

0.1 Vector Space

A vector space is an algebraic structure that consists of a set of vectors and two operations defined on these vectors: vector addition and scalar multiplication. The elements of the vector space are called vectors, which can be added together and multiplied by numbers, called scalars. The operations of vector addition and scalar multiplication must satisfy a set of axioms, which ensure that the resulting vectors remain within the vector space. It will be helpful to describe a vector space by first explaining the concept of a set and a group.

A set is a collection of mathematical objects such as numbers, lines or potentially other sets. They were first formalised by George Cantor in 1895 [1] as being either infinite or finite and containing distinct elements.

Definition 0.1.1 Let \mathcal{V} be a set and \otimes denote a binary operation between elements of this set such that $\otimes : \mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$. $G := (\mathcal{G}, \otimes)$ is called a group [2, page 36] if the following conditions are met and specific elements are present:

1. (Closure) The binary operation between any two elements of the set will result in an element which is also part of the set: $\otimes : \forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$.
2. (Associativity) The way in which elements of the set are combined within a larger expression does not affect the result: $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$.
3. (Neutral Element) A neutral element in a set is an element that, when

combined with any other element in the set using the group operation, results in the same element: $\exists e \in \mathcal{G} \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$.

4. (Inverse Element) The inverse of an element in a set is an element that when combined with the original element using the group operation, results in the neutral element. It allows for the reversal of the group operation: $\forall x \in \mathcal{G} \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$, where e is the neutral element. We can denote the inverse of an element x as x^{-1} .

Groups are very important in many areas of mathematics. Their rigorous and formal definition means that, if something is found to be a group, its properties can be better understood in the specific context in which it is relevant. An example of a group is the general linear group. This is the set of regular invertible matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$ with respect to matrix multiplication. Given a group G , if the order in which the group operation is performed does not matter i.e. $\forall x, y \in \mathcal{G} : x \otimes y = y \otimes x$, then $G = (\mathcal{G}, \otimes)$ is known as an *Abelian group* (commutative). An example of this would be $(\mathbb{Z}, +)$, the set of all integers under the addition operation.

A vector space [2, page 37] is a special type of group with some additional conditions.

Definition 0.1.2 (Vector Space) A real-valued *vector space* $V = (\mathcal{V}, +, \cdot)$ is a set \mathcal{V} with two operations:

$$+ : \mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V} \quad (\text{Inner Operation}) \tag{1}$$

$$\cdot : \mathbb{R} \times \mathcal{V} \rightarrow \mathcal{V} \quad (\text{Outer Operation}) \tag{2}$$

where:

1. $(\mathcal{V}, +)$ is an Abelian group
2. (Distributivity) The outer operation can be “Distributed” across elements either before or after the inner operation has occurred:
 - (a) $\forall \lambda \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V} : \lambda(\mathbf{x} + \mathbf{y}) = (\lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y})$
 - (b) $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : (\lambda + \psi) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \psi \cdot \mathbf{x}$
3. Associativity (outer operation): $\forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \lambda \cdot (\psi \cdot \mathbf{x}) = (\lambda\psi) \cdot \mathbf{x}$
4. Neutral element with respect to the outer operation: $\forall \mathbf{x} \in \mathcal{V} : 1 \cdot \mathbf{x} = \mathbf{x}$

We commonly think of vectors as being mathematical objects with both direction and magnitude. However, these are only geometric vectors. More generally, any set of objects which follows the definition of a vector space is known as a vector. For example, polynomials are also vectors. Two can be added together, resulting in another polynomial and they can be multiplied by a scalar $\lambda \in \mathbb{R}$ which again results in another polynomial.

0.2 Linear Independence

Linear independence [2, page 40] is a property of a set of vectors which describes whether there is any redundancy with respect to the linear combinations of these vectors.

Definition 0.2.1 (Linear Combination) Consider a vector space V and a finite number of vectors $\mathbf{x}_1, \dots, \mathbf{x}_k \in V$. Then, every $\mathbf{v} \in V$ of the form

$$\mathbf{v} = \lambda_1 \mathbf{x}_1 + \dots + \lambda_k \mathbf{x}_k = \sum_{i=1}^k \lambda_i \mathbf{x}_i \in V \quad (3)$$

with $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ is a *linear combination* of the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$.

Consider two vectors in \mathbb{R}^2 , $\mathbf{e}_1 = (1, 0)^\top$ and $\mathbf{e}_2 = (0, 1)^\top$. It is common to see these written as \mathbf{i} and \mathbf{j} respectively. We can represent any vector in \mathbb{R}^2 as a linear combination of these two vectors.

