

Some linear algebra basics

Sean Fitzpatrick

June 13, 2018

A fundamental principle underlying calculus (of one or many variables) is that any differentiable function $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$ can be locally approximated by a *linear* function. With this principle in mind, it's important to understand the behaviour of general linear functions, which requires a certain amount of linear algebra, which we will now discuss (without going into too much detail).

1 Vector spaces

Definition 1.1 A **vector space** over the real numbers¹ is a set V equipped with operators of addition

$$V \times V \rightarrow V : (\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} + \mathbf{y}$$

and scalar multiplication

$$\mathbb{R} \times V \rightarrow V : (c, \mathbf{x}) \mapsto c\mathbf{x}$$

such that the following axioms are satisfied, for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all scalars $c, d \in \mathbb{R}$:

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (Commutativity of addition)
2. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (Associativity of addition)
3. There exists an element $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x}$. (Existence of a zero element.)
4. There exists an element $-\mathbf{x} \in V$ such that $\mathbf{x} + (-\mathbf{x}) = -\mathbf{x} + \mathbf{x} = \mathbf{0}$. (Existence of additive inverses.)
5. $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$ (Distributive law for vector addition.)
6. $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$ (Distributive law for scalar addition.)
7. $c(d\mathbf{x}) = (cd)\mathbf{x}$ (Associativity of scalar multiplication.)
8. $1\mathbf{x} = \mathbf{x}$ (Multiplication by the unit scalar.)

¹One can also study vector spaces over other *fields*, such as the complex numbers, but we will only consider real vector spaces in this course.

From these axioms various other properties follow automatically; for example, $0\mathbf{x} = \mathbf{0}$ and $(-1)\mathbf{x} = -\mathbf{x}$ for any $\mathbf{x} \in V$. For us, the main example of a vector space is $V = \mathbb{R}^n$, where

we view \mathbb{R}^n not as a set of points (x_1, \dots, x_n) , but as the set of column vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$,

equipped with the addition $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}$ and scalar multiplication

$c\mathbf{x} = c \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} cx_1 \\ \vdots \\ cx_n \end{bmatrix}$. Another example of a vector space is the set of all polynomials

$p(x) = a_n x^n + \dots + a_1 x + a_0$ of degree less than or equal to n , equipped with the usual addition and scalar multiplication of functions.

Remark 1.2 It is also possible to identify \mathbb{R}^n with the vector space of *row* vectors $\mathbf{x} = [x_1 \ \dots \ x_n]$, although it will be more convenient to think of this space as being “dual” to \mathbb{R}^n in a sense that you will explore on your second assignment.

2 Bases of vector spaces

Two important notions in linear algebra are those of *linear independence* and *span*. Given vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ in a vector space V , a **linear combination** of these vectors is any expression of the form

$$\mathbf{x} = c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2 + \dots + c_n \mathbf{x}_n,$$

where $c_1, c_2, \dots, c_n \in \mathbb{R}$ are scalars. Note that the vector space axioms ensure that \mathbf{x} is again an element of V , and that the expression defining \mathbf{x} is unambiguous.

Definition 2.1 Let $S = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ be a set of vectors in a vector space V . The **span** of S is the set $\text{span}(S) \subseteq V$ consisting of all possible linear combinations of the vectors in S . That is,

$$\text{span}(S) = \{\mathbf{x} \in V \mid \mathbf{x} = c_1 \mathbf{x}_1 + \dots + c_k \mathbf{x}_k \text{ for some } c_1, \dots, c_k \in \mathbb{R}\}.$$

For example, the span of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ is a plane through the origin. Notice that the span of any finite set $S \subseteq V$ is again a vector space (exercise). A subset of a vector space which is again a vector space is known as a *subspace*.

Remark 2.2 Note that since a vector space must contain a zero vector, a subset of a vector space V can only be a subspace if it contains the zero vector of V . In particular, lines and planes in \mathbb{R}^3 only define subspaces if they pass through the origin. Lines and planes that do not pass through the origin are sometimes known as *affine* spaces: spaces which are *geometrically* the same as vector spaces, but not *algebraically*, since in particular they do not have a preferred choice of origin. Given a fixed vector $\mathbf{x}_0 \in V$, a typical affine space would be of the form $U = \{\mathbf{x}_0 + \mathbf{x} \mid \mathbf{x} \in \text{span}(S)\}$.

Definition 2.3 A set $S = \{\mathbf{x}_1, \dots, \mathbf{x}_k\} \subseteq V$ is called **linearly independent** if none of the elements of S can be written as a linear combination of the others. Equivalently, S is linearly independent if the only solution to the equation $c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k = \mathbf{0}$ is $c_1 = \dots = c_k = 0$.

