STEP Support Programme

STEP 2 Matrices Topic Notes

These notes are designed to help support the study of the matrices topic, which has been introduced into the new STEP syllabus (first examination in 2019) in light of the changes to the A-level specifications.

They begin with a brief summary of the relevant definitions and related results, which should mostly be familiar. There is an extensive section at the end of the notes on invariant points and lines.

Definitions

Elements: The individual items in a matrix (usually numbers, which might be unknown and represented by letters). Elements of a matrix are also known as *entries*.

Dimensions: The "size" of a matrix. A matrix with m rows and n columns has dimensions $m \times n$ ("m by n"). Most of the time you will be dealing with **square** matrices, which are ones which have the same number of rows and columns.

Conformable: For addition (or subtraction), two matrices are called conformable if they can be added (or subtracted), which is if they have the same dimensions. For multiplication, two matrices are called conformable if they can be multiplied in a specified order, which is if the number of columns in the first matrix is equal to the number of rows in the second matrix.

Zero Matrix: A **zero** or **null** matrix is one that has 0 for every element. It is often written as $\mathbf{O}$. Adding the (conformable) zero matrix to matrix $\mathbf{A}$ gives $\mathbf{A} + \mathbf{O} = \mathbf{O} + \mathbf{A} = \mathbf{A}$, whereas multiplying $\mathbf{A}$ by a conformable zero matrix results in a zero matrix: $\mathbf{AO} = \mathbf{OA} = \mathbf{O}$ (where the different zero matrices in this equation may have different dimensions).

Transpose: The **transpose** of a matrix $\mathbf{A}$ is what you get if you swap the rows and columns round (so that the first row becomes the first column and so on). It is written $\mathbf{A}^{\mathrm{T}}$.

Identity: An **identity** matrix (usually written as $\mathbf{I}$) is one which has 1's on the *leading diagonal* (i.e., the diagonal from top left to bottom right) and 0's everywhere else. An **identity** matrix must be a **square** matrix. If you multiply a matrix $\mathbf{A}$ by the (conformable) identity $\mathbf{I}$ then you get $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ (where the different identity matrices in this equation will have different dimensions if $\mathbf{A}$ is not square).

Determinant: Every square matrix has a number associated with it called its **determinant**. In the case of a 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the determinant is denoted by $\det \mathbf{A}$ or $|\mathbf{A}|$ and is given by the formula $\det \mathbf{A} = |\mathbf{A}| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$. Note that when writing out the determinant explicitly, we generally only write vertical lines; we wouldn't usually write $\left|\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right|$.

Inverse: The **inverse** of a square matrix $\mathbf{A}$ is the matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. In the case of a 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\mathbf{A}$ has an inverse if $\det \mathbf{A} \neq 0$, and in this case, the inverse is given by $\mathbf{A}^{-1} = \dfrac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \dfrac{1}{\det \mathbf{A}}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

Singular matrix: A square matrix $\mathbf{A}$ which has $\det \mathbf{A} = 0$ is called **singular**. If a matrix $\mathbf{A}$ is singular then the inverse $\mathbf{A}^{-1}$ does not exist. A matrix which is not singular is called **non-singular**.

Manipulating Matrices

To add or subtract two **conformable** matrices you add/subtract the corresponding elements of the two matrices, ending up with a matrix of the same dimensions as the original matrices.

To multiply a matrix by a **scalar**, multiply every element by the scalar.

Multiplying two matrices together: if matrix $\mathbf{A}$ has dimensions $p \times q$ and matrix $\mathbf{B}$ has dimensions $q \times r$ then matrix $\mathbf{AB}$ will have dimensions $p \times r$. Matrix $\mathbf{BA}$ will only exist if $p = r$. See here for an explanation of how to multiply two matrices together.

Note that matrix multiplication is **not commutative**, that is in general $\mathbf{AB} \neq \mathbf{BA}$, even if both sides are defined. Matrix multiplication is **associative**, that is we have $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ whenever either side is defined.

We can also expand brackets as with normal algebra, though we must now be careful about the order of multiplication. For example, assuming that all matrices are conformable:

$$\begin{aligned}
\mathbf{A}(\mathbf{B}+\mathbf{C}) &= \mathbf{AB}+\mathbf{AC} \\
(\mathbf{A}+\mathbf{B})\mathbf{C} &= \mathbf{AC}+\mathbf{BC} \\
(\mathbf{A}+\mathbf{B})^2 &= \mathbf{A}^2+\mathbf{AB}+\mathbf{BA}+\mathbf{B}^2
\end{aligned}$$

The first two of these say that multiplication *distributes* over addition. Note that we cannot, in general, simplify the right hand side of the third line, as $\mathbf{AB}$ and $\mathbf{BA}$ may be different.