Definition 0.2.2 (Linear (In)dependence) Let us consider a vector space V with $k \in \mathbb{R}$ and $\mathbf{x}_1, \dots, \mathbf{x}_k \in V$. If there is a non-trivial combination, such that $\mathbf{0} = \sum_{i=1}^k \lambda_i \mathbf{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ are *linearly dependent*. If only the trivial solution exists, i.e., $\lambda_1 = \dots = \lambda_k = 0$ the vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ are *linearly independent*.

Lets consider the following set \mathcal{V} of vectors $\mathbf{v}_1, \mathbf{v}_2$ and \mathbf{v}_3 where (respectively):

$$\mathcal{V} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ -4 \end{pmatrix}, \begin{pmatrix} -4 \\ -1 \end{pmatrix} \right\} \quad (4)$$

And let $\mathbf{y} \in \mathbb{R}^2$ be any linear combination that can be made from these vectors. Writing this out explicitly with $\alpha, \beta, \omega \in \mathbb{R}$:

$$\mathbf{y} = \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} -2 \\ -4 \end{pmatrix} + \omega \begin{pmatrix} -4 \\ -1 \end{pmatrix} \quad (5)$$

$$= \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 2\beta \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \omega \begin{pmatrix} -4 \\ -1 \end{pmatrix} \quad (6)$$

$$= (\alpha - 2\beta) \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \omega \begin{pmatrix} -4 \\ -1 \end{pmatrix} \quad (7)$$

$$(8)$$

What we have shown is that if a vector \mathbf{y} can be represented as a linear combination of $\mathbf{v}_1, \mathbf{v}_2$ and \mathbf{v}_3 then we can represent \mathbf{y} as a linear combination of \mathbf{v}_2 and \mathbf{v}_3 . This is because \mathbf{v}_1 is a scaled version of \mathbf{v}_2 and vice versa. More formally:

$$\begin{pmatrix} -2 \\ -4 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \mathbf{0} \quad (9)$$

which following from definition 0.2.2 demonstrates that we have a linearly dependent set of vectors.

0.3 Generating Set and Span

A generating set and the span [2, page 44] of a set of vectors are concerned with the vector space produced by the linear combination of all the vectors in the set.

Definition 0.3.1 (Generating Set and Span) Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and set of vectors $\mathcal{A} = \{\mathbf{x}_1, \dots, \mathbf{x}_k\} \subseteq V$. If every vector $\mathbf{v} \in V$ can be expressed as a linear combination of vectors of $\mathbf{x}_1, \dots, \mathbf{x}_k$, \mathcal{A} is called the generating set of V . The set of all linear combinations of vectors in \mathcal{A} is called the span of \mathcal{A} . If \mathcal{A} spans the vector space V , we write $V = \text{span}[\mathcal{A}]$.

Consider the vector space $V = (\mathbb{R}^2, +, \cdot)$ under matrix addition and scalar multiplication. Previously we have said that every vector $v \in V$ can be represented as a linear combination of $\mathbf{e}_1 = (1, 0)^\top$ and $\mathbf{e}_2 = (0, 1)^\top$. By definition 0.3.1, the set of vectors $\mathcal{B} = \{\mathbf{e}_1, \mathbf{e}_2\}$ is a generating set of V . On the other hand, the set of two-dimensional vectors \mathbb{R}^2 contains all linear combinations of the set B meaning that the $\text{span}[\mathcal{B}] = \mathbb{R}^2$. In this sense, a generating set and the span of a set are two sides of the same coin.

0.4 Basis

A basis is concerned with the minimum number of vectors that would be needed to span a particular vector space.

Definition 0.4.1 (Basis) Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and $\mathcal{A} \subseteq \mathcal{V}$. A generating set \mathcal{A} of V is called *minimal* if there exists no smaller set

$\tilde{\mathcal{A}} \subseteq \mathcal{A} \subseteq \mathcal{V}$ that spans V . Every linearly independent generating set of V is minimal and is called a *basis* of V .

Consider the vector space $V = (\mathbb{R}^3, +, \cdot)$. A potential generating set for V is the set \mathcal{A} where:

$$\mathcal{A} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \quad (10)$$

This is because every vector in \mathbb{R}^3 can be represented as linear combinations of the vectors in \mathcal{A} . Since this set of vectors is linearly independent, it forms a *minimal* generating set. We've described previously that the *span* of a set of vectors can be thought of as the collection of all potential linear combinations of the set. \mathcal{A} forms a *minimal* generating set because no vectors may be removed without eliminating some linear combinations / reducing the overall span of the set. This is what we mean by a basis for the vector space V . As well as this, the number of dimensions of the vector space spanned by the basis is equal to the number of vectors in the basis for the vector space.

