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. Vectors 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 $\bigotimes$ denote a binary operation between elements of this set such that $\bigotimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$. $G := (\mathcal{G}, \bigotimes)$ is called a group [2, page 36] if the following conditions are met:

1. (Closure) The binary operation between any two elements of the set will result in an element which is also part of the set: $\bigotimes : \forall x, y \in \mathcal{G} : x \bigotimes y \in \mathcal{G}$.

2. (Associativity) The order in which group operations are performed on a set of elements is irrelevant: $\forall x, y, z \in \mathcal{G} : (x \bigotimes y) \bigotimes z = x \bigotimes (y \bigotimes z)$.

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: $\exists e \in \mathcal{G} \forall x \in \mathcal{G} : x \bigotimes e = x$ and $e \bigotimes x = x$.

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: $\forall x \in \mathcal{G} \exists y \in \mathcal{G} : x \bigotimes y = e$ and $y \bigotimes x = e$, 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 $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ where the group operation is matrix multiplication. The inverse elements would be the inverse of the matrix and the neutral element would be the identity matrix. However, the order in which the group operation is performed will affect the result i.e $\boldsymbol{AB} \neq \boldsymbol{BA}$. Given a group G, if the order in which the group operation is performed does not matter i.e. $\forall x, y \in \mathcal{G} : x \bigotimes y = y \bigotimes x$, then $G = (\mathcal{G}, \bigotimes)$ 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:

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

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

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: $\forall \mathbf{x} \in \mathcal{V} : 1 \cdot \mathbf{x} = \mathbf{x}$

It is also the case that non-real valued vector spaces are permitted.

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 of the same degree are also vectors. Two can be added together, resulting in another polynomial of the same degree 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 $\mathbf{x}_i, \ldots, \mathbf{x}_k \in V$. Then, every $\mathbf{v} \in V$ of the form

$$\mathbf{v} = \lambda_1 \mathbf{x}_1 + \cdots + \lambda_k \mathbf{x}_k = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i \in V \tag{3}$$

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

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:

Proof: We can express any vector in $\mathbb{R}^2$ as $\mathbf{v} = (\alpha, \beta)^\top$ where $\alpha, \beta \in \mathbb{R}$. It is the case that $\boldsymbol{I}\mathbf{v} = \mathbf{v}$ for all $\mathbf{v}$. Therefore:

$$\boldsymbol{I}\mathbf{v} = \mathbf{v} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \alpha\mathbf{e}_1 + \beta\mathbf{e}_2 \tag{4}$$

Definition 0.2.2 (Linear Dependence) Consider a vector space V with $k \in \mathbb{N}$ and $\mathbf{x}_1, \cdots, \mathbf{x}_k \in V$. 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 $\mathbf{x}_1, \cdots, \mathbf{x}_k$ are *linearly dependent*.

Definition 0.2.3 (Linear Independence) Consider a vector space V with $k \in \mathbb{N}$ and $\mathbf{x}_1, \cdots, \mathbf{x}_k \in V$. If only the trivial solution exists, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i$ with all $\lambda_i = 0$, the vectors $\mathbf{x}_1, \cdots, \mathbf{x}_k$ are *linearly independent*.

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

$$\mathcal{V} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -4 \\ -1 \end{bmatrix} \right\} \tag{5}$$

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

$$\mathbf{y} = \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \beta \begin{bmatrix} -2 \\ -4 \end{bmatrix} + \omega \begin{bmatrix} -4 \\ -1 \end{bmatrix} \tag{6}$$

$$= \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 2\beta \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \omega \begin{bmatrix} -4 \\ -1 \end{bmatrix} \tag{7}$$

$$= (\alpha - 2\beta) \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \omega \begin{bmatrix} -4 \\ -1 \end{bmatrix} \tag{8}$$

$$\tag{9}$$

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:

$$\begin{bmatrix} -2 \\ -4 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \mathbf{0} \tag{10}$$

which following from definition 0.2.3 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 $\mathcal{A} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subseteq V$. If every vector $\mathbf{v} \in V$ can be expressed as a linear combination of vectors of $\mathbf{x}_1, \ldots, \mathbf{x}_k$, $\mathcal{A}$ is called a 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 $V = span[\mathcal{A}]$.