A linearly independent set can be thought of as one that does not contain any “unnecessary” vectors: if one of the vectors, say \mathbf{x}_k , in S was a linear combination of the others, then the span of $\{\mathbf{x}_1, \dots, \mathbf{x}_{k-1}\}$ would be the same as the span of S . In \mathbb{R}^3 , two vectors are linearly independent if they are not parallel, and three vectors are linearly independent if they do not all lie in the same plane.

Finally, we can define a **basis** for a vector space V to be a set $S = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ such that S is linearly independent, and $V = \text{span}(S)$. A basis S for V has the property that every element of V can be written as a *unique* linear combination of the elements of S . Note also that the number of vectors in S is “just right”: if we removed any element of S , the remaining vectors would no longer span V , and if we added another vector, the resulting set would no longer be linearly independent. One of the most important theorems in linear algebra is that every basis for a given vector space V must contain the same number of elements. This number is known as the **dimension** of V .

Remark 2.4 Note that a choice of basis $S = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ gives us an identification between elements of V and elements of \mathbb{R}^n : if $\mathbf{x} = c_1\mathbf{e}_1 + \dots + c_n\mathbf{e}_n$, we can identify \mathbf{x} with the column

vector $\mathbf{x}_S = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. Thus, every vector space of V is “the same” as \mathbb{R}^n , but only up to a choice of basis.

3 Matrices

A **matrix** of size $m \times n$ is an array of mn real numbers arranged into m rows and n columns. We denote the entries of a matrix A of size $m \times n$ by a_{ij} , where the index $i \in \{1, \dots, m\}$ indicates the row in which the number a_{ij} is found, and the index $j \in \{1, \dots, n\}$ indicates the column in which a_{ij} is found:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

For example, $A = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 0 & -5 \end{bmatrix}$ is a 2×3 matrix. The set of all $m \times n$ matrices is in fact a vector space of dimension mn ; addition and scalar multiplication are defined component-wise, as for column vectors in \mathbb{R}^n (which are just matrices of size $n \times 1$).

As well as addition and scalar multiplication, it’s possible to define the product of two matrices, but only for certain compatible pairs. The product AB of two matrices A and B is defined whenever the number of columns of A is equal to the number of rows of B ; that

is, if A is an $m \times n$ matrix, and B is an $n \times p$ matrix. The product AB is then an $m \times p$ matrix, defined as follows: if c_{ij} denotes the $(i, j)^{\text{th}}$ entry of AB (that is, the entry in the i^{th} row and j^{th} column of AB), then

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

You might find it helpful to think of c_{ij} as the dot product of the i^{th} row of A with the j^{th} column of B . (And since the dot product is only defined between vectors of the same size, we need A to have as many columns as B has rows.)

Example 3.1 Let $A = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 0 & -5 \end{bmatrix}$, and let $B = \begin{bmatrix} 1 & -3 \\ -2 & 2 \\ 3 & -1 \end{bmatrix}$. Since A is 2×3 and B is 3×2 , the product AB is defined, and is a 2×2 matrix, with entries

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} = 1(1) + (-2)(-2) + 3(3) = 14$$

$$c_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} = 1(-3) + (-2)(2) + 3(-1) = -10$$

$$c_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} = 4(1) + 0(-2) + (-5)(3) = -11$$

$$c_{22} = a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} = 4(-3) + 0(2) + (-5)(-1) = -7,$$

which gives $AB = \begin{bmatrix} 14 & -10 \\ -11 & -7 \end{bmatrix}$.

In particular, we should make note of the following special case: if A is an $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$ is an $n \times 1$ matrix, then $\mathbf{y} = A\mathbf{x}$ is an $m \times 1$ matrix; that is, $\mathbf{y} \in \mathbb{R}^m$:

$$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

Matrix multiplication satisfies most of the usual properties we expect for multiplication. In particular, it is distributive and associative: $A(BC) = (AB)C$, $A(B + C) = AB + AC$, and $(A + B)C = AC + BC$, whenever these products are defined, and matrix multiplication also respects scalar multiplication: for any real number c , we have $(cA)B = A(cB) = c(AB)$. Note however that matrix multiplication is **not** commutative: in general $AB \neq BA$ (and in fact, if A is $m \times n$ and B is $n \times p$, BA is not even *defined*, unless $p = m$).