If a matrix $\mathbf{A}$ has an inverse, the the inverse is unique. We show this by assuming that $\mathbf{A}$ has two different inverses, $\mathbf{B}$ and $\mathbf{C}$. If we can show that $\mathbf{B} = \mathbf{C}$, this will be a contradiction, so $\mathbf{A}$ can have at most one inverse. So to show that $\mathbf{B} = \mathbf{C}$, consider $\mathbf{BAC}$. By associativity, we have

$$\begin{aligned}
\mathbf{BAC} &= (\mathbf{BA})\mathbf{C} \\
&= \mathbf{IC} \qquad \text{as } \mathbf{B} \text{ is an inverse of } \mathbf{A} \\
&= \mathbf{C} \\
\text{and} \quad \mathbf{BAC} &= \mathbf{B}(\mathbf{AC}) \\
&= \mathbf{BI} \qquad \text{as } \mathbf{C} \text{ is an inverse of } \mathbf{A} \\
&= \mathbf{B}
\end{aligned}$$

so $\mathbf{B} = \mathbf{C}$, and $\mathbf{A}$ has at most one inverse.

If the matrices $\mathbf{A}$ and $\mathbf{B}$ are both **non-singular** then $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. Note that the order of the matrices is reversed. We can prove this as follows: the inverse of $\mathbf{AB}$ is a matrix $\mathbf{X}$ which satisfies $(\mathbf{AB})\mathbf{X} = \mathbf{X}(\mathbf{AB}) = \mathbf{I}$. Letting $\mathbf{X} = \mathbf{AB}$ gives:

$$\begin{aligned}
(\mathbf{AB})\mathbf{X} &= (\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) \\
&= \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} \\
&= \mathbf{AIA}^{-1} \\
&= \mathbf{AA}^{-1} \\
&= \mathbf{I}
\end{aligned}$$

and $\mathbf{X}(\mathbf{AB}) = \mathbf{I}$ similarly. Therefore the inverse of $\mathbf{AB}$ is $\mathbf{B}^{-1}\mathbf{A}^{-1}$.

Likewise, for any two conformable matrices $\mathbf{A}$ and $\mathbf{B}$, we have the identity $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$. This is relatively easy to show, just by thinking about how an element of $\mathbf{AB}$ is calculated, and what happens when the matrix is then transposed.

Another useful result is that for any square matrix $\mathbf{A}$, we have $\det \mathbf{A}^{\mathrm{T}} = \det \mathbf{A}$. A quick calculation shows that it is true for 2×2 matrices. (To prove this for $n \times n$ matrices requires a more general definition of determinant.)

The pair of simultaneous equations $ax + by = e$ and $cx + dy = f$ can be written in the form $\mathbf{Ax} = \mathbf{q}$ where $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ and $\mathbf{q} = \begin{pmatrix} e \\ f \end{pmatrix}$. If $\det \mathbf{A} \neq 0$ then the simultaneous equations have a unique solution given by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{q}$. If $\det \mathbf{A} = 0$, then $\mathbf{A}$ does not have an inverse. In this case, either the two lines represented by the equations are parallel and so do not meet, meaning that there are no solutions, or the two lines are the same line and there are infinitely many solutions.

We can see this algebraically: $\det \mathbf{A} = ad - bc$, so if $ad - bc = 0$, we have (assuming a and b are non-zero) $\frac{d}{b} = \frac{c}{a}$, so $c = ak$ and $d = bk$ for some k. Therefore the second equation is $akx + bky = f$, so (assuming $k \neq 0$) $ax + by = \frac{f}{k}$. If $\frac{f}{k} = e$, then the two equations are just multiples of each other and there are infinitely many solutions, while if $\frac{f}{k} \neq e$, the two lines are parallel and there are no solutions. The cases where $k = 0$ or $a = 0$ or $b = 0$ can be dealt with similarly.

Transformations

A transformation T is called a **linear transformation** if T has the two properties that $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ and $T(\lambda\mathbf{x}) = \lambda(T(\mathbf{x}))$ whenever $\mathbf{x}$ and $\mathbf{y}$ are vectors and λ is a scalar. Linear transformations fix the origin (i.e., a point at the origin stays where it is under the transformation) and transform straight lines into straight lines or points. (The proofs of these are given below.) Common examples of linear transformations include rotations about the origin and reflections about lines through the origin. Translations are not linear transformations as they do not fix the origin.

