

Week 1 — Vectors and Matrices

Richard Earl*
Mathematical Institute, Oxford, OX1 2LB,

October 2003

Abstract

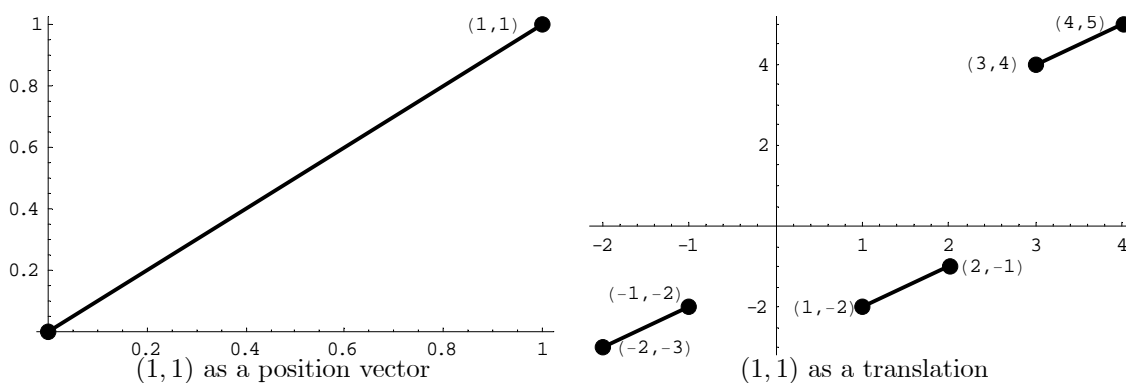
Algebra and geometry of vectors. The algebra of matrices. 2x2 matrices. Inverses. Determinants. Simultaneous linear equations. Standard transformations of the plane.

Notation 1 The symbol \mathbb{R}^2 denotes the set of ordered pairs (x, y) — that is the xy -plane. Similarly \mathbb{R}^3 denotes the set of ordered triples (x, y, z) — that is, three-dimensional space described by three co-ordinates x, y and z — and \mathbb{R}^n denotes a similar n -dimensional space.

1 Vectors

A *vector* can be thought of in two different ways. Let's for the moment concentrate on vectors in the xy -plane.

- From one point of view a vector is just an ordered pair of numbers (x, y) . We can associate this vector with the point in \mathbb{R}^2 which has co-ordinates x and y . We call this vector the *position vector* of the point.
- From the second point of view a vector is a 'movement' or translation. For example, to get from the point $(3, 4)$ to the point $(4, 5)$ we need to move 'one to the right and one up'; this is the same movement as is required to move from $(-2, -3)$ to $(-1, -2)$ or from $(1, -2)$ to $(2, -1)$. Thinking about vectors from this second point of view, all three of these movements are the same vector, because the same translation 'one right, one up' achieves each of them, even though the 'start' and 'finish' are different in each case. We would write this vector as $(1, 1)$. Vectors from this second point of view are sometimes called *translation vectors*.



*These handouts are produced by Richard Earl, who is the Schools Liaison and Access Officer for mathematics, statistics and computer science at Oxford University. Any comments, suggestions or requests for other material are welcome at earl@maths.ox.ac.uk

Likewise in three (or higher) dimensions the triple (x, y, z) can be thought of as the point in \mathbb{R}^3 , which is x units along the x -axis, y units along the y -axis and z units along the z -axis, or it can represent the translation which would take the origin to that point.

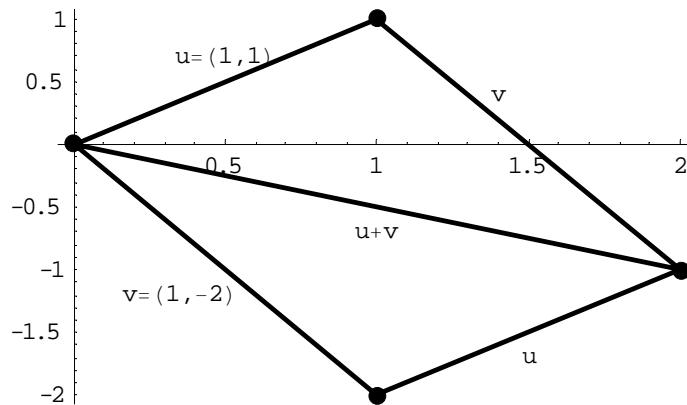
Notation 2 For ease of notation vectors are often denoted by a single letter, but to show that this is a vector quantity, rather than just a single number, the letter is either underlined such as \underline{v} or written in bold as \mathbf{v} .