0.5 Row Rank and Column Rank

Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, the row rank (denoted as $rk_{row}(\mathbf{A})$) and the column rank (denoted as $rk_{col}(\mathbf{A})$) describe the number of linearly independent row vectors and column vectors respectively.

Proposition 0.5.1 For a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$:

$$rk_{col}(\mathbf{A}) = \dim(\text{span}[\mathbf{A}_{.1}, \dots, \mathbf{A}_{.m}]) \quad (11)$$

Example: Consider a matrix $A = [\mathbf{A}_{.1} | \mathbf{A}_{.2}]$ where $\mathbf{A}_{.1}$ and $\mathbf{A}_{.2}$ are linearly dependent. The *span* of $\mathcal{B} = \{\mathbf{A}_{.1}, \mathbf{A}_{.2}\}$ is all the linear combinations that can be created from the two vectors. Let $s \in \text{span}[B]$:

$$s = \lambda_1 \mathbf{A}_{.1} + \lambda_2 \mathbf{A}_{.2}, \quad \forall \lambda_1, \lambda_2 \in \mathbb{R} \quad (12)$$

Because the column vectors are linearly dependent, by definition 0.2.2, $\exists \lambda_3, \lambda_4$ such that:

$$\lambda_3 \mathbf{A}_{.1} + \lambda_4 \mathbf{A}_{.2} = \mathbf{0} \quad (13)$$

$$\mathbf{A}_{.1} = \frac{\lambda_4}{\lambda_3} \mathbf{A}_{.2} \quad (14)$$

Meaning that every linear combination can be written as:

$$s = \lambda_1 \mathbf{A}_{.1} + \lambda_2 \mathbf{A}_{.2} \quad (15)$$

$$= (-\lambda_1 \frac{\lambda_4}{\lambda_3} + \lambda_2) \mathbf{A}_{.2} \quad (16)$$

This means that $\mathcal{C} = \{\mathbf{A}_{.2}\}$ forms a basis. Because the number of elements in this set is 1, the dimension of $\text{span}[C]$ is 1. Because both column vectors of \mathbf{A} are linearly independent, the $rk_{col}(\mathbf{A}) = 1$ and so proposition 0.5.1 is shown to be true for the case of a matrix with two column vectors. Similar

methods can show that this is true for larger matrices. Furthermore, the same reasoning can be used to conclude that $rk_{row}(\mathbf{A}) = \dim(\text{span}[\mathbf{A}_1, \dots, \mathbf{A}_n])$.

Proposition 0.5.2 Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$, the following statement is always true:

$$rk_{row}(\mathbf{A}) = rk_{col}(\mathbf{A}) \quad (17)$$

where $rk_{row}(\mathbf{A})$ and $rk_{col}(\mathbf{A})$ denote the row rank and the column rank respectively.

Proof: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a matrix and let $u = rk_{col}(\mathbf{A})$. There exists a basis $C = \{\mathbf{c}_1, \dots, \mathbf{c}_u\}$ of $m \times 1$ column vectors that spans the same space spanned by the columns of \mathbf{A} . Let $\mathbf{B} = [\mathbf{c}_1 | \mathbf{c}_2 | \dots | \mathbf{c}_{u-1} | \mathbf{c}_u]$ be a matrix where $[\mathbf{v}_1 | \mathbf{v}_2]$ denotes the concatenation between two vectors \mathbf{v}_1 and \mathbf{v}_2 . By definition 0.2.2, each column of \mathbf{A} can be expressed as a linear combination of vectors in \mathbf{B} . If we collect the coefficients of these linear combinations into a matrix \mathbf{D} , we can represent \mathbf{A} as $\mathbf{A} = \mathbf{B}\mathbf{D}$ where $\mathbf{D} \in \mathbb{R}^{m \times n}$. Equally, we can see that the rows of $\mathbf{B}\mathbf{D}$ can be expressed as a linear combination of \mathbf{D} with coefficients taken from \mathbf{B} . \therefore , the span of the rows of \mathbf{A} is no greater than the span of the rows of \mathbf{D} because linear combinations of the rows of \mathbf{A} can be written as linear combinations of the rows in \mathbf{D} . \mathbf{D} has u rows. If they are linearly independent, the dimension of their span is u . Otherwise, it has dimension $< u$. \therefore , the row rank is $\leq u$ and so $rk_{row}(\mathbf{A}) \leq rk_{col}(\mathbf{A})$ for any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. If we let $\mathbf{K} \in \mathbb{R}^{n \times m}$ such that $\mathbf{K} = \mathbf{A}^\top$, our previous result demonstrates that $rk_{row}(\mathbf{K}) \leq rk_{col}(\mathbf{K})$. Because $\mathbf{K} = \mathbf{A}^\top$, this means that $rk_{col}(\mathbf{A}) \leq rk_{row}(\mathbf{A})$. This means that $rk_{col}(\mathbf{A}) \leq rk_{row}(\mathbf{A}) \leq rk_{col}(\mathbf{A})$ and