Consider the vector space $V = (\mathbb{R}^2, +, \cdot)$ under vector 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 $\mathcal{B}$ meaning that the $span[\mathcal{B}] = \mathbb{R}^2$. 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 (with a lower cardinality) $\tilde{\mathcal{A}} \subseteq \mathcal{A} \subseteq \mathcal{V}$ 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:

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \tag{11}$$

This is because every vector in $\mathbb{R}^3$ can be represented as linear combinations of the vectors in $\mathcal{A}$: *Proof*: Given a vector $\mathbf{v} \in \mathbb{R}^3$:

$$\mathbf{v} = \boldsymbol{I}\mathbf{v} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{12}$$

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 thus 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 defined as the number of vectors in the basis for the vector space.

0.5 Row Rank and Column Rank

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

Definition 0.5.1 For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ where $\boldsymbol{A}_{\cdot i}$ denotes the ith column vector:

$$rk_{col}(\boldsymbol{A}) = dim(span[\{\boldsymbol{A}_{\cdot 1}, \ldots, \boldsymbol{A}_{\cdot m}\}]) \tag{13}$$

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

$$s = \lambda_1 \boldsymbol{A}_{.1} + \lambda_2 \boldsymbol{A}_{.2}, \quad \forall \lambda_1, \lambda_2 \in \mathbb{R} \tag{14}$$

Because the column vectors are linearly dependent, by definition 0.2.3, $\exists \lambda_3, \lambda_4$ where $\lambda_3, \lambda_4 \neq 0$ such that:

$$\lambda_3 \boldsymbol{A}_{.1} + \lambda_4 \boldsymbol{A}_{.2} = \boldsymbol{0} \tag{15}$$

$$\boldsymbol{A}_{.1} = \frac{\lambda_4}{\lambda_3} \boldsymbol{A}_{.2} \tag{16}$$

Meaning that every linear combination can be written as:

$$s = \lambda_1 \boldsymbol{A}_{.1} + \lambda_2 \boldsymbol{A}_{.2} \tag{17}$$

$$= (-\lambda_1 \frac{\lambda_4}{\lambda_3} + \lambda_4) \boldsymbol{A}_{.2} \tag{18}$$

This means that $\mathcal{C} = \{\boldsymbol{A}_{.2}\}$ forms a basis for the set of column vectors of $\boldsymbol{A}$. Because the number of elements in this set is 1, the dimension of $span[C]$ is 1. Because both column vectors of $\boldsymbol{A}$ are linearly independent, the $rk_{col}(\boldsymbol{A}) = 1$.

Proposition 0.5.1 Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, the following statement is always true:

$$rk_{row}(\boldsymbol{A}) = rk_{col}(\boldsymbol{A}) \tag{19}$$

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

Proof: Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a matrix and let $u = rk_{col}(\boldsymbol{A})$. There exists a basis

$C = \{\mathbf{c}_1, \ldots, \mathbf{c}_u\}$ of $m \times 1$ column vectors that spans the same space spanned by the columns of $\mathbf{A}$. Let $\boldsymbol{B} = [\mathbf{c}_1|\mathbf{c}_2|\ldots|\mathbf{c}_{u-1}\mathbf{c}_u]$ 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 0.2.3, each column of $\boldsymbol{A}$ can be expressed as a linear combination of vectors in $\boldsymbol{B}$. If we collect the coefficients of these linear combinations into a matrix $\boldsymbol{D}$, we can represent $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{D}$ where $\boldsymbol{D} \in \mathbb{R}^{m \times n}$. Equally, we can see that the rows of $\boldsymbol{B}\boldsymbol{D}$ can be expressed as a linear combination of $\boldsymbol{D}$ with coefficients taken from B. $\therefore$, the span of the rows of $\boldsymbol{A}$ is no greater than the span of the rows of $\boldsymbol{D}$ because linear combinations of the rows of $\boldsymbol{A}$ can be written as linear combinations of the rows in $\boldsymbol{D}$. $\boldsymbol{D}$ 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 $rk_{row}(\boldsymbol{A}) \leq rk_{col}(\boldsymbol{A})$ for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. If we let $\boldsymbol{K} \in \mathbb{R}^{n \times m}$ such that $\boldsymbol{K} = \boldsymbol{A}^{\top}$, our previous result demonstrates that $rk_{row}(\boldsymbol{K}) \leq rk_{col}(\boldsymbol{K})$. Because $\boldsymbol{K} = \boldsymbol{A}^{\top}$, this means that $rk_{col}(\boldsymbol{A}) \leq rk_{row}(\boldsymbol{A})$. This means that $rk_{col}(\boldsymbol{A}) \leq rk_{row}(\boldsymbol{A}) \leq rk_{col}(\boldsymbol{A})$ and so proposition 0.5.1 that the $rk_{col}(\boldsymbol{A}) = rk_{row}(\boldsymbol{A})$ is proved.