Matrices can be used to represent linear transformations (see the proof below). To find out what a 2-dimensional transformation matrix[1] does, it is often a good idea to look at the effect on the unit square. Apply the transformation matrix $\mathbf{A}$ to the vectors $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\end{pmatrix}$ and $\begin{pmatrix}1\\1\end{pmatrix}$ and you should be able to see what the transformation does by comparing the original square (the **object**) with its **image**.

Conversely, to find a transformation matrix for a given linear transformation consider the effect of the transformation on the vectors $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$. The image of $\begin{pmatrix}1\\0\end{pmatrix}$ will give the first column of the matrix and the image of $\begin{pmatrix}0\\1\end{pmatrix}$ will give the second column, as $\begin{pmatrix}a & b\\c & d\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}a\\c\end{pmatrix}$ and so on.

The matrix $\mathbf{AB}$ represents the transformation that results from the transformation represented by $\mathbf{B}$ **followed by** the transformation represented by $\mathbf{A}$ (note the order!). The inverse matrix $\mathbf{A}^{-1}$ will give the inverse transformation to $\mathbf{A}$. Note that some transformations (such as reflections) are **self-inverse**, i.e., $\mathbf{A}^{-1} = \mathbf{A}$.

The determinant of a 2×2 transformation matrix gives the area scale factor of the transformation. If the determinant is negative then the *orientation* of the shape has been reversed, such as what happens when the transformation is a reflection.

It follows from the last two paragraphs that the determinant of $\mathbf{AB}$ is the area scale factor of the transformation $\mathbf{AB}$, which is the product of the scale factors of $\mathbf{B}$ and $\mathbf{A}$, so we have the identity

$$\det(\mathbf{AB}) = (\det\mathbf{A})(\det\mathbf{B})$$

for any 2×2 square matrices $\mathbf{A}$ and $\mathbf{B}$. This identity is true for square matrices of any dimension.

Some proofs

It is unlikely that a STEP question would require candidates to produce a proof like the following. Nevertheless, STEP questions do frequently expect proofs, so it is valuable to study the following, asking yourself at each point: What is the purpose of this stage of the argument?

We first show that the origin is fixed by any linear transformation T.

[1] More correctly, we should say "a matrix representing a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$", where $\mathbb{R}^2$ is two-dimensional space. This is because we could have a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ (three-dimensional space), for example, which would be represented by a 3×2 matrix. But we will only be considering transformations from $\mathbb{R}^2$ to itself in STEP 2, so we will be imprecise and talk about 2-dimensional transformation matrices.

Let $\mathbf{x}$ be any vector, and note that $\mathbf{0} = 0\mathbf{x}$ (the zero vector, representing the origin). Then we have

$$T(0\mathbf{x}) = 0(T(\mathbf{x})) = \mathbf{0}$$

where we have used the rule $T(\lambda\mathbf{x}) = \lambda T(\mathbf{x})$ with $\lambda = 0$. Therefore $T(\mathbf{0}) = \mathbf{0}$.

Next, we show that straight lines are transformed into straight lines or points. Consider the straight line with vector equation $\mathbf{r} = \mathbf{a} + s\mathbf{b}$. Then an arbitrary point on this line is transformed to

$$T(\mathbf{r}) = T(\mathbf{a} + s\mathbf{b}) = T(\mathbf{a}) + T(s\mathbf{b}) = T(\mathbf{a}) + sT(\mathbf{b}),$$

using both $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ and $T(\lambda\mathbf{x}) = \lambda T(\mathbf{x})$. Thus the image of any arbitrary point lies on the line through $T(\mathbf{a})$ in the direction of $T(\mathbf{b})$ if this is non-zero, and the image of the line is just the single point $T(\mathbf{a})$ if $T(\mathbf{b}) = \mathbf{0}$.[2]

We now show that we can represent the (2-dimensional) linear transformation T by a matrix.[3] Let

$$\begin{pmatrix} a \\ c \end{pmatrix} = T\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) \quad \text{and} \quad \begin{pmatrix} b \\ d \end{pmatrix} = T\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$$

and then set $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so that $\mathbf{M}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = T\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$ and $\mathbf{M}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = T\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)$.

We want to show that $\mathbf{Mv} = T(\mathbf{v})$ for any vector $\mathbf{v}$. Letting $\mathbf{v} = \begin{pmatrix} r \\ s \end{pmatrix}$, we have:

$$\begin{aligned}
\mathbf{Mv} &= \mathbf{M}\left(r\begin{pmatrix} 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\
&= r\mathbf{M}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + s\mathbf{M}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= rT\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + sT\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\
&= T\left(r\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + T\left(s\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\
&= T\left(r\begin{pmatrix} 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\
&= T(\mathbf{v})
\end{aligned}$$

so the matrix $\mathbf{M}$ behaves exactly the same as the linear transformation T.[4]

[2] You might be asking yourself why we need to prove so many things in mathematics. One reason is that it helps us to avoid making invalid assumptions by forcing us to be explicit in our thinking. For example, in a draft of these notes, it was stated that "linear transformations transform straight lines into straight lines". We overlooked the possibility that a straight line might be transformed into a point. But by writing the proof down explicitly, it became clear that we had overlooked the case of $T(\mathbf{b}) = \mathbf{0}$.