so proposition 0.5.2 that the $rk_{col}(\mathbf{A}) = rk_{row}(\mathbf{A})$ is proved.

Bringing it all together, we get the definition for the rank of a matrix:

Definition 0.5.1 (Rank) The number of linearly independent columns of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ equals the number of linearly independent rows and is called the *rank* of \mathbf{A} and is denoted by $rk(\mathbf{A})$.

0.6 Row Space and Column Space

Given a set of vectors \mathcal{C} , the $span[\mathcal{C}]$ is the set of all possible linear combinations of vectors in \mathcal{C} . This also forms a vector space with the set \mathcal{C} which in some sense is a consequence of the definition of the *span* being linear combinations of vectors.

Proposition 0.6.1 Given a set of vectors $\mathcal{S} = \{v_1, v_2, \dots, v_{k-1}, v_k\}$, the $span[\mathcal{S}]$ forms a vector space.

Proof: From definition 0.3.1 we know that all elements in $span[\mathcal{S}]$ can be represented as linear combinations of vectors in \mathcal{S}

$$\sum_{i=0}^k a_i \mathbf{v}_i \in \mathcal{S}, \forall a_1, a_2, \dots, a_k \in \mathbb{R} \quad (18)$$

For $span[\mathcal{S}]$ to be a vector space, it must meet certain conditions:

1. $(\mathcal{S}, +)$ forms an Abelian Group (commutative). This means it must meet the conditions for a group and be commutative:

(a) (Closure) $\otimes : \forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$:

$$\sum_{i=0}^k a_i \mathbf{v}_i + \sum_{i=0}^k b_i \mathbf{v}_i = \left[\sum_{i=0}^k (a_i + b_i) \mathbf{v}_i \right] \in \mathcal{S}, \forall a_i, b_i \in \mathbb{R} \quad (19)$$

(b) (Associativity) $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$:

$$\sum_{i=0}^k a_i \mathbf{v}_i + \left(\sum_{i=0}^k b_i \mathbf{v}_i + \sum_{i=0}^k c_i \mathbf{v}_i \right) = \sum_{i=0}^k (a_i + b_i + c_i) \mathbf{v}_i = \quad (20)$$

$$\left(\sum_{i=0}^k a_i \mathbf{v}_i + \sum_{i=0}^k b_i \mathbf{v}_i \right) + \sum_{i=0}^k c_i \mathbf{v}_i, \forall a_i, b_i, c_i \in \mathbb{R} \quad (21)$$

(c) (Neutral Element) $\exists e \in \mathcal{G} \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$:

$$\sum_{i=0}^k a_i \mathbf{v}_i + \sum_{i=0}^k (0) \mathbf{v}_i = \sum_{i=0}^k (0) \mathbf{v}_i + \sum_{i=0}^k a_i \mathbf{v}_i = \sum_{i=0}^k a_i \mathbf{v}_i \quad (22)$$

(d) (Inverse) $\forall x \in \mathcal{G} \exists y \in \mathcal{G} : x \otimes y = e$ and $y \otimes x = e$ where e is the inverse element:

$$\sum_{i=0}^k a_i \mathbf{v}_i + \sum_{i=0}^k -a_i \mathbf{v}_i = \sum_{i=0}^k -a_i \mathbf{v}_i + \sum_{i=0}^k a_i \mathbf{v}_i = \quad (23)$$

$$\sum_{i=0}^k (a_i - a_i) \mathbf{v}_i = \sum_{i=0}^k (0) \mathbf{v}_i \quad (24)$$

(e) (Commutativity) $\forall x, y \in \mathcal{G}, x \otimes y = y \otimes x$:

$$\sum_{i=0}^k a_i \mathbf{v}_i + \sum_{i=0}^k b_i \mathbf{v}_i = \sum_{i=0}^k b_i \mathbf{v}_i + \sum_{i=0}^k a_i \mathbf{v}_i = \sum_{i=0}^k (a_i + b_i) \mathbf{v}_i \quad (25)$$

Therefore, $(\mathcal{S}, +)$ forms an Abelian Group.