1.1 Algebra of Vectors

Again we return to xy -plane. Given two vectors $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$ we can add them, much as you would expect, by adding their co-ordinates. That is:

$$\mathbf{v} + \mathbf{w} = (v_1, v_2) + (w_1, w_2) = (v_1 + w_1, v_2 + w_2).$$

This is easiest to see from a diagram:



A geometric interpretation of the vector sum

The sum is also easiest to interpret when we consider the vectors as translations. The translation $(v_1 + w_1, v_2 + w_2)$ is the composition of doing the translation (v_1, w_1) first and then doing the translation (v_2, w_2) or it can be achieved by doing the translations in the other order — that is, vector addition is *commutative*: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$. Note that it makes sense to add two vectors in \mathbb{R}^2 , or two vectors from \mathbb{R}^3 etc., but that we can make no obvious sense of adding a vector in \mathbb{R}^2 to one from \mathbb{R}^3 — they both need to be of the same type.

Given a vector $\mathbf{v} = (v_1, v_2)$ and a real number (a scalar) k then the *scalar multiple* $k\mathbf{v}$ is defined as

$$k\mathbf{v} = (kv_1, kv_2).$$

When k is a positive integer then we can think of $k\mathbf{v}$ as the translation achieved when we translate by \mathbf{v} k times. Note that the points $k\mathbf{v}$, as k varies, make up the line which passes through the origin and the point \mathbf{v} .

We write $-\mathbf{v}$ for $(-1)\mathbf{v} = (-a, -b)$ and this is the inverse operation of translating by \mathbf{v} . And the difference of two vectors $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = (w_1, w_2)$ is defined as

$$\mathbf{v} - \mathbf{w} = \mathbf{v} + (-\mathbf{w}) = (v_1 - w_1, v_2 - w_2).$$

Put another way $\mathbf{v} - \mathbf{w}$ is the vector that translates the point with position vector \mathbf{w} to the point with position vector \mathbf{v} .

Note that there is also a special vector $\mathbf{0} = (0, 0)$ which may be viewed either as a special point *the origin*, where the axes cross, or as the *null translation*, the translation that doesn't move anything.

The vectors $(1, 0)$ and $(0, 1)$ form the *standard* or *canonical basis* for \mathbb{R}^2 . They are denoted by the symbols \mathbf{i} and \mathbf{j} respectively. Note that any vector $\mathbf{v} = (x, y)$ can be written uniquely as a linear combination of \mathbf{i} and \mathbf{j} : that is

$$(x, y) = x\mathbf{i} + y\mathbf{j}$$

and this is the only way to write (x, y) as a sum of scalar multiples of \mathbf{i} and \mathbf{j} . Likewise the vectors $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ form the canonical basis for \mathbb{R}^3 and are respectively denoted as \mathbf{i} , \mathbf{j} and \mathbf{k} .

Proposition 3 *Vector addition and scalar multiplication satisfy the following properties. These properties verify that \mathbb{R}^2 is a real vector space (cf. Michaelmas Linear Algebra). Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ and $\lambda, \mu \in \mathbb{R}$. Then*

$$\begin{array}{lll} \mathbf{u} + \mathbf{0} = \mathbf{u} & \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} & \mathbf{0}\mathbf{u} = \mathbf{0} \\ \mathbf{u} + (-\mathbf{u}) = \mathbf{0} & (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) & \mathbf{1}\mathbf{u} = \mathbf{u} \\ (\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u} & \lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v} & \lambda(\mu\mathbf{u}) = (\lambda\mu)\mathbf{u} \end{array}$$

Everything above generalises in an obvious way to the case of \mathbb{R}^n and from now on we will discuss this general case.

1.2 Geometry of Vectors

As vectors represent geometric ideas like points and translations, they have important geometric properties as well as algebraic ones. The length of a vector $\mathbf{v} = (v_1, v_2, \dots, v_n)$, which is written $|\mathbf{v}|$, is defined to be

$$|\mathbf{v}| = \sqrt{\sum_{i=1}^n (v_i)^2}.$$

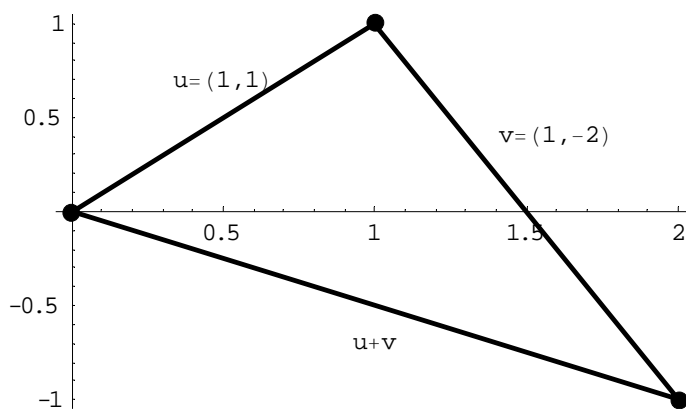
This is exactly what you'd expect it to be: from Pythagoras' Theorem we see this is the distance of the point \mathbf{v} from the origin, or equivalently how far a point, translated by \mathbf{v} , is from where it started. So if \mathbf{p} and \mathbf{q} are two points in the plane, then the vector that will translate \mathbf{p} to \mathbf{q} is $\mathbf{q} - \mathbf{p}$, and the distance between them is $|\mathbf{q} - \mathbf{p}|$ (or equally $|\mathbf{p} - \mathbf{q}|$).