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

Definition 0.5.2 (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 $span[C]$ 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 $\mathcal{S} = \{v_1, v_2, \ldots, v_{k-1}, v_k\}$, 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 $\mathcal{S}$

$$\sum_{i=0}^{k} a_i \mathbf{v}_i \in \mathcal{S}, \forall a_1, a_2, \ldots a_k \in \mathbb{R} \tag{20}$$

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) $\bigotimes : \forall x, y \in \mathcal{G} : x \bigotimes y \in \mathcal{G}$:

   $$\sum_{i=0}^{k} a_i \mathbf{v}_i + \sum_{i=0}^{k} b_i \mathbf{v}_i = \left[\sum_{i=0}^{k} (a_i + b_i) \mathbf{v}_i \right] \in \mathcal{S}, \forall a_i, b_i \in \mathbb{R} \tag{21}$$

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

   $$\sum_{i=0}^{k} a_i \mathbf{v}_i + \left(\sum_{i=0}^{k} b_i \mathbf{v}_i + \sum_{i=0}^{k} c_i \mathbf{v}_i \right) = \sum_{i=0}^{k} (a_i + b_i + c_i) \mathbf{v}_i = \tag{22}$$

   $$\left(\sum_{i=0}^{k} a_i \mathbf{v}_i + \sum_{i=0}^{k} b_i \mathbf{v}_i \right) + \sum_{i=0}^{k} c_i \mathbf{v}_i, \forall a_i, b_i, c_i \in \mathbb{R} \tag{23}$$

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

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

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

inverse element:

$$\sum_{i=0}^{k} a_i\mathbf{v}_i + \sum_{i=0}^{k} -a_i\mathbf{v}_i = \sum_{i=0}^{k} -a_i\mathbf{v}_i + \sum_{i=0}^{k} a_i\mathbf{v}_i = \tag{25}$$

$$\sum_{i=0}^{k} (a_i - a_i)\mathbf{v}_i = \sum_{i=0}^{k} (0)\mathbf{v}_i \tag{26}$$

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

$$\sum_{i=0}^{k} a_i\mathbf{v}_i + \sum_{i=0}^{k} b_i\mathbf{v}_i = \sum_{i=0}^{k} b_i\mathbf{v}_i + \sum_{i=0}^{k} a_i\mathbf{v}_i = \sum_{i=0}^{k} (a_i + b_i)\mathbf{v}_i \tag{27}$$

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

2. (Distributivity):

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

$$\lambda \left[\sum_{i=0}^{k} a_i\mathbf{v}_i + \sum_{i=0}^{k} b_i\mathbf{v}_i \right] = \lambda \sum_{i=0}^{k} a_i\mathbf{v}_i + \lambda \sum_{i=0}^{k} b_i\mathbf{v}_i, \forall \lambda \in \mathbb{R} \tag{28}$$

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

$$(\lambda + \phi) \sum_{i=0}^{k} a_i\mathbf{v}_i = \lambda \sum_{i=0}^{k} a_i\mathbf{v}_i + \phi \sum_{i=0}^{k} a_i\mathbf{v}_i \tag{29}$$

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

$$1 \cdot \sum_{i=0}^{k} a_i\mathbf{v}_i = \sum_{i=0}^{k} a_i\mathbf{v}_i \tag{30}$$

Thus demonstrating that the $span[\mathcal{S}]$ 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 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. The row space of the matrix $\mathcal{A}$ 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 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. The column space of the matrix $\boldsymbol{A}$ 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 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the row space is not always equal to the column space.

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

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \\ 3 & 1 \end{bmatrix} \tag{31}$$

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

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix} \tag{32}$$

It is clear that there exists no linear combination of the row vectors that could result in the vector $\mathbf{v}_1$. Therefore, the $span[\mathcal{R}]$ does not contain a vector in $span[\mathcal{C}]$ 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 $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ 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.