2. (Distributivity):

$$(a) \quad \forall \lambda \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V} : \lambda(\mathbf{x} + \mathbf{y}) = (\lambda \cdot \mathbf{x} + \lambda \cdot \mathbf{y}):$$

$$\lambda \left[\sum_{i=0}^k a_i \mathbf{v}_i + \sum_{i=0}^k b_i \mathbf{v}_i \right] = \lambda \sum_{i=0}^k a_i \mathbf{v}_i + \lambda \sum_{i=0}^k b_i \mathbf{v}_i, \forall \lambda \in \mathbb{R} \quad (26)$$

$$(b) \quad \forall \lambda, \psi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : (\lambda + \psi) \cdot \mathbf{x} = \lambda \cdot \mathbf{x} + \psi \cdot \mathbf{x}:$$

$$(\lambda + \phi) \sum_{i=0}^k a_i \mathbf{v}_i = \lambda \sum_{i=0}^k a_i \mathbf{v}_i + \phi \sum_{i=0}^k a_i \mathbf{v}_i \quad (27)$$

3. (Neutral Element) $\exists e \in \mathcal{G} \forall x \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$:

$$1 \cdot \sum_{i=0}^k a_i \mathbf{v}_i = \sum_{i=0}^k a_i \mathbf{v}_i \quad (28)$$

Thus demonstrating that the $\text{span}[\mathcal{S}]$ is a vector space and proving proposition 0.6.1.

We need to introduce two more definitions:

Definition 0.6.1 (Row Space) Let \mathcal{C} be a set of vectors of length n where each vector is a row of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. The row space of the matrix \mathbf{A} is the vector space of the span of the set of vectors \mathcal{C} .

Definition 0.6.2 (Column Space) Let \mathcal{C} be a set of vectors of length m where each vector is a column of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. The column space of the matrix \mathbf{A} is the vector space of the span of the set of vectors \mathcal{C} .

Proposition 0.6.2 (Row Space and Column Space) Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the row space is not always equal to the column space.

Proof: Let \mathbf{A} be a matrix defined as follows with the set of all row vectors being \mathcal{R} and the set of all column vectors being \mathcal{C} :

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{pmatrix} \quad (29)$$

The column space contains a vector \mathbf{v}_1 such that:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix} \quad (30)$$

It is clear that there exists no linear combination of the row vectors that could result in the vector \mathbf{v}_1 . Therefore, the $\text{span}[\mathcal{R}]$ does not contain a vector in $\text{span}[\mathcal{C}]$ and they cannot be equal. More generally, the row space and column space may not always be the same since they do not contain the same set of vectors.

0.7 Eigenvalues and Eigenvectors

One of the most important properties of a matrix in potentially all of linear algebra is the calculation / study of eigenvalues and eigenvectors. Its hard to overstate their significance. We will introduce them here and discuss a method for calculating them at a later point.

Definition 0.7.1 Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ where $n \in \mathbb{N}$ and two variables, a vector $\mathbf{u} \in \mathbb{R}^n$ and a scalar $\lambda \in \mathbb{R}$, if these variables solve the following equation, λ is an “eigenvalue” and \mathbf{u} is an “eigenvector” of λ .

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \tag{31}$$

There may be up to n eigenvalues and corresponding eigenvectors for the matrix \mathbf{A} .

Before we can delve into a method for calculating the eigenvalues and eigenvectors, we need to explore another property of a matrix known as the “determinant”.

0.8 Determinant

The determinant is a property of a matrix that describes whether it is a non-singular (invertible) matrix. Consider the following set of simultaneous equations:

$$2x - 3y = 4 \tag{32}$$

$$5x + 7y = 9 \tag{33}$$

A standard way to solve this is by rearranging one of the equations,

substituting this into the other equation and putting the derived value back into the original:

$$\begin{aligned}
 x &= \frac{1}{2}(4 + 3y) \\
 5\frac{1}{2}(4 + 3y) + 7y &= 9 \\
 10 + \frac{15}{2}y + 7y &= 9 \\
 10 + \frac{29}{2}y &= 9 \\
 y &= -\frac{29}{2}
 \end{aligned} \tag{34}$$

$$\begin{aligned}
 x &= \frac{1}{2}\left(4 + 3\left(-\frac{29}{2}\right)\right) \\
 x &= \frac{55}{29}
 \end{aligned} \tag{35}$$