Note that $|\mathbf{v}| \geq 0$ and that $|\mathbf{v}| = 0$ if and only if $\mathbf{v} = \mathbf{0}$. Also $|\lambda\mathbf{v}| = |\lambda| |\mathbf{v}|$ for any $\lambda \in \mathbb{R}$.

Proposition 4 (*Triangle Inequality*) *Let $\mathbf{u} = (u_1, u_2, \dots, u_n)$, $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$. Then*

$$|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$$

with equality when \mathbf{u} and \mathbf{v} are multiples of one another.



Geometric interpretation of the triangle inequality

Proof. Note that for $t \in \mathbb{R}$

$$0 \leq |\mathbf{u} + t\mathbf{v}|^2 = \sum_{i=1}^n (u_i + tv_i)^2 = |\mathbf{u}|^2 + 2t \sum_{i=1}^n u_i v_i + t^2 |\mathbf{v}|^2.$$

The RHS of the above inequality is a quadratic in t which is always non-negative and so has non-positive discriminant (i.e. $b^2 \leq 4ac$). Hence

$$4 \left(\sum_{i=1}^n u_i v_i \right)^2 \leq 4 |\mathbf{u}|^2 |\mathbf{v}|^2$$

and so

$$\left| \sum_{i=1}^n u_i v_i \right| \leq |\mathbf{u}| |\mathbf{v}|. \quad (1)$$

Hence

$$|\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + 2 \sum_{i=1}^n u_i v_i + |\mathbf{v}|^2 \leq |\mathbf{u}|^2 + 2 |\mathbf{u}| |\mathbf{v}| + |\mathbf{v}|^2 = (|\mathbf{u}| + |\mathbf{v}|)^2$$

to give the required result. The inequality 1 is known as the *Cauchy-Schwarz Inequality*. Note that we have equality in $b^2 \leq 4ac$ if and only if the quadratic $|\mathbf{u} + t\mathbf{v}|^2 = 0$ has a unique real solution, say $t = t_0$. So $\mathbf{u} + t_0\mathbf{v} = \mathbf{0}$ and we see that \mathbf{u} and \mathbf{v} are multiples of one another. Hence equality only occurs in the triangle inequality when \mathbf{u} and \mathbf{v} are multiples of one another. ■

1.3 The Scalar Product

Given two vectors $\mathbf{u} = (u_1, u_2, \dots, u_n), \mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$, the *scalar product* $\mathbf{u} \cdot \mathbf{v}$, also known as the *dot product* or *Euclidean inner product*, is given by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n.$$

The following properties of the scalar product are easy to verify and are left as exercises. (Note (v) has already been proven in the previous section).

Proposition 5 Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then

- (i) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$;
- (ii) $(\lambda \mathbf{u}) \cdot \mathbf{v} = \lambda (\mathbf{u} \cdot \mathbf{v})$;
- (iii) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$;
- (iv) $\mathbf{u} \cdot \mathbf{u} = |\mathbf{u}|^2 \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$;
- (v) (*Cauchy-Schwarz Inequality*) $|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}| |\mathbf{v}|$ with equality when \mathbf{u} and \mathbf{v} are multiples of one another.

We see that the length of a vector \mathbf{u} can be written in terms of the scalar product, namely

$$|\mathbf{u}| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$$

We can also define *angle* using the scalar product in terms of their scalar product.

Definition 6 Given two non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ the angle between them is given by the expression

$$\cos^{-1} \left(\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}| |\mathbf{v}|} \right).$$

Note that the formula makes sense as

$$-1 \leq \mathbf{u} \cdot \mathbf{v} / (|\mathbf{u}| |\mathbf{v}|) \leq 1$$

from the Cauchy-Schwarz Inequality. If we take the principal values of \cos^{-1} to be in the range $0 \leq \theta \leq \pi$ then this formula measures the smaller angle between the vectors. Note that two vectors \mathbf{u} and \mathbf{v} are perpendicular precisely when $\mathbf{u} \cdot \mathbf{v} = 0$.

2 Matrices

At its simplest a *matrix* is just a two-dimensional array of numbers: for example

$$\begin{pmatrix} 1 & 2 & -3 \\ \sqrt{2} & \pi & 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1.2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

are all matrices. The examples above are respectively a 2×3 matrix, a 3×1 matrix and a 2×2 matrix (read ‘2 by 3’ etc.); the first figure refers to the number of *rows* and the second to the number of *columns*. So vectors like (x, y) and (x, y, z) are also matrices, respectively 1×2 and 1×3 matrices.