[3] This argument can easily be extended to 3-dimensional linear transformations.

[4] After this, you may be wondering why we bothered: what, after all, is the difference between a linear transformation and a matrix? One sophisticated answer (which is beyond the requirements of STEP) is as follows: a linear transformation describes what happens to the points, and it only needs the ideas of an origin and vectors. It doesn't need any particular coordinate system. The matrix representing it, on the other hand, requires a coordinate system to be in place. If we choose a different coordinate system, then the matrix representing the linear transformation and the coefficients of any vectors may well be different, but the physical vectors involved will be unchanged.

As an example of different coordinate systems that you may well have come across without even realising it, imagine a particle on a rough sloping plane. When trying to work out the forces involved, one common coordinate system takes x to be horizontal and y to be vertical, and while another one takes x to be parallel to the plane and y to be perpendicular to the plane. The weight of the particle will have different representations as a vector in these two coordinate systems, even those it is same physical force.

2×2 Transformation Matrices

The following are 2×2 transformation matrices for standard transformations. They can all be deduced as described above.

Reflection in the x-axis: $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

Reflection in the y-axis: $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

Reflection in the line $y = x$: $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Reflection in the line $y = -x$: $\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$

Reflection in the line $y = x \tan\theta$: $\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$

Rotation about the origin by angle θ anticlockwise: $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

Enlargement centre the origin, scale factor λ: $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$

Stretch parallel to the x-axis, scale factor λ: $\begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}$

Stretch parallel to the y-axis, scale factor λ: $\begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$

Shear parallel to the x-axis, scale factor λ (x-axis invariant): $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$

Shear parallel to the y-axis, scale factor λ (y-axis invariant): $\begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$

3×3 Transformation Matrices

Only certain 3×3 transformation matrices are expected to be known for STEP papers 2 and 3; these are as follows.

Reflection in the plane...

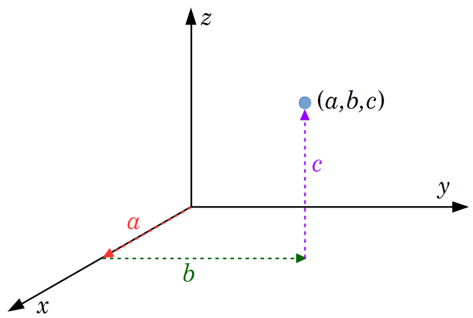
$x = 0$	$y = 0$	$z = 0$
$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$

For the following transformations, the direction of positive rotation is taken to be anticlockwise when looking towards the origin from the positive side of the axis of rotation.

Rotation of angle θ about the...

x-axis	y-axis	z-axis
$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$	$\begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}$	$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$

The 3 axes are arranged according to the "right-hand" convention, as shown in this diagram. This means that if you let your thumb be the x-axis, first finger the y-axis and second finger the y-axis, and make the three perpendicular to each other (without breaking or straining anything), then the fingers point in the directions of the axes.



Just as with 2×2 matrices, it is probably best not to try to remember all these, but to generate them from considering the effect of the transformation on the 3 unit vectors $\mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{j} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. The image of $\mathbf{i}$ is the first column of the transformation matrix, the image of $\mathbf{j}$ is the second column and the image of $\mathbf{k}$ is the third column.

Invariant Points and Lines

An **invariant point** of a transformation is one that satisfies $\mathbf{Ax} = \mathbf{x}$.

To find the invariant points, write $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ and then use $\mathbf{Ax} = \mathbf{x}$ to find 2 simultaneous equations in x and y. Use a similar technique in 3 (or more) dimensions.

Note that the point $(0, 0)$ is an invariant point for every transformation represented by a matrix, since $\mathbf{A}\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. If $\mathbf{A}$ is a 2×2 matrix, then any other invariant point will be on a line of invariant points, and this line will pass through the origin. If $\mathbf{A}$ is any matrix other than the identity, there can be at most one such line (though the proof of this fact is not part of the A-level course).