This method is sufficient when we are solving two equations with two unknowns but it becomes more tedious when we are working with larger systems of equations. An alternative method is to setup a matrix equation and use methods from linear algebra to determine the unknowns. We first setup our system by collecting together the coefficients rewriting the system:

$$\begin{array}{l}
 2x - 3y = 4 \\
 5x + 7y = 9
 \end{array} \xrightarrow{\text{Matrix Representation}} \begin{bmatrix} 2 & -3 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \tag{36}$$

This representation allows us to use all the tools that linear algebra provides. If we take our matrix equation and rearrange the variables, we get a

way of deriving the solution for x and y :

$$2x - 3y = 4$$

$$5x + 7y = 9$$

$$\downarrow$$

$$\begin{bmatrix} 2 & -3 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

$$\mathbf{A}\mathbf{v} = \mathbf{B}$$

$$\mathbf{v} = \mathbf{A}^{-1}\mathbf{B}$$

(37)

The inverse of a matrix \mathbf{A} , denoted as \mathbf{A}^{-1} , is a matrix that satisfies the equation $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ where \mathbf{I} is the identity matrix. For the case of a matrix $\mathbf{A} \in \mathbb{R}^{2 \times 2}$, the inverse can be calculated using the “inverse matrix formula”:

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad (38)$$

With this, we can calculate the solution to our system of linear equations:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ 5 & 7 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad (39)$$

$$= \frac{1}{(2)(7) - (-3)(5)} \begin{bmatrix} 7 & 3 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad (40)$$

$$= \frac{1}{29} \begin{bmatrix} 7 & 3 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 9 \end{bmatrix} \quad (41)$$

$$= \begin{bmatrix} 55 \\ -2 \end{bmatrix} \quad (42)$$

$$= \begin{bmatrix} \frac{55}{29} \\ -\frac{2}{29} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (43)$$

This method could be considered a better choice because it generalises for much larger matrices. A part of the solution involved calculating the inverse of a matrix. In certain cases, the inverse of a matrix may not exist. This brings us to the determinant. The determinant is a scalar value derived as a function of all the elements of a matrix. It is only defined on squares matrices and if it is non-zero, the matrix is invertible. For a matrix $\mathbf{A} \in \mathbb{R}^{2 \times 2}$, the determinant of \mathbf{A} , denoted by $\det(\mathbf{A})$, is given by:

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \det(\mathbf{A}) = ad - bc \quad (44)$$

This makes sense for in the case of our system of linear equations. You can see

in equation 38 where the determinant features in the equation for the inverse of the matrix. If $\det(\mathbf{A}) = 0$ then the inverse cannot be calculated. This same condition applies to equations for the inverse of larger matrices that don't explicitly involve $\det(\mathbf{A})$. For square matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times m}$, the following properties hold:

1. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
2. $\det(\mathbf{A})^{-1} = \frac{1}{\det(\mathbf{A})}$
3. $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$

We can now revisit our discussion from section 0.7 on eigenvalues and eigenvectors. We can use the determinant to calculate these values. We can take the definition 0.7.1 and rewrite it into a form that we can solve:

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \tag{45}$$

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{I}\mathbf{u} \tag{46}$$

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0} \tag{47}$$

Proposition 0.8.1 For a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and an eigenvalue $\lambda \in \mathbb{R}$ with corresponding eigenvector $\mathbf{u} \in \mathbb{R}^n$:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{48}$$

Proof: Lets presume that the matrix $\mathbf{A} - \lambda\mathbf{I}$ has an inverse so that $(\mathbf{A} - \lambda\mathbf{I})^{-1}$

exists. Given 47, the following equation would be true:

$$\mathbf{u} = \mathbf{I}\mathbf{u} \quad (49)$$

$$= ((\mathbf{A} - \lambda\mathbf{I})^{-1}(\mathbf{A} - \lambda\mathbf{I}))\mathbf{u} \quad (50)$$

$$= (\mathbf{A} - \lambda\mathbf{I})^{-1}((\mathbf{A} - \lambda\mathbf{I})\mathbf{u}) \quad (51)$$

$$= (\mathbf{A} - \lambda\mathbf{I})^{-1}\mathbf{0} \quad (52)$$

$$= \mathbf{0} \quad (53)$$

However, given that $\mathbf{u} \neq \mathbf{0}$, we have a contradiction and the inverse of $\mathbf{A} - \lambda\mathbf{I}$ cannot exist. Therefore $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. We now have an equation involving the eigenvalues of a matrix that we can use to calculate that and the eigenvectors. Consider the following example for a matrix $\mathbf{A} \in \mathbb{R}^{2 \times 2}$:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \quad (54)$$

