replaced by the inverse of **A**.

We can now use the inverse matrix to solve the following equations:

$$\begin{aligned} x \quad\quad\;\; - z &= 5 \\ -2x + y \quad\quad\;\; &= 1 \\ x - y + 2z &= -10 \end{aligned} \qquad \text{let } \mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ -2 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 5 \\ 1 \\ -10 \end{pmatrix}$$

then the equations are $\mathbf{A\,x} = \mathbf{b}$ and the solution is $\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{b}$

i.e. $\mathbf{x} = \begin{pmatrix} 2 & 1 & 1 \\ 4 & 3 & 2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 5 \\ 1 \\ -10 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix}$ – check by substitution into the equations.

It is clear there is more work in calculating the inverse than in solving the equations directly. **You do not calculate an inverse matrix merely to solve equations!**

A formula for the inverse matrix

It is sometimes convenient to express the inverse matrix as a formula. Consider the square matrix $\mathbf{A} = (a_{ij})$.

For each element a_{ij} there corresponds a **cofactor** c_{ij} obtained by evaluating the determinant resulting from the elimination of the ith row and jth column together with its 'place sign' = $(-1)^{i+j}$.

e.g. for $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ some cofactors c_{ij} are $c_{11} = +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$ $\quad c_{12} = -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$

$$c_{21} = -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} \qquad c_{22} = +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$$

Let the matrix of cofactors be $\mathbf{C} = (c_{ij})$ then the adjoint matrix is $\text{adj}(\mathbf{A}) = \mathbf{C}^T$

and the inverse matrix is: $\boxed{\mathbf{A}^{-1} = \dfrac{\text{adjoint}}{\text{determinant}} = \dfrac{\text{adj}(\mathbf{A})}{\det(\mathbf{A})}}$

Let us use this formula to confirm the inverse matrix obtained previously:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ -2 & 1 & 0 \\ 1 & -1 & 2 \end{pmatrix} \qquad C = \begin{pmatrix} 2 & 4 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 1 \end{pmatrix} \qquad \det(\mathbf{A}) = 2 - 2 + 1 = 1$$

$$\mathbf{A}^{-1} = \frac{\text{adj}(\mathbf{A})}{\det(\mathbf{A})} = \begin{pmatrix} 2 & 1 & 1 \\ 4 & 3 & 2 \\ 1 & 1 & 1 \end{pmatrix}$$

Note that if $\det(\mathbf{A}) = 0$ there is no inverse and the matrix is said to be **singular**.

Rank of a matrix

The rank of a matrix is a measure of the information it contains. It is the number of independent rows or columns of the matrix. It is most efficiently determined by Gauss elimination:

$$\begin{pmatrix} 1 & 1 & 0 & -2 \\ 2 & 0 & 2 & 2 \\ 4 & 1 & 3 & 1 \end{pmatrix} \Rightarrow \begin{matrix} \\ (2)-2(1) \\ (3)-4(1) \end{matrix} \begin{pmatrix} 1 & 1 & 0 & -2 \\ 0 & -2 & 2 & 6 \\ 0 & -3 & 3 & 9 \end{pmatrix} \Rightarrow \begin{matrix} \\ \\ (3)-3(2)/2 \end{matrix} \begin{pmatrix} 1 & 1 & 0 & -2 \\ 0 & -2 & 2 & 6 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The 3rd row can be expressed as a linear combination of the first two, so the rank is **2**.

Note that row (3) = $\frac{3}{2} \times (2) + (1)$ – it is a linear combination of the other rows and so does not contain any new information.

Underdetermined sets of equations

Sometimes there is insufficient information to provide a unique solution to the equations. Solve the following equations by Gauss elimination:

$$\begin{aligned} x &\; + 2y &\; + z &\; = 1 \\ 2x &\; - y &\; - 3z &\; = 2 \\ 4x &\; + 3y &\; - z &\; = 4 \end{aligned}$$

$$\begin{pmatrix} 1 & 2 & 1 & 1 \\ 2 & -1 & -3 & 2 \\ 4 & 3 & -1 & 4 \end{pmatrix} \Rightarrow \begin{array}{c} \\ (2)-2(1) \\ (3)-4(1) \end{array} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & -5 & -5 & 0 \\ 0 & -5 & -5 & 0 \end{pmatrix} \Rightarrow \begin{array}{c} \\ \\ (3)-(2) \end{array} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & -5 & -5 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The last equation becomes $0 = 0$, which is true but not very useful. We therefore have 3 unknowns but only 2 useful equations. This allows us to express 2 of the variables in terms of the 3rd:

Let $z = z$ then back substitution gives: $y = -z$, $x = 1 + z$.

We thus have an infinite number of solutions in terms of a single variable. This is because the equations are linearly dependent: (3) = 2 × (1) + (2)

Note that the determinant of the left-hand-side coefficients is zero: $\begin{vmatrix} 1 & 2 & 1 \\ 2 & -1 & -3 \\ 4 & 3 & -1 \end{vmatrix} = 0$

Inconsistent equations

Making a small alteration to the rhs of one equation gives: $\begin{aligned} x &\; + 2y &\; + z &\; = 1 \\ 2x &\; - y &\; - 3z &\; = 2 \\ 4x &\; + 3y &\; - z &\; = 0 \end{aligned}$

and solving them gives: $\begin{array}{c} \\ (2)-2(1) \\ (3)-4(1) \end{array} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & -5 & -5 & 0 \\ 0 & -5 & -5 & 4 \end{pmatrix} \Rightarrow \begin{array}{c} \\ \\ (3)-(2) \end{array} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & -5 & -5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$

The last equation is now $0 = 4$ which is false. The equations are therefore **inconsistent** and there is no solution. Note that the determinant is zero as before. The equations are inconsistent when the rank of the augmented matrix is greater than the rank of the lhs matrix.