The 3×1 matrix above could just as easily be thought of as a vector — it is after all just a list of three numbers, but written down rather than across. This is an example of a *column* vector. When we need to differentiate between the two, the vectors we have already met like (x, y) and (x, y, z) are called *row* vectors.

Notation 7 *The set of $1 \times n$ row vectors (or n -tuples) is written as \mathbb{R}^n . When we need to differentiate between row and column vectors we write $(\mathbb{R}^n)'$ or $(\mathbb{R}^n)^*$ for the set of $n \times 1$ column vectors. If there is no chance of confusion between the two (because we are only using row vectors, or only column vectors) then \mathbb{R}^n can denote either set.*

Notation 8 *If you are presented with an $m \times n$ matrix $A = (a_{ij})$ then the notation here simply means that a_{ij} is the (i, j) th entry. That is, the entry in the i th row and j th column is a_{ij} . Note that i can vary between 1 and m , and that j can vary between 1 and n . So*

$$i\text{th row} = (a_{i1}, \dots, a_{in}) \quad \text{and} \quad j\text{th column} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Notation 9 *As with vectors there is a zero $m \times n$ matrix, whose every entry is 0, which we denote by 0 unless we need to specify its size.*

2.1 Algebra of Matrices

We shall see though that a matrix is much more than simply an array of numbers. Where vectors can be thought of as points in space, we shall see that a matrix is really a map between spaces: specifically a $m \times n$ matrix can be thought of as a map from $(\mathbb{R}^n)'$ to $(\mathbb{R}^m)'$. To see this we need to first talk about how matrices add and multiply.

1. **Addition:** Let A be an $m \times n$ matrix (recall: m rows and n columns) and B be an $p \times q$ matrix. We would like to add them (as we added vectors) by adding their corresponding entries. So to add them the two matrices have to be the same size, that is $m = p$ and $n = q$. In which case we have

$$(A + B)_{ij} = a_{ij} + b_{ij} \quad \text{for } 1 \leq i \leq m \text{ and } 1 \leq j \leq n.$$

2. **Scalar Multiplication:** Let A be an $m \times n$ matrix and k be a constant (a scalar). Then the matrix kA has as its (i, j) th entry

$$(kA)_{ij} = ka_{ij} \quad \text{for } 1 \leq i \leq m \text{ and } 1 \leq j \leq n.$$

That is, we multiply each of the entries of A by k to get the new matrix kA .

3. **Matrix Multiplication:** Based on how we added matrices then you might think that we multiply matrices in a similar fashion, namely multiplying corresponding entries, but we do not. At first glance the rule for multiplying matrices is going to seem rather odd, but we will soon discover why we multiply them as we do.

The rule is this: we can multiply an $m \times n$ matrix A with an $p \times q$ matrix B if $n = p$ and we produce an $m \times q$ matrix AB with (i, j) th entry

$$(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj} \text{ for } 1 \leq i \leq m \text{ and } 1 \leq j \leq q.$$

It may help a little to write the rows of A as $\mathbf{r}_1, \dots, \mathbf{r}_m$ and the columns of B as $\mathbf{c}_1, \dots, \mathbf{c}_q$ and the above rule says that

$$(AB)_{ij} = \mathbf{r}_i \cdot \mathbf{c}_j \text{ for } 1 \leq i \leq m \text{ and } 1 \leq j \leq q.$$

We dot the rows of A with the columns of B .

This will, I am sure, seem pretty bizarre, let alone easy to remember — so here are some examples.

Example 10 Show that $2(A + B) = 2A + 2B$ for the following matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -2 \\ 5 & 1 \end{pmatrix}.$$

Here we are checking the distributive law in a specific example. Generally it is the case that $c(A + B) = cA + cB$. But to check it here we note that

$$\begin{aligned} A + B &= \begin{pmatrix} 1 & 0 \\ 8 & 5 \end{pmatrix}, \text{ and so } 2(A + B) = \begin{pmatrix} 2 & 0 \\ 16 & 10 \end{pmatrix}; \\ 2A &= \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}, \text{ and } 2B = \begin{pmatrix} 0 & -4 \\ 10 & 2 \end{pmatrix}, \text{ so that } 2A + 2B = \begin{pmatrix} 2 & 0 \\ 16 & 10 \end{pmatrix}. \end{aligned}$$

Example 11 Where possible, calculate pairwise the products of the following matrices.

$$\underbrace{A = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix}}_{2 \times 2}, \quad \underbrace{B = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}}_{2 \times 3}, \quad \underbrace{C = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}}_{2 \times 2}.$$

First up, the products we can form are: AA, AB, AC, CA, CB, CC . Let's slowly go through the product AC .