If we put this into $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ then we can calculate our eigenvalues:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \det \left(\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \quad (55)$$

$$= \det \left(\begin{bmatrix} -\lambda & 1 \\ -2 & -3 - \lambda \end{bmatrix} \right) \quad (56)$$

$$= (-\lambda)(-3 - \lambda) - (1)(-2) \quad (57)$$

$$0 = \lambda^2 + 3\lambda + 2 \quad (58)$$

$$0 = (\lambda + 2)(\lambda + 1) \Rightarrow \lambda_1 = -2, \lambda_2 = -1 \quad (59)$$

Now that we have our eigenvalues, we can use equation 47 to find the corre-

sponding eigenvectors:

$$\text{for } \lambda_1 = -1 : (\mathbf{A} - \lambda_1 \mathbf{I}) \mathbf{u}_1 = \mathbf{0} \quad (60)$$

$$\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_{1,1} \\ \mathbf{u}_{1,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (61)$$

$$\therefore \mathbf{u}_{1,1} = -\mathbf{u}_{1,2} \Rightarrow \mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (62)$$

$$(63)$$

$$\text{for } \lambda_1 = -2 : (\mathbf{A} - \lambda_1 \mathbf{I}) \mathbf{u}_2 = \mathbf{0} \quad (64)$$

$$\begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_{2,1} \\ \mathbf{u}_{2,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (65)$$

$$\therefore \mathbf{u}_{2,1} = -2\mathbf{u}_{2,2} \Rightarrow \mathbf{u}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad (66)$$

Proposition 0.8.2 Given a matrix \mathbf{A} with distinct eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_{k-1}, \lambda_k\}$ and corresponding eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k-1}, \mathbf{v}_k\}$, the set of eigenvectors are linearly independent.

Proof: From definition 0.7.1 we know that each eigenvalue and eigenvector pair satisfies $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i \forall i = 1, \dots, k$. from definition 0.2.2 we know that a set of vectors is linearly independent if no non-trivial linear combinations (see definition 0.2.1) equate to the zero vector. Knowing these definitions, we can use induction to prove the proposition. Consider the case when $k = 1$. A set with only one vector \mathbf{v}_1 is always linearly independent. Now presume that in the case of $k = i$, the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{i-1}, \mathbf{v}_i\}$ is linearly

independent. if we now set $k = i + 1$, we are concerned with determining if there is any non-trivial linear combinations that result in the zero vector in equation 67.

$$a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_{i-1} \mathbf{v}_{i-1} + a_i \mathbf{v}_i + a_{i+1} \mathbf{v}_{i+1} = \mathbf{0} \quad (67)$$

Firstly, we can alter this equation in two ways:. Firstly, we multiply both sides by \mathbf{A} and simplify:

$$\begin{aligned} \mathbf{A}(a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_{i-1} \mathbf{v}_{i-1} + a_i \mathbf{v}_i + a_{i+1} \mathbf{v}_{i+1}) &= \mathbf{A} \mathbf{0} \\ a_1 \mathbf{A} \mathbf{v}_1 + a_2 \mathbf{A} \mathbf{v}_2 + \dots + a_{i-1} \mathbf{A} \mathbf{v}_{i-1} + a_i \mathbf{A} \mathbf{v}_i + a_{i+1} \mathbf{A} \mathbf{v}_{i+1} &= \mathbf{0} \\ a_1 \lambda_1 + a_2 \lambda_2 + \dots + a_{i-1} \lambda_{i-1} + a_i \lambda_i + a_{i+1} \lambda_{i+1} &= \mathbf{0} \end{aligned} \quad (68)$$

Secondly, we can multiply both sides by λ_{i+1} :

$$\begin{aligned} \lambda_{i+1}(a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_{i-1} \mathbf{v}_{i-1} + a_i \mathbf{v}_i + a_{i+1} \mathbf{v}_{i+1}) &= \lambda_{i+1} \mathbf{0} \\ a_1 \lambda_{i+1} \mathbf{v}_1 + a_2 \lambda_{i+1} \mathbf{v}_2 + \dots + a_i \lambda_{i+1} \mathbf{v}_i + a_{i+1} \lambda_{i+1} \mathbf{v}_{i+1} &= \mathbf{0} \end{aligned} \quad (69)$$

If we subtract equation 69 from 68:

$$\begin{aligned} a_1(\lambda_1 - \lambda_{i+1}) \mathbf{v}_1 + a_2(\lambda_2 - \lambda_{i+1}) \mathbf{v}_2 + \dots \\ + a_i(\lambda_i - \lambda_{i+1}) \mathbf{v}_i + a_{i+1}(\lambda_{i+1} - \lambda_{i+1}) \mathbf{v}_{i+1} &= \mathbf{0} \end{aligned} \quad (70)$$

Since the difference between any two distinct eigenvalues is non-zero and all eigenvectors are non-zero, $a_1 = a_2 = \dots = a_i = 0$. substituting these into equation 67, This is reduced down to $a_{i+1} \mathbf{v}_{i+1} = \mathbf{0}$. Since $\mathbf{v}_{i+1} \neq 0$,

$a_{i+1} = 0$. This means that in the case of $k = i + 1$, the eigenvectors are linearly independent and proposition 0.8.2 is proven by induction.