Summary

det(**A**) ≠ 0 gives a unique solution.
det(**A**) = 0 gives <u>either</u> an infinite number of solutions (consistent equations).
<u>or</u> no solution (inconsistent equations).

Homogeneous equations

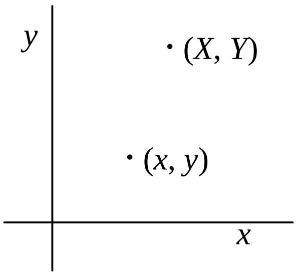
Homogeneous equations are **A x = 0**

giving the trivial solution **x = 0**

However, an infinite number of solutions becomes possible when det(**A**) = 0.

Linear transformations in a plane

We consider the linear transformation which moves the point (x, y) to (X, Y) and has the form



$$X = a_{11}\,x + a_{12}\,y$$

$$Y = a_{21}\,x + a_{22}\,y$$

or $$\begin{pmatrix} X \\ Y \end{pmatrix} = \mathbf{A}\begin{pmatrix} x \\ y \end{pmatrix}$$

We can regard the matrix $\mathbf{A}$ as an operator which transforms (x, y) into (X, Y).

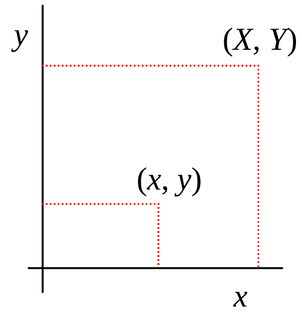
We could also think of $\begin{pmatrix} x \\ y \end{pmatrix}$ as the components of a vector and the operation changes both the magnitude and direction of the vector.

The operation performed by $\mathbf{A}^{-1}$ undoes the operation performed by $\mathbf{A}$:

$$\mathbf{A}^{-1}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

Dilation

Dilation changes the scales along the x and y axes:



$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x \\ 3y \end{pmatrix}$$

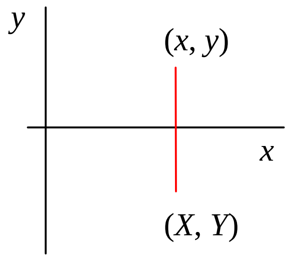
Let the dilation matrix be $\mathbf{D}_{a,b} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$

Then the inverse matrix is $\mathbf{D}_{a,b}^{-1} = \begin{pmatrix} 1/a & 0 \\ 0 & 1/b \end{pmatrix}$

The product of two dilations is also a dilation and dilation matrices commute.

Reflection

A reflection in the x-axis is given by:



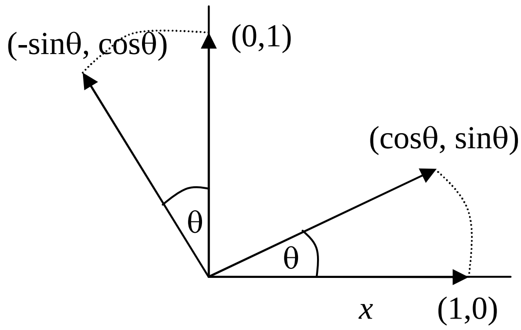
$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$$

Let the reflection matrix be $\mathbf{Q}$ and it is clear that $\mathbf{Q}^{-1} = \mathbf{Q}$ so that $\mathbf{Q}^2 = \mathbf{I}$

Rotation

Consider the rotation of the unit vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ by an angle θ about the origin.

It can be seen from the diagram that if the rotation is performed by a 2×2 matrix **R** then



$$\mathbf{R}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} \quad \text{and} \quad \mathbf{R}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}$$

Let $\mathbf{R} = \begin{pmatrix}a & b\\ c & d\end{pmatrix}$ then

$$\mathbf{R}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}a\\c\end{pmatrix} = \begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix} \quad \text{and} \quad \mathbf{R}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}b\\d\end{pmatrix} = \begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}$$

so the rotation matrix is $\boxed{\mathbf{R}_\theta = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}}$

and this performs a positive rotation of an object about the origin.

The inverse is a rotation of –θ giving $\mathbf{R}_\theta^{-1} = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}$

Clearly, $\mathbf{R}^{-1} = \mathbf{R}^{T}$ i.e. $\mathbf{R}_\theta^T\,\mathbf{R}_\theta = \mathbf{I}$

A matrix with this property is said to be **orthonormal** – the dot product of any pair of columns is zero (they are orthogonal) and regarding a column as elements of a vector, they all have a magnitude of unity.