An **invariant line** is one where every point on the line goes to some point on the same line. Algebraically, this means either that every point on $y = mx+k$ goes to another point on $y = mx+k$, or for vertical lines, that every point on $x = q$ goes to a point on $x = q$.[5]

If we have $\mathbf{A}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix}$, then for an invariant (non-vertical) line we must have $y = mx + k$ and $Y = mX + k$. We can find possible values for m and k as shown in the examples below.

For a **vertical** line of the form $x = q$ to be an invariant line, then any point with x co-ordinate q must map to another point with x coordinate q.

In this case, if we set $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} q \\ y \end{pmatrix} = \begin{pmatrix} q \\ Y \end{pmatrix}$, then we have $aq+by = q$ and $cq+dy = Y$. For the first equation to be true for every value of y, we must have $b = 0$. Substituting $b = 0$ in this equation gives $aq = q$, or $(a-1)q = 0$. We must therefore have either $a = 1$ or $q = 0$. If $a = 1$, then the first equation holds for any q. If $q = 0$, then the second equation becomes $dy = Y$, which simply gives a formula for Y in terms of y.

In conclusion, the vertical line through the origin, $x = 0$, is an invariant line whenever the matrix has the form $\begin{pmatrix} a & 0 \\ c & d \end{pmatrix}$.

And if one vertical line not passing through the origin is an invariant line, then all vertical lines are invariant; this is the case if and only if the matrix has the form $\begin{pmatrix} 1 & 0 \\ c & d \end{pmatrix}$.

An **line of invariant points** is a line in which every point on the line stays where it is, whereas an **invariant line** is one where every point on the line stays on the line (but might move to a different position on the line). A line of invariant points is always an invariant line, but the reverse is not true.

[5]Textbooks often neglect to consider the vertical line case. Lines of the form $y = mx + k$ cannot be vertical. It is possible to consider lines of the form $ax + by = c$, which includes all straight lines, but the algebra can be somewhat trickier to handle. An alternative approach using this formula is given at the end of these notes. Note also that we have used k rather than c in the general formula $y = mx + k$ so as to not get confused with the c in the general matrix expression for $\mathbf{A}$. In the examples below c is not used in the expression for $\mathbf{A}$ so we can use $y = mx + c$ freely.

Examples

Example 1: Find the invariant **lines**[6] of $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}$.

First note that this matrix does not have any vertical invariant lines (either from the explanation above, or by noting that for a line of the form $x = k$ we would need $3k + 2y = k$ which is not true for all values of y).

Using $\mathbf{A} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix}$ we have:

$$3x + 2y = X \quad \text{and} \quad x + 2y = Y\,.$$

We also have:

$$y = mx + c \quad \text{and} \quad Y = mX + c\,.$$

Substituting for X and Y in the last equation gives $x + 2y = m(3x + 2y) + c$, and then substituting for y we have $x + 2(mx + c) = m(3x + 2(mx + c)) + c$. This can be rearranged as:

$$x(2m^2 + m - 1) + c(2m - 1) = 0\,. \tag{1}$$

In this equation x is a variable which can take any real value, whereas m and c are constants taking specific values (if unknown at the moment). For (1) to be equal to 0 for all values of x, we need $2m^2 + m - 1 = 0$ and also $c(2m - 1) = 0$.

The first of these conditions can be written as $(2m - 1)(m + 1) = 0$, i.e., $m = \frac{1}{2}$ or $m = -1$. When $m = \frac{1}{2}$, the second condition is satisfied for any value of c, while when $m = -1$ then the second condition gives $c = 0$. This means that there is one family of invariant lines and one other invariant line not in the family; the equations of these lines are

$$y = \tfrac{1}{2}x + c \quad \text{and} \quad y = -x\,.$$

Example 2: Find the invariant points of $\mathbf{A} = \begin{pmatrix} 4 & 1 & 0 \\ 3 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$.

Using $\mathbf{A} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ we have:

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

The first two equations each simplify to the equation $3x + y = 0$, and the third equation simplifies to $x + z = 0$. So if we take $x = \lambda$, we have $y = -3\lambda$ and $z = -\lambda$. We therefore have a line of invariant points of the form $(\lambda, -3\lambda, -\lambda)$.

[6] To find the invariant **points** of **A** you need to solve $\mathbf{A} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$. If you try to solve this, you should end up with $x + y = 0$, i.e., there is a line of invariant points with equation $y = -x$. This must therefore also be an invariant line.

Example 3: Find the invariant lines of $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$.

This looks very similar to Example 1, so we work in the same way. Again, we note that this matrix does not have any vertical invariant lines.