$$\begin{pmatrix} \boxed{1} & \boxed{2} \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \boxed{1} & -1 \\ \boxed{1} & -1 \end{pmatrix} = \begin{pmatrix} 1 \times 1 + 2 \times 1 & ?? \\ ?? & ?? \end{pmatrix} = \begin{pmatrix} 3 & ?? \\ ?? & ?? \end{pmatrix}.$$

This is how we calculate the $(1, 1)$ th entry of AC . We the first row of A and the first column of C and we dot them together. We complete the remainder of the product as follows:

$$\begin{aligned} \begin{pmatrix} \boxed{1} & \boxed{2} \\ -1 & 0 \end{pmatrix} \begin{pmatrix} \boxed{1} & \boxed{-1} \\ \boxed{1} & \boxed{-1} \end{pmatrix} &= \begin{pmatrix} 2 & \boxed{1 \times (-1) + 2 \times (-1)} \\ ?? & ?? \end{pmatrix} = \begin{pmatrix} 3 & \boxed{-3} \\ ?? & ?? \end{pmatrix}; \\ \begin{pmatrix} \boxed{1} & \boxed{2} \\ -1 & \boxed{0} \end{pmatrix} \begin{pmatrix} \boxed{1} & -1 \\ \boxed{1} & -1 \end{pmatrix} &= \begin{pmatrix} 2 & 4 \\ \boxed{-1 \times 1 + 0 \times 1} & ?? \end{pmatrix} = \begin{pmatrix} 3 & -3 \\ -1 & ?? \end{pmatrix}; \\ \begin{pmatrix} \boxed{1} & \boxed{2} \\ -1 & \boxed{0} \end{pmatrix} \begin{pmatrix} \boxed{1} & \boxed{-1} \\ \boxed{1} & \boxed{-1} \end{pmatrix} &= \begin{pmatrix} 2 & 4 \\ 0 & \boxed{(-1) \times (-1) + 0 \times (-1)} \end{pmatrix} = \begin{pmatrix} 3 & -3 \\ -1 & \boxed{1} \end{pmatrix}. \end{aligned}$$

So finally

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

We complete the remaining examples more quickly but still leaving a middle stage in the calculation to help see the process:

$$\begin{aligned}
 AA &= \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1-2 & 2+0 \\ -1+0 & -2+0 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ -1 & -2 \end{pmatrix}; \\
 AB &= \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1+6 & 2+4 & 3+2 \\ -1+0 & -2+0 & -3+0 \end{pmatrix} = \begin{pmatrix} 7 & 6 & 5 \\ -1 & -2 & -3 \end{pmatrix}; \\
 CA &= \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1+1 & 2-0 \\ 1+1 & 2-0 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}; \\
 CB &= \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1-3 & 2-2 & 3-1 \\ 1-3 & 2-2 & 3-1 \end{pmatrix} = \begin{pmatrix} -2 & 0 & 2 \\ -2 & 0 & 2 \end{pmatrix}; \\
 CC &= \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1-1 & 1-1 \\ 1-1 & 1-1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.
 \end{aligned}$$

N.B. There are certain important things to note here, which make matrix algebra different from the algebra of numbers.

1. Note the $AC \neq CA$. That is, matrix multiplication is not generally *commutative*.
2. It is, though, *associative*, which means that

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

whenever this product makes sense. So we can write down a product like $A_1A_2 \dots A_n$ without having to specify how we go about multiplying all these matrices or needing to worry about bracketing. But we do have to keep the order in mind. We shall not prove this fact here.

3. Note that $CC = 0$ even though C is non-zero — not something that happens with numbers.
4. The distributive laws also hold for matrix multiplication, namely

$$A(B + C) = AB + AC \quad \text{and} \quad (A + B)C = AC + BC$$

whenever these products make sense.

Notation 12 We write A^2 for the product AA and similarly we write A^n for the product $\underbrace{AA \cdots A}_{n \text{ times}}$. Note that A must be a square matrix (same number of rows and columns) for this to make sense.

3 Matrices as Maps

We have seen then how we can multiply an $m \times n$ matrix A and an $n \times p$ matrix B to form a product AB which is an $m \times p$ matrix. Now given the co-ordinates of a point in n -dimensional space, we can put them in a column to form a $n \times 1$ column vector, an element $\mathbf{v} \in (\mathbb{R}^n)'$. If we *pre*multiply this vector \mathbf{v} by the $m \times n$ matrix A we get a $m \times 1$ column vector $A\mathbf{v} \in (\mathbb{R}^m)'$.

So premultiplying by the matrix A gives a map: $(\mathbb{R}^n)' \rightarrow (\mathbb{R}^m)' : \mathbf{v} \mapsto A\mathbf{v}$.