Combination of operations – two rotations

A rotation of θ, given by $\mathbf{R}_\theta$, followed by a rotation of ϕ, $\mathbf{R}_\phi$, is represented by

$$\mathbf{R}_\theta\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}X\\Y\end{pmatrix} \quad \text{then} \quad \mathbf{R}_\phi\begin{pmatrix}X\\Y\end{pmatrix} = \begin{pmatrix}X'\\Y'\end{pmatrix} \quad \Rightarrow \quad \mathbf{R}_\phi\,\mathbf{R}_\theta\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}X'\\Y'\end{pmatrix}$$

and the complete rotation is given by $\mathbf{R}_\phi\mathbf{R}_\theta$ – note the order of the matrices.

$$\mathbf{R}_\phi\,\mathbf{R}_\theta = \begin{pmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} = \begin{pmatrix}\cos\theta\cos\phi - \sin\theta\sin\phi & -\sin\theta\cos\phi - \cos\theta\sin\phi\\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi\end{pmatrix}$$

Geometrically, the composite rotation is through an angle of θ+ϕ and the matrix is

$$\mathbf{R}_{\theta+\phi} = \begin{pmatrix}\cos(\theta+\phi) & -\sin(\theta+\phi)\\ \sin(\theta+\phi) & \cos(\theta+\phi)\end{pmatrix}$$

Equating elements of these matrices gives the familiar trigonometric identities:

$\cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi$ and $\sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi$

It is also clear that: $\mathbf{R}_\theta\,\mathbf{R}_\phi = \mathbf{R}_\phi\,\mathbf{R}_\theta = \mathbf{R}_{\theta+\phi}$

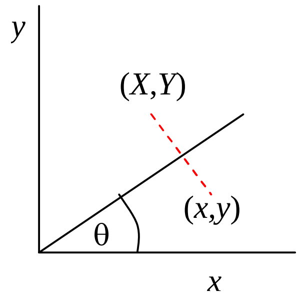
However, be careful – rotations only commute if they are about the same axis.

Combination of operations – reflection through any line

Let us derive the matrix representing a reflection in the line making an angle θ with the x-axis:

The combination of operations required is:

1. Rotate the line to the x-axis.
2. Reflect in the x-axis.
3. Rotate the line back to its original position.



If $\mathbf{Q}_\theta$ is the desired matrix, it is obtained by:

$$\mathbf{Q}_\theta = \mathbf{R}_\theta\, \mathbf{Q}_0\, \mathbf{R}_{-\theta} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos^2\theta - \sin^2\theta & 2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \sin^2\theta - \cos^2\theta \end{pmatrix}$$

$$\boxed{\therefore\ \mathbf{Q}_\theta = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}}$$

Combination of operations – two reflections

Let us now derive the composite of two reflections and interpret the result:

$$\mathbf{Q}_\phi\, \mathbf{Q}_\theta = \begin{pmatrix} \cos 2\phi & \sin 2\phi \\ \sin 2\phi & -\cos 2\phi \end{pmatrix}\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$$

$$= \begin{pmatrix} \cos 2\phi\cos 2\theta + \sin 2\phi\sin 2\theta & \cos 2\phi\sin 2\theta - \sin 2\phi\cos 2\theta \\ \sin 2\phi\cos 2\theta - \cos 2\phi\sin 2\theta & \sin 2\phi\sin 2\theta + \cos 2\phi\cos 2\theta \end{pmatrix}$$

$$\therefore\ \mathbf{Q}_\phi\, \mathbf{Q}_\theta = \begin{pmatrix} \cos(2\phi - 2\theta) & -\sin(2\phi - 2\theta) \\ \sin(2\phi - 2\theta) & \cos(2\phi - 2\theta) \end{pmatrix}$$

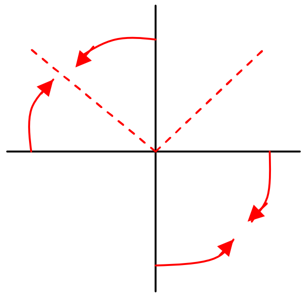
This is a rotation matrix corresponding to a positive rotation of 2(ϕ-θ), so that

$$\mathbf{Q}_\phi\, \mathbf{Q}_\theta = \mathbf{R}_{2(\phi-\theta)} \qquad \text{also} \qquad \mathbf{Q}_\theta\, \mathbf{Q}_\phi = \mathbf{R}_{2(\theta-\phi)}$$

If the order of the reflections is reversed, it represents a rotation in the opposite direction, i.e. reflections do not commute.

Eigenvectors and eigenvalues

A general linear transformation in the plane is expected to change the magnitude and direction of any vector on which it operates. However, if we try several transformations and plot them, a pattern emerges:



$$\begin{pmatrix} 3 & -2 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \quad \text{similarly} \quad \begin{pmatrix} -1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} -3 \\ 1 \end{pmatrix}$$

$$\text{also} \quad \begin{pmatrix} 3 & -2 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ -1 \end{pmatrix} \rightarrow \begin{pmatrix} 2 \\ -2 \end{pmatrix}$$

Rotations occur in opposite senses. There must therefore be boundaries between opposite rotations where the sense of rotation is reversed and a vector on the boundary is not rotated.

Consider these special cases:

$$\begin{pmatrix} 3 & -2 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -8 \\ 4 \end{pmatrix} = 4\begin{pmatrix} -2 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 & -2 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Notice that the vectors are not rotated. These special cases are of the form:

$$\mathbf{A}\ \mathbf{x} = \lambda\ \mathbf{x}$$

Square matrix Vector Scalar Vector

The vectors and scalars which satisfy this relationship are the **eigenvectors** and **eigenvalues** of the matrix.