$$\boldsymbol{A}\mathbf{u} = \lambda\mathbf{u} \tag{33}$$

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

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

14

equations:

$$2x - 3y = 4 \tag{34}$$

$$5x + 7y = 9 \tag{35}$$

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:

$$x = \frac{1}{2}(4 + 3y)$$
$$5\frac{1}{2}(4 + 3y) + 7y = 9$$
$$10 + \frac{15}{2}y + 7y = 9$$
$$10 + \frac{29}{2}y = 9$$
$$y = -\frac{29}{2} \tag{36}$$

$$x = \frac{1}{2}(4 + 3(-\frac{29}{2}))$$
$$x = \frac{55}{29} \tag{37}$$

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:

$$\begin{matrix} 2x - 3y = 4 \\ 5x + 7y = 9 \end{matrix} \xrightarrow{\text{Matrix Representation}} \begin{bmatrix} 2 & -3 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \tag{38}$$

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:

$$2x - 3y = 4$$
$$5x + 7y = 9$$

$$\downarrow \tag{39}$$

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

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

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

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":

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{40}$$

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

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ 5 & 7 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 9 \end{bmatrix} \tag{41}$$

$$= \frac{1}{(2)(7) - (-3)(5)} \begin{bmatrix} 7 & 3 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 9 \end{bmatrix} \tag{42}$$

$$= \frac{1}{29} \begin{bmatrix} 7 & 3 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 9 \end{bmatrix} \tag{43}$$

$$= \begin{bmatrix} 55 \\ -2 \end{bmatrix} \tag{44}$$

$$= \begin{bmatrix} \frac{55}{29} \\ -\frac{2}{29} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \tag{45}$$

This method could be considered a better choice because it generalises for much larger systems of linear equations. 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 $\boldsymbol{A} \in \mathbb{R}^{2 \times 2}$, the determinant of $\boldsymbol{A}$, denoted by $\det(\boldsymbol{A})$, is given by:

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \det(\boldsymbol{A}) = ad - bc \tag{46}$$

This makes sense for in the case of our system of linear equations. You can see in equation 40 where the determinant features in the equation for the inverse of the matrix. If $\det(\boldsymbol{A}) = 0$ then the inverse cannot be calculated. This same condition applies to equations for the inverse of larger matrices that

don't explicitly involve $\det(\boldsymbol{A})$. For square matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}, \boldsymbol{B} \in \mathbb{R}^{m \times m}$, the following properties hold:

1. $\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$

2. $\det(\boldsymbol{A})^{-1} = \frac{1}{\det(\boldsymbol{A})}$

3. $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\top})$

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:

$$\boldsymbol{A}\mathbf{u} = \lambda\mathbf{u} \tag{47}$$

$$\boldsymbol{A}\mathbf{u} = \lambda\boldsymbol{I}\mathbf{u} \tag{48}$$

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\mathbf{u} = \mathbf{0} \tag{49}$$

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

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0 \tag{50}$$

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

exists. Given 49, the following equation would be true:

$$\mathbf{u} = \boldsymbol{I}\mathbf{u} \tag{51}$$

$$= ((\boldsymbol{A} - \lambda\boldsymbol{I})^{-1}(\boldsymbol{A} - \lambda\boldsymbol{I}))\mathbf{u} \tag{52}$$

$$= (\boldsymbol{A} - \lambda\boldsymbol{I})^{-1}((\boldsymbol{A} - \lambda\boldsymbol{I})\mathbf{u}) \tag{53}$$

$$= (\boldsymbol{A} - \lambda\boldsymbol{I})^{-1}0 \tag{54}$$

$$= 0 \tag{55}$$

However, given that $\mathbf{u} \neq 0$, we have a contradiction and the inverse of $\boldsymbol{A} - \lambda\boldsymbol{I}$ cannot exist. Therefore $\det(\boldsymbol{A} - \boldsymbol{\lambda I}) = 0$. 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 $\boldsymbol{A} \in \mathbb{R}^{2 \times 2}$:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \tag{56}$$

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

$$\det(\boldsymbol{A} - \boldsymbol{\lambda I}) = \det\left(\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \tag{57}$$

$$= \det\left(\begin{bmatrix} -\lambda & 1 \\ -2 & -3 - \lambda \end{bmatrix} \right) \tag{58}$$

$$= (-\lambda)(-3 - \lambda) - (1)(-2) \tag{59}$$

$$0 = \lambda^2 + 3\lambda + 2 \tag{60}$$

$$0 = (\lambda + 2)(\lambda + 1) \Rightarrow \lambda_1 = -2, \lambda_2 = -1 \tag{61}$$