Among all *square* matrices (those of size $n \times n$), there is the matrix I_n whose entries are given by $I_{ii} = 1$ for $i = 1, \dots, n$, and $I_{ij} = 0$ if $i \neq j$. (So I_n has ones along its “main diagonal” and zeros everywhere else.) The matrix I_n is called the **identity matrix** since it has the property that $I_n A = A$ for any $n \times p$ matrix A , and $B I_n = B$ for any $m \times n$ matrix B .

In the case of a square matrix A of size $n \times n$, we can ask whether or not A has a multiplicative inverse A^{-1} such that $AA^{-1} = A^{-1}A = I_n$. Unlike the real numbers, it is not the case that every nonzero matrix A has an inverse. One consequence of this is that if we

know that $AB = AC$, we can **not** conclude that $B = C$, unless we know in advance that A^{-1} exists.

Finally, we should mention one other operations on matrices, called the **transpose**. Given an $m \times n$ matrix A , we can define an $n \times m$ matrix A^T , the transpose of A , such that the rows of A^T are the columns of A and vice versa. (In other words, the $(i, j)^{\text{th}}$ entry of A^T is equal to the $(j, i)^{\text{th}}$ entry of A .)

4 Determinants

Given an $n \times n$ (square) matrix A , we can compute a number $\det A$ (sometimes denoted $|A|$) called the **determinant** of A . Determinants are computed recursively via “cofactor expansion”. A 1×1 matrix A is simply a number a , and in this case we set $\det A = a$. For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ we define $\det A = ad - bc$. Note that this is the difference of the products of the entries on the two *diagonals* of A . For an $n \times n$ matrix A , the determinant of A is defined by²

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} A_{ij},$$

where A_{ij} is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the row and column of A containing a_{ij} . (Each $(n-1) \times (n-1)$ determinant can then be computed in terms of $(n-2) \times (n-2)$ determinants, and so on, until we reach 2×2 determinants.)

We will not need to know how to compute general determinants in this course, so we will not discuss methods for doing so here. It will, however, be useful to be aware of some of the properties of determinants, which we list here:

- A matrix A is invertible if and only if $\det A \neq 0$.
- For any two $n \times n$ matrices A and B , $\det(AB) = \det A \det B$.
- If we swap any two rows or columns in A , then $\det A$ is multiplied by -1 .
- If we multiply any row or column of A by a scalar c , then $\det A$ is multiplied by c as well.

5 Linear transformations

A **linear transformation** T is a map $T : V \rightarrow W$ between vector spaces V and W that respects the addition and scalar multiplication in both spaces, in the sense that for any vectors $\mathbf{x}, \mathbf{y} \in V$ and any scalars $a, b \in \mathbb{R}$, we have

$$T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y}).$$

²This formula gives the determinant in terms of expansion along the i^{th} row. It’s possible to show that the choice of row does not matter; moreover, one can also expand instead along any column of A .

For example, any $m \times n$ matrix A determines a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ given by $T(\mathbf{x}) = A\mathbf{x}$. In fact, this example is the canonical one. As soon as we make a choice of basis B for V and C for W , we can determine a matrix A such that if $\mathbf{y} = T(\mathbf{x})$, then $\mathbf{y}_C = A\mathbf{x}_B$, where \mathbf{x}_B denotes the element of \mathbb{R}^n determined by the basis B , and \mathbf{y}_C denotes the element of \mathbb{R}^m determined by the basis C (see Remark 2.4 above).

A linear transformation that is both one-to-one and onto is called an *isomorphism*. There exists an isomorphism between two vector spaces V and W if and only if V and W have the same dimension. For example, you should check that the association $\mathbf{x} \mapsto \mathbf{x}_S$ given by choosing a basis S for a vector space V is an isomorphism from V to \mathbb{R}^n (note that in particular this requires verifying that this map is linear).

Notice that any linear transformation $T : V \rightarrow W$ must take the zero vector of V to the zero vector of W . In the case of a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$, if we view \mathbb{R}^n as a set of points, then any linear transformation must “preserve the origin”. In particular, the definition of a linear transformation does not quite coincide with the usual notion of a linear *function* in calculus: usually we think of a linear function $f : \mathbb{R} \rightarrow \mathbb{R}$ as being a function of the form $f(x) = ax + b$, whereas f can only be a linear transformation if $b = 0$. When it becomes necessary to make the distinction, we will always reserve the term *linear transformation* to refer to the definition from linear algebra given above, and allow the more generic term *linear function* to refer to functions defined by the formula $f(\mathbf{x}) = A\mathbf{x} + \mathbf{y}$, where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, and A is an $m \times n$ matrix. (These functions are sometimes referred to as *affine* transformations - see Remark 2.2 above.)