Determination of eigenvalues

Let us determine the values of λ which satisfy the above relationship:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda\begin{pmatrix} x \\ y \end{pmatrix} \quad \Rightarrow \quad \begin{matrix} ax + by = \lambda x \\ cx + dy = \lambda y \end{matrix} \quad \Rightarrow \quad \begin{matrix} (a-\lambda)\,x + by = 0 \\ cx + (d-\lambda)\,y = 0 \end{matrix}$$

These equations are most conveniently expressed in matrix notation as:

$(\mathbf{A} - \lambda\ \mathbf{I})\ \mathbf{x} = \mathbf{0}$ which is a system of linear simultaneous equations.

Since the right-hand-side of the equations is zero, the equations have the trivial solution that $\mathbf{x} = \mathbf{0}$. This is clearly not the required result.

Instead of seeking a unique solution, an infinite number of solutions may be obtained if $\det(\mathbf{A}-\lambda\mathbf{I}) = 0$, thus avoiding the trivial solution.

For the matrix above, we have $\begin{vmatrix} 3-\lambda & -2 \\ -1 & 2-\lambda \end{vmatrix} = 0 \quad \Rightarrow \quad (3-\lambda)(2-\lambda)-2 = 0$

This gives $\lambda^2 - 5\lambda + 4 = 0 \quad \Rightarrow \quad (\lambda-1)(\lambda-4) = 0 \quad \Rightarrow \quad \lambda = 1 \text{ or } 4$

Clearly, the determinant represents a polynomial in λ of degree equal to the order of the matrix. Such a polynomial will have *n* roots so that **an order *n* matrix has *n* eigenvalues and *n* corresponding eigenvectors**.

The polynomial is known as the **characteristic equation** of the matrix.

Note that the eigenvalues we have obtained are precisely the values in the previous numerical example.

Determination of eigenvectors

Given values for λ, let us now determine the corresponding eigenvectors. These are obtained by solving the equations $(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0}$ for $\mathbf{x}$.

For λ = 1: $$\begin{pmatrix} 3-1 & -2 \\ -1 & 2-1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The equations are: $$\begin{aligned} 2x - 2y &= 0 \\ -x + y &= 0 \end{aligned}$$ Note that the second equation is merely a repeat of the first. This is to be expected – the equations must be linearly dependent because the determinant is zero.

We can therefore determine y in terms of x, giving the solution $x = x$, $x = y$ where x is arbitrary. So, with $x = 1$, the eigenvector is $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

This determines the direction of the eigenvector, but not its magnitude.

For λ = 4: $$\begin{pmatrix} -1 & -2 \\ -1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \begin{aligned} -x - 2y &= 0 \\ -x - 2y &= 0 \end{aligned} \Rightarrow x = x,\ x = -2y$$

So, with $y = 1$, the eigenvector is $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

Note that these are the eigenvectors in the numerical example.

Since the relationship $\mathbf{Ax} = \lambda\mathbf{x}$ does not define the magnitude of $\mathbf{x}$, it is conventional to normalise the vectors to a magnitude of unity. The two eigenvectors we have just determined should therefore be given as $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

In general, the elements of the vector are scaled such that $\sum_{i=1}^{n} x_i^2 = 1$ or, in matrix notation, $\mathbf{x}^{\mathrm{T}}\,\mathbf{x} = 1$.

The eigenvalues of a Hermitian (or symmetric) matrix are real

The eigenvalues of a matrix are the roots of a polynomial and therefore may be complex even if all the elements in the matrix are real. However, for symmetric and Hermitian matrices, the eigenvalues are real, as shown here.

The Hermitian transpose of a matrix is obtained as the transpose of its complex conjugate. So, if $\mathbf{A}^{\mathrm{H}} = \mathbf{A}$ then the matrix is Hermitian, i.e. $a_{ij} = a_{ji}^*$

Start with the definition of eigenvectors and eigenvalues: $\mathbf{A}\,\mathbf{x}_i = \lambda_i\,\mathbf{x}_i$

Take the Hermitian transpose of both sides: $(\mathbf{A}\,\mathbf{x}_i)^H = \mathbf{x}_i^H\,\mathbf{A}^H = \lambda_i^*\,\mathbf{x}_i^H$

Postmultiply by $\mathbf{x}_i$ and use $\mathbf{A}^{\mathrm{H}} = \mathbf{A}$: $\mathbf{x}_i^H\,\mathbf{A}\,\mathbf{x}_i = \lambda_i^*\,\mathbf{x}_i^H\,\mathbf{x}_i$

Now premultiply the first equation by $\mathbf{x}_i^H$: $\mathbf{x}_i^H \mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i^H \mathbf{x}_i$

A comparison of the last two equations shows that $\lambda_i = \lambda_i^*$ so λ_i must be real.

The eigenvectors of a Hermitian (or symmetric) matrix are orthogonal

If $\mathbf{x}_i$ and $\mathbf{x}_j$ are two vectors, the condition for orthogonality is $\mathbf{x}_i^H \mathbf{x}_j = 0$ for $i \neq j$.