Why have we started using column vectors all of a sudden? This is really just a matter of choice: we could just as easily take row vectors and *post*multiply by matrices on the right, and this is the convention some books choose. But, as we are used to writing functions on the left, we will use column vectors and premultiply by matrices.

Importantly though, we can now answer the question of why we choose to multiply matrices as we do. Take an $m \times n$ matrix A and an $n \times p$ matrix B . We have two maps associated with premultiplication by A and B ; let's call them α , given by:

$$\alpha : (\mathbb{R}^n)' \rightarrow (\mathbb{R}^m)' : \mathbf{v} \mapsto A\mathbf{v}$$

and β , given by

$$\beta : (\mathbb{R}^p)' \rightarrow (\mathbb{R}^n)' : \mathbf{v} \mapsto B\mathbf{v}.$$

We also have their composition $\alpha \circ \beta$, that is we do β first and then α , given by

$$\alpha \circ \beta : (\mathbb{R}^p)' \rightarrow (\mathbb{R}^m)' : \mathbf{v} \mapsto A(B\mathbf{v}).$$

But we have already commented that matrix multiplication is associative, so that

$$A(B\mathbf{v}) = (AB)\mathbf{v}.$$

That is,

the composition $\alpha \circ \beta$ is premultiplication by the product AB .

So if we think of matrices more as maps, rather than just simple arrays of numbers, matrix multiplication is quite natural and simply represents the composition of the corresponding maps.

3.1 Linear Maps

Let A be an $m \times n$ matrix and consider the associated map

$$\alpha : (\mathbb{R}^n)' \rightarrow (\mathbb{R}^m)' : \mathbf{v} \mapsto A\mathbf{v}$$

which is just pre-multiplication by A . Because of the distributive laws that hold for matrices, given column vectors $\mathbf{v}_1, \mathbf{v}_2 \in (\mathbb{R}^n)'$ and scalars $c_1, c_2 \in \mathbb{R}$ then

$$\begin{aligned} \alpha(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) &= A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) \\ &= c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 \\ &= c_1\alpha(\mathbf{v}_1) + c_2\alpha(\mathbf{v}_2). \end{aligned}$$

This means α is what is called a *linear map* — multiplication by a matrix leads to a linear map. The important thing is that converse also holds true — any linear map $(\mathbb{R}^n)' \rightarrow (\mathbb{R}^m)'$ has an associated matrix.

To see this most clearly we will return to 2×2 matrices. Let $\alpha : (\mathbb{R}^2)' \rightarrow (\mathbb{R}^2)'$ be a linear map — we'll try to work out what its associated matrix is. Let's suppose we're right and α is just premultiplication by some 2×2 matrix; let's write it as

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

Note that if we multiply the canonical basis vectors

$$\mathbf{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

by this matrix we get

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix}.$$

So, *if* this matrix exists, the first column is $\alpha(\mathbf{i})$ and the second column is $\alpha(\mathbf{j})$. But if we remember that α is linear then we see now that we have the right matrix, let's call it

$$A = \left(\alpha(\mathbf{i}) \quad \alpha(\mathbf{j}) \right).$$

Then

$$\begin{aligned} A \begin{pmatrix} x \\ y \end{pmatrix} &= x\alpha(\mathbf{i}) + y\alpha(\mathbf{j}) \\ &= \alpha(x\mathbf{i} + y\mathbf{j}) \quad [\text{as } \alpha \text{ is linear}] \\ &= \alpha \begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

We shall make use of this later when we calculate the matrices of some standard maps.

The calculations we performed above work just as well generally: if $\alpha : (\mathbb{R}^n)' \rightarrow (\mathbb{R}^m)'$ is a linear map then it is the same as premultiplying by an $m \times n$ matrix. In the columns of this matrix are the images of the canonical basis of \mathbb{R}^n under α .

4 Simultaneous Linear Equations — 2×2 Inverses

Suppose we are given two linear equations in two variables x and y . These might have the form

$$2x + 3y = 1 \tag{2}$$

and

$$3x + 2y = 2 \tag{3}$$

To have solved these in the past, you might have argued along the lines:

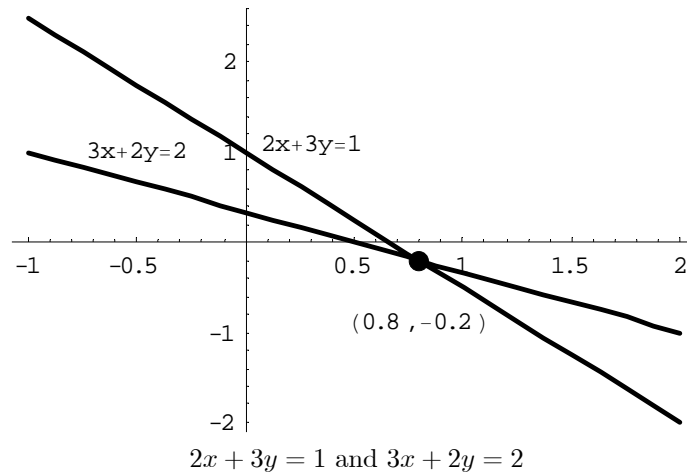
$$\begin{aligned} \text{eliminating } x \text{ by considering } 3 \times (2) - 2 \times (3) & : (6x + 9y) - (6x + 4y) = 3 - 4; \\ \text{simplifying:} & \quad 5y = -1; \\ \text{so} & : y = -0.2; \\ \text{substituting back in (2)} & : 2x - 0.6 = 1; \\ \text{solving} & : x = 0.8. \end{aligned}$$