In equation 58, we have a polynomial, the solutions of which are the eigenvalues of the matrix. This is called the characteristic equation. It is always the case that a square matrix of size n will have an n th degree polynomial as its characteristic equation. This leads us to an important finding.

Proposition 0.8.3 Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ where $n \in \mathbb{N}$ and eigenvalues $\lambda_1, \dots, \lambda_k$:

$$\det(\mathbf{A}) = \prod_{i=1}^k \lambda_i \quad (71)$$

where $k \leq n$.

Proof:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = P(\lambda) \quad (72)$$

$$= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_{n-1} - \lambda)(\lambda_n - \lambda) \quad (73)$$

If we let $\lambda = 0$:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = P(\lambda) \quad (74)$$

$$\det(\mathbf{A}) = P(0) \quad (75)$$

$$= (\lambda_1 - 0)(\lambda_2 - 0) \dots (\lambda_{n-1} - 0)(\lambda_n - 0) \quad (76)$$

$$= \prod_{i=1}^n \lambda_i \quad (77)$$

And therefore, proposition 0.8.3 is true. Furthermore, this fact can allow us to demonstrate the connection between the determinant of a matrix and whether it is invertible.

Proposition 0.8.4 Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ where $n \in \mathbb{N}$, \mathbf{A}^{-1} exists $\iff \det(\mathbf{A}) \neq 0$.

Proof: Lets presume that \mathbf{A}^{-1} exists. Then $\mathbf{A}\mathbf{w} = \mathbf{f}$ has a unique solution $\mathbf{w} = \mathbf{A}^{-1}\mathbf{f}$ for every \mathbf{f} . We know that from proposition 0.8.3 the only way $\det(\mathbf{A}) = 0$ is if one of its eigenvalues is 0. From definition 0.8.1, we know that $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$ for any eigenvalue λ of \mathbf{A} and its corresponding eigenvector \mathbf{v} . If one of our eigenvalues is equal to 0, then $(\mathbf{A} - 0\mathbf{I})\mathbf{v} = 0$ and $\mathbf{A}\mathbf{v} = 0$. If $\mathbf{A}\mathbf{w} = \mathbf{f}$ then we can say that $\mathbf{A}(\mathbf{w} + \mathbf{v}) = \mathbf{f}$. This means that if \mathbf{A} has an inverse and $\det(\mathbf{A}) = 0$, $\mathbf{A}\mathbf{w} = \mathbf{f}$ has multiple solutions, namely \mathbf{w} and $\mathbf{w} + \mathbf{v}$. Therefore, either the inverse does not exist or $\det(\mathbf{A}) \neq 0$. In other words, the inverse of \mathbf{A} only exists when $\det(\mathbf{A}) \neq 0$ and proposition 0.8.4 is proven.

We can summarise how to calculate determinants for larger matrices. Given a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$:

1. $n = 1$: $\det(\mathbf{A}) = \det(a_{11})$.
2. $n = 2$: $\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$.
3. $n = 3$: $\det(\mathbf{A}) = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$.
4. For $n > 3$:

Theorem 0.8.1 (Laplace Expansion). Consider a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$.

Then, for all $j = 1, \dots, n$:

(a) Expansion along column j :

$$\det(\mathbf{A}) = \sum_{k=1}^n (-1)^{k+j} a_{kj} \det(\mathbf{A}_{k,j}) \quad (78)$$

(b) Expansion along row j :

$$\det(\mathbf{A}) = \sum_{k=1}^n (-1)^{k+j} a_{jk} \det(\mathbf{A}_{j,k}) \quad (79)$$

Where $\mathbf{A}_{k,j} \in \mathbb{R}^{(n-1) \times (n-1)}$ is the submatrix of \mathbf{A} that we obtain when deleting row k and column j .

Bibliography

- [1] Georg Cantor. “Beiträge zur Begründung der transfiniten Mengenlehre”.
In: *Mathematische Annalen* 46.4 (1895), pp. 481–512.
- [2] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020, p. 36.