Let $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$ then $\mathbf{x}_j^H \mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_j^H \mathbf{x}_i$

But $\left(\mathbf{A}^H \mathbf{x}_j\right)^H = \mathbf{x}_j^H \mathbf{A}$ $\therefore \left(\mathbf{A}^H \mathbf{x}_j\right)^H \mathbf{x}_i = \lambda_i \mathbf{x}_j^H \mathbf{x}_i$ . . . . (1)

Because $\mathbf{A}^H = \mathbf{A}$ (Hermitian matrix) and $\lambda_j = \lambda_j^*$ (eigenvalues are real) we have

$\mathbf{A}^H \mathbf{x}_j = \lambda_j^* \mathbf{x}_j$

Take the Hermitian transpose of both sides: $\left(\mathbf{A}^H \mathbf{x}_j\right)^H = \lambda_j \mathbf{x}_j^H$

and postmultiply by $\mathbf{x}_i$: $\left(\mathbf{A}^H \mathbf{x}_j\right)^H \mathbf{x}_i = \lambda_j \mathbf{x}_j^H \mathbf{x}_i$ . . . . (2)

From (1) and (2) $\therefore \lambda_j \mathbf{x}_j^H \mathbf{x}_i = \lambda_i \mathbf{x}_j^H \mathbf{x}_i$

$$\therefore \left(\lambda_j - \lambda_i\right)\mathbf{x}_j^H \mathbf{x}_i = 0$$

So if $\lambda_j \neq \lambda_i$ we have $\mathbf{x}_j^H \mathbf{x}_i = 0$ and the eigenvectors are orthogonal.

Modal matrix

Since $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$, we can write the complete set of eigenvectors and eigenvalues as

$$\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}$$

where the $\mathbf{x}_i$ are the columns of $\mathbf{X}$ and $\mathbf{\Lambda}$ is a diagonal matrix containing the λ_i.

X is known as the **modal matrix** of **A**.

When **A** is symmetric (or Hermitian), the columns of **X** are orthogonal and, in standard form, will be normalised to unit length. This makes **X** an orthonormal matrix so that $\mathbf{X}^{-1} = \mathbf{X}^T$.

Matrix diagonalisation

When $\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{\Lambda}$ then $\mathbf{X}^T \mathbf{A}\mathbf{X} = \mathbf{\Lambda}$ and this has **diagonalised** the matrix **A**.

The matrix **X** which performs this operation must have eigenvectors of the symmetric matrix **A** as columns and the diagonal matrix **Λ** contains the corresponding eigenvalues.

Conversely, a matrix which has a given set of eigenvalues and eigenvectors may be constructed from $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^T$.

This is known as a **similarity transform**. Matrices with the same set of eigenvalues are said to be **similar** and the eigenvalues of **A** and **Λ** are the same.

Properties of eigenvalues

1. Starting with the similarity transform $\mathbf{X}^{\mathrm{T}}\,\mathbf{A}\,\mathbf{X} = \mathbf{\Lambda}$ we evaluate the determinant of both sides. For the left-hand-side, we have:

$$\det(\mathbf{X}^{\mathrm{T}}\,\mathbf{A}\,\mathbf{X}) = \det(\mathbf{X}^{\mathrm{T}})\times\det(\mathbf{A})\times\det(\mathbf{X}) = \det(\mathbf{X}^{\mathrm{T}}\,\mathbf{X})\times\det(\mathbf{A}) = \det(\mathbf{A})$$

and the right-hand-side gives: $\det(\mathbf{\Lambda}) = \prod_{i=1}^{n} \lambda_i$

So the product of the eigenvalues is equal to the value of the determinant.

$$\boxed{\det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i}$$

2. The trace of a matrix is the sum of its diagonal elements: $\mathrm{trace}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$

Using the similarity transform again: $\mathrm{trace}(\mathbf{\Lambda}) = \sum_{i=1}^{n} \lambda_i = \mathrm{trace}(\mathbf{X}^T\,\mathbf{A}\,\mathbf{X})$

so let us determine $\mathrm{trace}(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X})$:

Let $\mathbf{AX} = \mathbf{B}$ then $b_{ij} = \sum_k a_{ik}\, x_{kj}$

Let $\mathbf{X}^{\mathrm{T}}\mathbf{B} = \mathbf{X}^{\mathrm{T}}\mathbf{AX} = \mathbf{C}$ then $c_{ij} = \sum_k x_{ik}^T\, b_{kj} = \sum_k x_{ki} \sum_l a_{kl}\, x_{lj}$

The diagonal element is $c_{ii} = \sum_k \sum_l x_{ki}\, x_{li}\, a_{kl}$

so the trace is $\sum_i c_{ii} = \sum_i \sum_k \sum_l x_{ki}\, x_{li}\, a_{kl}$

But $\sum_i x_{ki}\, x_{li} = 0 \quad k \neq l$ orthogonal eigenvectors

$= 1 \quad k = l$ vectors of unit length

i.e. $\sum_i x_{ki}\, x_{li} = \delta_{kl}$ which is the Kronecker delta.

$$\therefore \sum_i c_{ii} = \mathrm{trace}(\mathbf{X}^T\,\mathbf{A}\,\mathbf{X}) = \sum_k \sum_l \delta_{kl}\, a_{kl} = \sum_k a_{kk} = \mathrm{trace}(\mathbf{A})$$

We therefore have:

$$\boxed{\text{the sum of the eigenvalues = the trace of the matrix}}$$

Case study

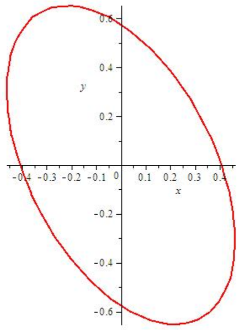
Whenever the properties of a physical system are expressed in terms of a matrix, the mathematical description of the system is nearly always simplified by diagonalising the matrix. A diagonal matrix clearly has fewer non-zero elements than one which is fully populated and so leads to a simpler mathematical model. This requires calculating the eigenvectors and eigenvalues of the matrix. Typically, the eigenvectors will define the directions of new (rotated) axes used to describe the system and the eigenvalues will give the values of essential parameters of the simplified model. This should become clearer by considering the following mathematical example.