So we see we have a unique solution: $x = 0.8, y = -0.2$.

You may even have seen how this situation could be solved graphically by drawing the lines

$$2x + 3y = 1 \text{ and } 3x + 2y = 2;$$

the solution then is their unique intersection.



We can though, put the two equations (2) and (3) into matrix form. We do this by writing

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

The two simultaneous equations in the scalar variables x and y have now been replaced by a single equation in vector quantities — we have in this vector equation two 2×1 vectors (one on each side of the equation), and for the vector equation to be true *both* co-ordinates of the two vectors must agree.

We know that it is possible to *undo* this equation to obtain the solution

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.8 \\ -0.2 \end{pmatrix}$$

because we have already solved this system of equations. In fact there is a very natural way of unravelling these equations using matrices.

Definition 13 Let A be a 2×2 matrix. We say that B is an inverse of A if

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

- The matrix I_2 is called the *identity* matrix — or more specifically it is the 2×2 identity matrix. There is an $n \times n$ identity matrix which has 1s down the diagonal from top left to bottom right and 0s elsewhere.
- The identity matrices have the property that

$$AI_n = A = I_nA$$

for any $n \times n$ matrix A .

- If an inverse B for A exists then it is unique. This is easy to show: suppose B and C were two inverses then note that

$$C = I_nC = (BA)C = B(AC) = BI_n = B.$$

- We write A^{-1} for the inverse of A if an inverse exists.
- If $BA = I_n$ then, in fact, it will follow that $AB = I_n$. The converse is also true. We will not prove this here.

Proposition 14 *The matrix*

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

has an inverse precisely when $ad - bc \neq 0$. If $ad - bc \neq 0$ then

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

Proof. Note for any values of a, b, c, d , that

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = (ad - bc)I_2.$$

So if $ad - bc \neq 0$ then we can divide by this scalar and we have found our inverse.

But if $ad - bc = 0$ then we have found a matrix

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

such that

$$BA = 0_2$$

the 2×2 zero matrix. Now if an inverse C for A existed, we'd have that

$$0_2 = 0_2C = (BA)C = B(AC) = BI_2 = B.$$

So each of a, b, c and d must be zero. So $A = 0_2$ which makes it impossible for AC to be I_2 — multiplying any C by the 0_2 will always lead to 0_2 , not the identity. ■

Let's return now to a pair of simultaneous equations. Say they have the form

$$\begin{aligned} ax + by &= k_1, \\ cx + dy &= k_2. \end{aligned}$$

Then these can be rewritten as a single vector equation:

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}.$$

If the matrix A has an inverse A^{-1} then we can premultiply both sides by this and we find

$$\begin{pmatrix} x \\ y \end{pmatrix} = I_2 \begin{pmatrix} x \\ y \end{pmatrix} = A^{-1}A \begin{pmatrix} x \\ y \end{pmatrix} = A^{-1} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$$

and we have found our unique solution.

What happens though if A doesn't have an inverse? In this case $ad = bc$, or equally that $a : c = b : d$ and we see that

$$ax + by \quad \text{and} \quad cx + dy$$

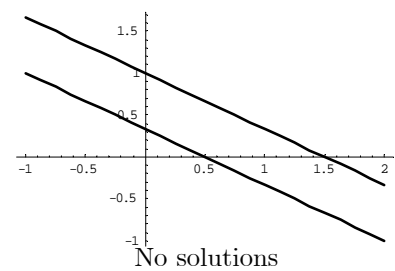
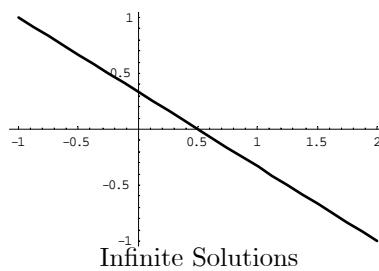
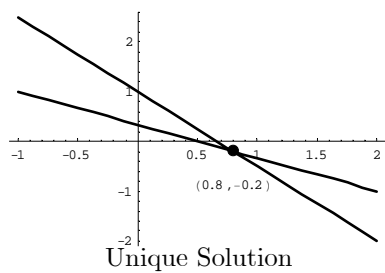
are multiples of one another.