Now using $\mathbf{A}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix}$ we have:

$$3x + 2y = X \quad \text{and} \quad x + 4y = Y\,.$$

We also have:

$$y = mx + c \quad \text{and} \quad Y = mX + c\,.$$

Substituting for X and Y in the last equation gives $x + 4y = m(3x + 2y) + c$, and then substituting for y we have $x + 4(mx + c) = m(3x + 2(mx + c)) + c$. This can be rearranged as:

$$x(2m^2 - m - 1) + c(2m - 3) = 0\,.$$

For this to be equal to 0 for all values of x, we need $2m^2 - m - 1 = 0$ and also $c(2m - 3) = 0$.

The first of these conditions can be written as $(2m + 1)(m - 1) = 0$, i.e., $m = -\frac{1}{2}$ or $m = 1$. In both cases, the second condition gives $c = 0$. This means that there are exactly two invariant lines, which have equations:

$$y = -\tfrac{1}{2}x \quad \text{and} \quad y = x\,.$$

Example 4: Find the invariant lines of $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ -2 & 3 \end{pmatrix}$.

Once again, we note that this matrix does not have any vertical invariant lines. Working as before, using $\mathbf{A}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix}$ we have:

$$3x + 2y = X \quad \text{and} \quad -2x + 3y = Y\,.$$

We also have:

$$y = mx + c \quad \text{and} \quad Y = mX + c\,.$$

Substituting for X and Y in the last equation gives $-2x+3y = m(3x+2y)+c$, and then substituting for y we have $-2x + 3(mx + c) = m(3x + 2(mx + c)) + c$. This can be rearranged as:

$$x(2m^2 + 2) + c(2m - 2) = 0\,.$$

For this to be equal to 0 for all values of x, we need $2m^2 + 2 = 0$ and also $c(2m - 2) = 0$. But there are no real values of m which satisfy the first of these equations, so there are no invariant lines.

Geometrically, this matrix represents a composition of an enlargement and a rotation; this explains why there are no invariant lines.

An alternative approach to invariants (for interest only)

This more sophisticated approach is offered here to show another way of thinking about this. It is not required for the STEP examinations, and no STEP question will assume that you have read and understood this section of these notes.

In this section, we offer a more sophisticated way to find invariant points and lines, which gives a different perspective on what is going on. We start by building some theory, and then apply it to the examples in the previous section. We deal with invariant points and invariant lines through the origin first, and we consider the 2-dimensional case of invariant lines not passing through the origin afterwards.

You may well find that you need to read these notes several times with a pen and paper to hand to jot down thoughts and ideas as you go. This is a good mathematical skill to develop! As you read these notes, ask yourself questions such as: "Why are they doing this? What are they trying to achieve? How does this help move us towards our goal? Why does this work?" These will help you to develop a solid understanding of the approach shown here.

Invariant points and invariant lines through the origin

The first thing to note is that if $\mathbf{A}$ is a square matrix, $\mathbf{x}$ is a vector of the same dimension as $\mathbf{A}$ and k is a scalar, then

$$\mathbf{A}(k\mathbf{x}) = k(\mathbf{A}\mathbf{x}).$$

So if $\mathbf{x}$ is an invariant point, which means $\mathbf{A}\mathbf{x} = \mathbf{x}$, then so is $k\mathbf{x}$ for every k. Also, if $\mathbf{x}$ lies on an invariant line through the origin, which means that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $k\mathbf{x}$ lies on the same invariant line for every k.

These two cases can therefore be considered together: they both solve the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, where $\lambda = 1$ in the case of invariant points, and λ can be anything in the case of invariant lines.[7]

We now use a standard trick: we write $\mathbf{x} = \mathbf{I}\mathbf{x}$ on the right-hand side, where $\mathbf{I}$ is the identity matrix. This allows us to rearrange the equation as $\mathbf{A}\mathbf{x} - \lambda\mathbf{I}\mathbf{x} = \mathbf{0}$ (where the right-hand side is the zero vector), or:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}. \tag{2}$$

This always has a solution $\mathbf{x} = \mathbf{0}$, for any choice of λ; this says that $\mathbf{0}$ is always a fixed point.

We could consider solving the equation in (2) for a given λ. If $\mathbf{A} - \lambda\mathbf{I}$ is invertible, then we have $\mathbf{x} = (\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{0} = \mathbf{0}$, which brings us back to the zero vector again. The only way we can possibly get a non-zero vector $\mathbf{x}$ solving this is if $\mathbf{A} - \lambda\mathbf{I}$ is *singular*, which is when it has determinant 0. It turns out that if $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, then we are guaranteed to find a non-zero vector $\mathbf{x}$ solving the equation, though we do not attempt to prove this result here.

[7] This equation has proven to be so important that mathematicians have given names to its solutions. A value of $\mathbf{x}$ which solves the equation is called an *eigenvector*, and the corresponding λ is called its *eigenvalue*. The equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ itself is called the *characteristic equation of* $\mathbf{A}$.