The matrix: $\begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}$ The matrix is symmetrical (typical of a physical system with real eigenvalues and orthogonal eigenvectors) and is of order 2×2, so we are in two-dimensional space.

Quadratic form: Use the matrix to give the coefficients in an algebraic expression.

$$(x \quad y)\begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (6x+2y \quad 2x+3y)\begin{pmatrix} x \\ y \end{pmatrix} = 6x^2+4xy+3y^2$$

Ellipse: Equate the expression to a constant (= 1) and plot the curve. This gives an ellipse. Clearly, the equation we have used to describe the ellipse is not in standard form and you can see that the axes of the ellipse are not aligned with the coordinate axes. Let us seek a simpler description of the ellipse.



Eigenvalues: Solve det(**A** - λ**I**) = 0 for λ.

$$\begin{vmatrix} 6-\lambda & 2 \\ 2 & 3-\lambda \end{vmatrix} = 18-9\lambda+\lambda^2-4 = \lambda^2-9\lambda+14$$
$$= (\lambda-2)(\lambda-7) = 0$$

The characteristic equation factorises and gives two roots: λ = 2 or 7.

Eigenvectors: Substitute the eigenvalues into (**A** - λ**I**) **x** = 0 and solve for **x**.

For $\lambda = 2$: $(6-2)x + 2y = 0 \Rightarrow y = -2x$

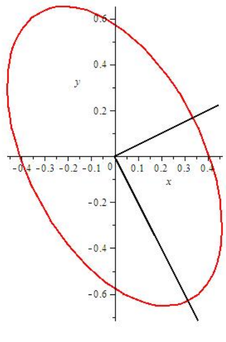
Only one equation is considered – underdetermined equations.

Let $x = 1$ then $y = -2$ so the normalised eigenvector is $\frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ -2 \end{pmatrix}$

For $\lambda = 7$: $(6-7)x + 2y = 0 \Rightarrow x = 2y$

Let $y = 1$ then $x = 2$ so the normalised eigenvector is $\frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}$

Plot the eigenvectors on the ellipse: To find the relationship between the eigenvectors and the ellipse, plot them on the graph. It is seen that the eigenvectors give the directions of the axes of the ellipse. If we rotate the coordinate axes to coincide with the directions of the eigenvectors, the ellipse will be in the standard orientation.



Rotate the axes: Set up the modal matrix (its columns are the eigenvectors) and note that it has the form of a rotation matrix. The new coordinates are then given by:

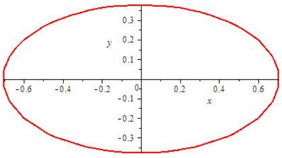
$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

Diagonalise the matrix: The above step is not normally necessary as the easy way to describe the ellipse on the rotated axes is to apply the rotation to the matrix itself, i.e. to diagonalise the matrix using a similarity transform, $\mathbf{X}^T\mathbf{A}\mathbf{X} = \mathbf{\Lambda}$. Note that the diagonal elements are the eigenvalues.

$$\frac{1}{5}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 7 \end{pmatrix}$$

Quadratic form on the new axes:

$$(x \quad y)\begin{pmatrix}2 & 0\\0 & 7\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = (2x \quad 7y)\begin{pmatrix}x\\y\end{pmatrix} = 2x^2 + 7y^2$$



Plot the ellipse on the new axes: Equate the above expression to a constant (= 1) and plot the graph. The axes of the ellipse now coincide with the coordinate axes and it has a simpler equation: $2x^2 + 7y^2 = 1$

The standard equation of an ellipse is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

We therefore have $\frac{1}{a^2} = 2$ so that $a = \frac{1}{\sqrt{2}} = 0.71$ and $b = \frac{1}{\sqrt{7}} = 0.38$

The graph confirms that these are the lengths of the semi-axes of the ellipse, so we have $length\ of\ semi-axis = \frac{1}{\sqrt{\lambda}}$

Cartesian tensors

Tensors come after vectors in the progression: scalar – vector – tensor. Scalars and vectors can also be called tensors. They are tensors of different **rank**. In 3-dimensional space, a scalar has $3^0 = 1$ components, a vector has $3^1 = 3$ components and a second rank tensor has $3^2 = 9$ components, which are written as a 3×3 array like a matrix. In general, a tensor of rank *n* has 3^n components, so a scalar is a tensor of rank 0 and a vector is a tensor of rank 1. We will consider mainly 2nd rank tensors referred to Cartesian axes and just call them tensors.