- If the two equations are entirely multiples of one another, i.e. $a : c = b : d = k_1 : k_2$ then we essentially just have the one equation and there are infinitely many solutions.
- If the two equations aren't entirely multiples of one another, just the left hand sides i.e. $a : c = b : d \neq k_1 : k_2$ then we have two contradictory equations and there are no solutions.

Geometrically these three cases correspond to the two lines

$$ax + by = k_1 \quad \text{and} \quad cx + dy = k_2$$

being non-parallel and intersecting once, parallel and concurrent, or parallel and distinct.



5 Determinants

The determinant of a square $n \times n$ matrix is a number which reflects the way a matrix (or rather its associated map α) stretch space. We will define this only for 2×2 matrices.

Definition 15 Given a 2×2 matrix

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

Then

$$\det A = ad - bc.$$

We have already seen that A has an inverse precisely when $\det A \neq 0$. More generally $|\det A|$ is an area-scaling factor for α . So if R is a region of the plane, then

$$\text{area}(\alpha(R)) = |\det A| \times \text{area}(R).$$

The sign of $\det A$ is also important (whether it is positive or negative). If $\det A$ is positive then α will preserve a sense of orientation (such as a rotation does) but if $\det A$ is negative then the sense of orientation will be reversed (such as a reflection does). If $\det A = 0$ then α collapses space: under α then the xy -plane will map to a line (or a single point in the case of $A = 0_2$). Viewed from this geometric point the following multiplicative property of determinants should not seem to surprising, if we recall that AB represents the composition of the maps represented by B then A .

Proposition 16 Let A and B be two 2×2 matrices. Then

$$\det(AB) = \det A \det B.$$

Proof. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} e & f \\ g & h \end{pmatrix}.$$

Then

$$\begin{aligned} \det(AB) &= \det \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix} \\ &= (ae + bg)(cf + dh) - (af + bh)(ce + dg) \\ &= bgcf + aedh - bhce - afdg \quad [\text{after cancelling}] \\ &= (ad - bc)(eh - fg) \\ &= \det A \det B. \end{aligned}$$

■

6 Transformations of the Plane

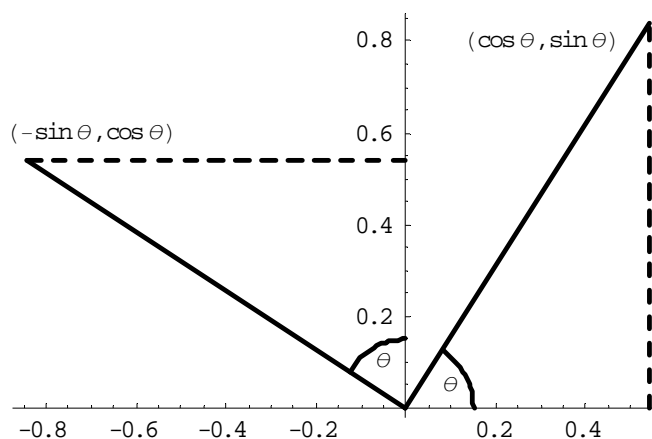
We end with a little geometry. We already noted in the section *Matrices as Maps* how a matrix A leads to a linear map α and how every linear map can be achieved by premultiplying by a matrix — we further saw how to calculate such matrices.

The following are all examples of linear maps of the xy -plane:

- Rotation R_θ anti-clockwise by θ about the origin;
- Reflection in a line through the origin;
- A stretch with an invariant line through the origin;
- A shear with an invariant line through the origin;

though this list is far from comprehensive.

We concentrate on the first example R_θ . Remember to find the matrix for R_θ we need to find the images $R_\theta(\mathbf{i})$ and $R_\theta(\mathbf{j})$. We note from the diagrams



that

$$R_\theta(\mathbf{i}) = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \quad \text{and} \quad R_\theta(\mathbf{j}) = \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}.$$

So, with a little abuse of notation, we can write

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

There are several important things to notice about this matrix:

- $\det R_\theta = +1$, unsurprisingly, as it preserves area and it preserves a sense of orientation.
- $R_\theta R_\phi = R_{\theta+\phi}$ — rotating by θ and then by ϕ is the same as rotating by $\theta + \phi$. This is one way of calculating the $\cos(\theta + \phi)$ and $\sin(\theta + \phi)$ formulas.
- the inverse of R_θ is $R_{-\theta}$ — rotating by $-\theta$ undoes the effect of rotating by θ .
- $R_\theta(\mathbf{v}) \cdot R_\theta(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$ for any two 2×1 vectors \mathbf{v} and \mathbf{w} . This perhaps is surprising: this equation says that R_θ is an *orthogonal* matrix (cf. Michaelmas Geometry I course). One consequence of this equation is that R_θ preserves distances and angles.