We can therefore find invariant lines by following the two steps:

(i) Find the value(s) of λ which give $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$.

(ii) For each such value of λ, find the non-zero vectors $\mathbf{x}$ which solve (2).

We show examples of this process below.

Invariant lines in 2 dimensions not passing through the origin

We start with an observation. If we have two column vectors,

$$\mathbf{v} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \qquad \text{and} \qquad \mathbf{w} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

then we can find the scalar (dot) product of them:

$$\mathbf{v}.\mathbf{w} = x_1x_2 + y_1y_2.$$

We cannot multiply these vectors as if they were matrices, since they are not conformable. However, if we transpose $\mathbf{v}$, we discover that

$$\mathbf{v}^{\mathrm{T}}\mathbf{w} = \begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = (x_1x_2 + y_1y_2),$$

a 1×1 matrix whose single entry is $\mathbf{v}.\mathbf{w}$! It is often convenient to write $\mathbf{v}^{\mathrm{T}}\mathbf{w} = \mathbf{v}.\mathbf{w}$ and to treat the 1×1 matrix and its entry as the same thing, and we will use this idea below. (This also means that matrix multiplication can be thought of as finding lots of scalar products, though we will not pursue this idea further here.) We could also have written $\mathbf{v}$ and $\mathbf{w}$ in the opposite order, as $\mathbf{v}.\mathbf{w} = \mathbf{w}.\mathbf{v}$, so overall we have

$$\mathbf{v}^{\mathrm{T}}\mathbf{w} = \mathbf{w}^{\mathrm{T}}\mathbf{v} = \mathbf{v}.\mathbf{w}$$

Returning to the problem of finding invariant lines, let our invariant line have equation $mx+ny = p$, where m, n and p are real numbers, with m and n not both zero. If (x_0, y_0) is a point on this line which is transformed to (X_0, Y_0) by the action of $\mathbf{A}$, then we want:

$$\begin{aligned} mx_0 + ny_0 &= p \\ mX_0 + nY_0 &= p. \end{aligned}$$

If we subtract these two simultaneous equations, we obtain

$$m(X_0 - x_0) + n(Y_0 - y_0) = 0 \tag{3}$$

Now the two points (x_0, y_0) and (X_0, Y_0) are related by the transformation matrix $\mathbf{A}$, so

$$\begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} = \mathbf{A} \begin{pmatrix} x_0 \\ x_0 \end{pmatrix},$$

which gives

$$\begin{aligned}\begin{pmatrix} X_0 - x_0 \\ Y_0 - y_0 \end{pmatrix} &= \begin{pmatrix} X_0 \\ Y_0 \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\ &= \mathbf{A} \begin{pmatrix} x_0 \\ x_0 \end{pmatrix} - \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\ &= (\mathbf{A} - \mathbf{I}) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\end{aligned}$$

using the result $\mathbf{Iv} = \mathbf{v}$ for any vector $\mathbf{v}$.

Now we can use our earlier observation about scalar products to express (3) in matrix form; since we can express the left-hand side of (3) as a scalar product:

$$m(X_0 - x_0) + n(Y_0 - y_0) = \begin{pmatrix} m \\ n \end{pmatrix} \cdot \begin{pmatrix} X_0 - x_0 \\ Y_0 - y_0 \end{pmatrix},$$

we can write (3) as

$$\begin{pmatrix} m & n \end{pmatrix} \begin{pmatrix} X_0 - x_0 \\ Y_0 - y_0 \end{pmatrix} = 0.$$

Now using the above calculation, this becomes

$$\begin{pmatrix} m & n \end{pmatrix} (\mathbf{A} - \mathbf{I}) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = 0.$$

This means that the (transpose of the) vector $\begin{pmatrix} m & n \end{pmatrix} (\mathbf{A} - \mathbf{I})$ must be perpendicular to $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ for every point (x_0, y_0) on the invariant line. Since this invariant line does not pass through the origin, this means that the vector $\begin{pmatrix} m & n \end{pmatrix} (\mathbf{A} - \mathbf{I})$ must be the zero vector (otherwise it could not be perpendicular to every point on the line). Solving the equation

$$\begin{pmatrix} m & n \end{pmatrix} (\mathbf{A} - \mathbf{I}) = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

for non-zero matrices $\begin{pmatrix} m & n \end{pmatrix}$ turns out to be straightforward, if it is possible at all, as we will see in the examples below. In this case, $mx + ny = p$ will be an invariant line for every possible value of p.