A useful application of tensors is to relate two vector quantities where the vectors are not necessarily parallel. Here are some examples in physics where this is the case:

Electrical conductivity
The microscopic version of Ohm's law is $\mathbf{J} = \sigma\, \mathbf{E}$ where the vector **J** is current density, **E** is electric field and σ is the conductivity. Taking the first components of each vector, we have $J_1 = \sigma\, E_1$ which is correct for an isotropic material (one whose properties are the same in all directions). However, if the conductivity is different in different directions (anisotropic material), **J** will depend upon components of the electric field in the *x*, *y* and *z* directions, i.e. $J_1 = \sigma_{11} E_1 + \sigma_{12} E_2 + \sigma_{13} E_3$.
Similarly $J_2 = \sigma_{21} E_1 + \sigma_{22} E_2 + \sigma_{23} E_3$
and $J_3 = \sigma_{31} E_1 + \sigma_{32} E_2 + \sigma_{33} E_3$
Therefore, in general, we have $J_i = \sum_{k=1}^{3} \sigma_{ik}\, E_k$ or $\mathbf{J} = \boldsymbol{\sigma}\, \mathbf{E}$ where $\boldsymbol{\sigma}$ is a 3×3 array, i.e. a second rank tensor.

Moment of inertia
Angular momentum, **L**, is related to the angular velocity, **ω**, by $\mathbf{L} = \mathrm{I}\, \boldsymbol{\omega}$, where I is moment of inertia. However, in the case of a tumbling object, the direction of **ω** will constantly change, whereas angular momentum is constant. Therefore, **L** and **ω** need not be parallel so the moment of inertia, I, must be a tensor quantity.

Electric permittivity
The electric flux density, **D**, is related to the electric field strength, **E**, by $\mathbf{D} = \varepsilon\, \mathbf{E}$, where ε is the electric permittivity. In an anisotropic medium, the polarisation need not be parallel to the field, so **D** and **E** need not be parallel. This requires ε to be a tensor. The **dielectric constant** contributes to the permittivity, so the dielectric constant is also a tensor quantity.

Stress
In the case of a wire supporting a weight, the stress (force per unit area) and the strain (fractional increase in length) are proportional to each other (Hooke's law). The force at a point in the wire can then be written as *force* = *stress* × *area*, i.e. $\mathbf{dF} = \mathrm{T}\, \mathbf{dS}$. However, for a 3-dimensional problem, there are not only normal forces (tensile stress) but also tangential forces (shear stress) which can twist or bend the body as well as stretch or compress it. This means that the force **dF** need not be parallel to the vector area, **dS**, so the stress, T, must be a tensor. Similarly, **strain** is also a tensor quantity.

We now need to look at some of the mathematical properties of tensors.

Addition and subtraction

Tensors can be added or subtracted like matrices, but in order to do so they must be of the same rank:

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \qquad\qquad a_{ij} + b_{ij} = c_{ij}$$

Multiplication

Tensors *do not* multiply like matrices. The most common multiplication is called a **direct product –** also known as an **outer product**.

If **U** and **V** are first rank tensors, then their direct product to obtain the tensor **W** can be written as: $u_i\, v_j = w_{ij}$ i.e. the direct product of two first rank tensors gives a second rank tensor.

If **U** is (u_1, u_2, u_3) and **V** is (v_1, v_2, v_3) then **W** is $\begin{pmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \end{pmatrix}$.

More generally, we may have $u_{ij}\, v_k = w_{ijk}$ so the direct product of a second rank tensor and a first rank tensor produces a third rank tensor. The rank of the resulting tensor is always the sum of the ranks of the original two tensors.

Transformation of vectors

Let **i**, **j**, **k** define the directions of the x, y, z axes and **i**', **j**', **k**' define the directions of the rotated system x', y', z'. The vector **r** can be written in terms of either set of components and basis vectors as

$$\mathbf{r} = x\,\mathbf{i} + y\,\mathbf{j} + z\,\mathbf{k} = x'\,\mathbf{i}' + y'\,\mathbf{j}' + z'\,\mathbf{k}'$$

The transformation between the two systems of axes is obtained as follows. Take the dot product of r with **i**':

$$\mathbf{r}\,.\,\mathbf{i}' = x\,\mathbf{i}\,.\,\mathbf{i}' + y\,\mathbf{j}\,.\,\mathbf{i}' + z\,\mathbf{k}\,.\,\mathbf{i}' = x'$$

Now $\mathbf{i}\,.\,\mathbf{i}'$, $\mathbf{j}\,.\,\mathbf{i}'$ and $\mathbf{k}\,.\,\mathbf{i}'$ are the cosines of the angles between the axes x and x', y and x' and z and x' respectively. Call these l_1, m_1 and n_1 so we have

$$x' = l_1\, x + m_1\, y + n_1\, z$$

Similarly, $$y' = l_2\, x + m_2\, y + n_2\, z$$

and $$z' = l_3\, x + m_3\, y + n_3\, z$$

These are the transformation equations between the coordinate system (x, y, z) and (x', y', z')

In the same way, dotting **r** with **i**, **j**, **k** in turn gives

$$x = l_1\, x' + l_2\, y' + l_3\, z'$$
$$y = m_1\, x' + m_2\, y' + m_3\, z'$$
$$z = n_1\, x' + n_2\, y' + n_3\, z'$$

These are more concisely expressed using matrix notation: $\mathbf{r}' = \mathbf{A}\,\mathbf{r}$ and $\mathbf{r} = \mathbf{A}^T\,\mathbf{r}'$

where **A** is the rotation matrix $\begin{pmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{pmatrix}$ which is orthonormal so that $\mathbf{A}^{-1} = \mathbf{A}^T$

These equations give the official mathematical definition of a vector. That is, a vector is a quantity that has components which transform according to these equations upon rotation of the axes.