So when does this equation have a non-zero solution? It helps to understand what is going on if we take the transpose of the equation, so that we are dealing with column vectors instead of row vectors. Remembering that for any matrices $\mathbf{B}$ and $\mathbf{C}$, we have $(\mathbf{BC})^{\mathrm{T}} = \mathbf{C}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}$, we get

$$(\mathbf{A} - \mathbf{I})^{\mathrm{T}} \begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

As we noted earlier, this equation has a non-zero solution if and only if the matrix $(\mathbf{A} - \mathbf{I})^{\mathrm{T}}$ is singular, that is, if and only if it has zero determinant. Since for any 2×2 matrix $\mathbf{B}$, we have $\det \mathbf{B}^{\mathrm{T}} = \det \mathbf{B}$, we see that the equation has a non-zero solution if and only if $\det(\mathbf{A} - \mathbf{I}) = 0$. But we have already found all the values of λ for which $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, so we already know whether or not the equation has any non-zero solutions.

Examples revisited

We use these techniques to present an alternative solution to the examples above.

Example 1: Find the invariant **lines** of $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}$.

As discussed above, we have two problems: one to find the invariant lines passing through the origin, the other to find invariant lines not necessarily passing through the origin.

To find the invariant lines passing through the origin, we first solve $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. This gives

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

which expands to $(3-\lambda)(2-\lambda) - 2 = 0$ or $\lambda^2 - 5\lambda + 4 = 0$, giving $\lambda = 1$ or $\lambda = 4$.

In the case $\lambda = 1$, we solve $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}$, which expands as

$$\begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and therefore $2x + 2y = x + y = 0$, so $y = -x$ is an invariant line, which happens to consist of invariant points (as $\lambda = 1$).

Likewise, when $\lambda = 4$, we have

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

giving $-x + 2y = x - 2y = 0$, and therefore $y = \frac{1}{2}x$ is also an invariant line (but the points on it other than the origin are not invariant).

To find any invariant lines which do not pass through the origin, we use the theory we developed above. Since $\lambda = 1$ is a solution of $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, we know that we have a family of invariant lines. So we solve $\begin{pmatrix} m & n \end{pmatrix} (\mathbf{A} - \mathbf{I}) = \begin{pmatrix} 0 & 0 \end{pmatrix}$, which in this case becomes

$$\begin{pmatrix} m & n \end{pmatrix} \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

The two columns each give the equation $2m + n = 0$, so we can take $m = 1$, $n = -2$, giving the family of invariant lines $x - 2y = p$, which can be rearranged to give $y = \frac{1}{2}x + c$ if we wish; here, c and p are arbitrary.

Example 2: Find the invariant points of $\mathbf{A} = \begin{pmatrix} 4 & 1 & 0 \\ 3 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$.

As we noted above, the invariant points satisfy $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}$, which gives

$$\begin{pmatrix} 3 & 1 & 0 \\ 3 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

As in the first approach, the first two rows both give $3x + y = 0$ while the third gives $x + z = 0$, and we therefore have a line of invariant points of the form $(k, -3k, -k)$.

Example 3: Find the invariant lines of $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$.

Again, we start with invariant lines passing through the origin. We solve $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, which gives

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

which expands to $(3-\lambda)(4-\lambda) - 2 = 0$ or $\lambda^2 - 7\lambda + 10 = 0$, giving $\lambda = 2$ or $\lambda = 5$.

In the case $\lambda = 2$, we solve $(\mathbf{A} - 2\mathbf{I})\mathbf{x} = \mathbf{0}$, which expands as

$$\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and therefore $x + 2y = 0$, so $y = -\frac{1}{2}x$ is an invariant line.

Likewise, when $\lambda = 5$, we have

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

giving $-2x + 2y = x - y = 0$, and therefore $y = x$ is a second invariant line.

Since $\lambda = 1$ was not a solution of the equation $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, there are no invariant lines which do not pass through the origin, so we are done.

We might also observe that the only invariant point in this case is the origin (which is always an invariant point): since $\lambda = 1$ is not a solution of $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, the equation $\mathbf{A}\mathbf{x} = \mathbf{x}$ has no non-zero solutions for $\mathbf{x}$.

Example 4: Find the invariant lines of $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ -2 & 3 \end{pmatrix}$.

We again start by solving $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$, which gives

$$\det\begin{pmatrix} 3-\lambda & 2 \\ -2 & 3-\lambda \end{pmatrix} = 0,$$

which expands to $(3-\lambda)(3-\lambda) + 4 = 0$ or $\lambda^2 - 6\lambda + 13 = 0$. Since this has negative discriminant, there are no real values of λ which give a zero determinant, and so there are no invariant lines.