Transformation of tensors

In the direct product of **U** and **V** above, these first rank tensors (vectors) transform as

$$\mathrm{u}'_k = \sum_{i=1}^{3} a_{ki}\, u_i \quad \text{and} \quad \mathrm{v}'_k = \sum_{i=1}^{3} a_{ki}\, v_i$$

so the components of the second rank tensor **UV** (= **W**) transform as

$$w'_{kl} = u'_k v'_l = \sum_{i=1}^{3} a_{ki}\, u_i \sum_{j=1}^{3} a_{lj}\, v_j = \sum_{i=1}^{3}\sum_{j=1}^{3} a_{ki}\, a_{lj}\, u_i\, v_j = \sum_{i=1}^{3}\sum_{j=1}^{3} a_{ki}\, a_{lj}\, w_{ij}$$

This relationship is easily generalised so, for example, a third rank tensor transforms as:

$$w'_{pqr} = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} a_{pi}\, a_{qj}\, a_{rk}\, w_{ijk}$$

Note that the components of a third rank tensor have three indices. These relationships give the formal mathematical definition of a tensor. That is, a tensor is a quantity that has components which transform according to these relationships upon rotation of the axes.

Summation convention

It should be clear by now that tensor equations make use of a lot of summation signs. It would be a simplification if we could survive without them. There is therefore a convention that the summation signs can be omitted so long as it is understood that a summation occurs over any index which appears exactly twice in one term.

The relationship $\mathrm{u}'_k = \sum_{i=1}^{3} a_{ki}\, u_i$ can therefore be abbreviated to $u'_k = a_{ki}\, u_i$

Thus a_{jj} means $a_{11} + a_{22} + a_{33}$

$x_i\, x_i$ means $x_1^2 + x_2^2 + x_3^2$

$a_{ij}\, b_{jk}$ means $a_{i1}\, b_{1k} + a_{i2}\, b_{2k} + a_{i3}\, b_{3k}$

Quotient rule

The quotient rule is used to determine whether some given quantities are the components of a tensor. To demonstrate it, we will prove that the electric permittivity of an anisotropic material is a tensor. It was previously stated that the electric flux density, **D**, was related to the electric field, **E**, in the presence of a dielectric material by $D_i = \varepsilon_{ij}\, E_j$

where ε is the permittivity. In a rotated coordinate system, we must have equations of the same form: $D'_i = \varepsilon'_{ij}\, E'_j$

but D'_i is a vector, so that $D'_i = a_{ik}\, D_k$ giving $a_{ik}\, D_k = \varepsilon'_{ij}\, E'_j$

and $D_k = \varepsilon_{kl}\, E_l$ $\quad \therefore\ a_{ik}\, \varepsilon_{kl}\, E_l = \varepsilon'_{ij}\, E'_j$

Similarly, E_l is a vector $E_l = a_{jl}\, E'_j$ giving $a_{ik}\, \varepsilon_{kl}\, a_{jl}\, E'_j = \varepsilon'_{ij}\, E'_j$

A rearrangement gives $\left(a_{ik}\, a_{jl}\, \varepsilon_{kl} - \varepsilon'_{ij}\right) E'_j = 0$

Since E'_j is arbitrary, this implies $\varepsilon'_{ij} = a_{ik}\, a_{jl}\, \varepsilon_{kl}$

and the permittivity ε_{kl} transforms as a tensor.

Contraction

Contraction of a tensor consists of setting two unlike indices equal to each other and then summing as implied by the summation convention. An example is afforded by the direct product of the two first rank tensors **U** and **V**. The general element of the product is $u_i\, v_j$. Making the indices the same gives $u_i\, v_i$, which, by the summation convention, gives the dot product, which is a scalar. It is more formally known as a **scalar invariant** as it does not change value upon rotation of axes. (The dot product is also known as a scalar product or an inner product.) The second rank tensor **UV** has therefore been contracted to a scalar, which is a tensor of zero rank. In general, contraction of a tensor reduces the rank by two.

A more general example of contraction is given by the transformation equations of a third rank tensor:

$$w'_{pqr} = a_{pi}\, a_{qj}\, a_{rk}\, w_{ijk}$$

Put $r = q$ which introduces an extra summation over q. We then have:

$$w'_{pqq} = a_{pi}\, a_{qj}\, a_{qk}\, w_{ijk}$$

Now $a_{qj}\, a_{qk}$ is the dot product of columns j and k of the rotation matrix **A**. Because the matrix is orthonormal, this dot product is 1 if $j = k$ and is 0 otherwise, i.e. $a_{qj}\, a_{qk} = \delta_{jk}$ (the Kronecker delta). Then $\delta_{jk}\, w_{ijk}$ becomes w_{ijj} since $\delta_{jk} = 0$ when $j \neq k$. This reduces the previous equation to:

$$w'_{pqq} = a_{pi}\, w_{ijj}$$

which is the transformation equation for a first rank tensor. Note that w_{ijj} has only one free index, i, since j is the dummy index of summation. The rank of the third rank tensor w_{ijk} has therefore been reduced by two.

This is the end of the course – I hope you have enjoyed it.