Now that we have our eigenvalues, we can use equation 49 to find the corre-

sponding eigenvectors:

$$\text{for } \lambda_1 = -1 : (\boldsymbol{A} - \lambda\boldsymbol{I})\mathbf{u}_1 = \mathbf{0} \tag{62}$$

$$\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_{1,1} \\ \mathbf{u}_{1,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{63}$$

$$\therefore \mathbf{u}_{1,1} = -\mathbf{u}_{1,2} \Rightarrow \mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{64}$$

$$\tag{65}$$

$$\text{for } \lambda_1 = -2 : (\boldsymbol{A} - \lambda\boldsymbol{I})\mathbf{u}_2 = \mathbf{0} \tag{66}$$

$$\begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{u}_{2,1} \\ \mathbf{u}_{2,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{67}$$

$$\therefore \mathbf{u}_{2,1} = -2\mathbf{u}_{2,2} \Rightarrow \mathbf{u}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \tag{68}$$

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

Proof: From definition 0.7.1 we know that each eigenvalue and eigenvector pair satisfies $\boldsymbol{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i \forall i = 1, \ldots, k$. from definition 0.2.3 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 $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_i\}$ 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 69.

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_{i-1}\mathbf{v}_{i-1} + a_i\mathbf{v}_i + a_{i+1}\mathbf{v}_{i+1} = \mathbf{0} \tag{69}$$

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

$$\mathbf{A}(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_{i-1}\mathbf{v}_{i-1} + a_i\mathbf{v}_i + a_{i+1}\mathbf{v}_{i+1}) = \mathbf{A0}$$

$$a_1\mathbf{A}\mathbf{v}_1 + a_2\mathbf{A}\mathbf{v}_2 + \ldots + a_{i-1}\mathbf{A}\mathbf{v}_{i-1} + a_i\mathbf{A}\mathbf{v}_i + a_{i+1}\mathbf{A}\mathbf{v}_{i+1} = \mathbf{0}$$

$$a_1\lambda_1 + a_2\lambda_2 + \ldots + a_{i-1}\lambda_{i-1} + a_i\lambda_i + a_{i+1}\lambda_{i+1} = \mathbf{0} \tag{70}$$

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

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

$$a_1\lambda_{i+1}\mathbf{v}_1 + a_2\lambda_{i+1}\mathbf{v}_2 + \ldots + +a_i\lambda_{i+1}\mathbf{v}_i + a_{i+1}\lambda_{i+1}\mathbf{v}_{i+1} = \mathbf{0} \tag{71}$$

If we subtract equation 71 from 70:

$$a_1(\lambda_1 - \lambda_{i+1})\mathbf{v}_1 + a_2(\lambda_2 - \lambda_{i+1})\mathbf{v}_2 + \ldots$$

$$+a_i(\lambda_i - \lambda_{i+1})\mathbf{v}_i + a_{i+1}(\lambda_{i+1} - \lambda_{i+1})\mathbf{v}_{i+1} = \mathbf{0} \tag{72}$$

Since the difference between any two distinct eigenvalues is non-zero and all eigenvectors are non-zero, $a_1 = a_2 = \ldots = a_i = 0$. substituting these into equation 69, 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 60, 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 nth degree polynomial as its characteristic equation. This leads us to an important finding.

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

$$\det(\boldsymbol{A}) = \prod_{i=1}^{k} \lambda_i \tag{73}$$

where $k \leq n$.

Proof:

$$\det(\boldsymbol{A} - \boldsymbol{\lambda I}) = P(\lambda) \tag{74}$$

$$= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_{n-1} - \lambda)(\lambda_n - \lambda) \tag{75}$$

If we let $\lambda = 0$:

$$\det(\boldsymbol{A} - \boldsymbol{\lambda I}) = P(\lambda) \tag{76}$$

$$\det(\boldsymbol{A}) = P(0) \tag{77}$$

$$= (\lambda_1 - 0)(\lambda_2 - 0) \ldots (\lambda_{n-1} - 0)(\lambda_n - 0) \tag{78}$$

$$= \prod_{i=1}^{n} \lambda_i \tag{79}$$

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 $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ where $n \in \mathbb{N}$, $\boldsymbol{A}^{-1}$ exists $\iff \det(\boldsymbol{A}) \neq 0$.

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

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

1. $n = 1$: $\det(\boldsymbol{A}) = \det(\boldsymbol{a_{11}})$.

2. $n = 2$: $\det(\boldsymbol{A}) = a_{11}a_{22} - a_{12}a_{21}$.

3. $n = 3$: $\det(\boldsymbol{A}) = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$.

4. For $n > 3$:

Theorem 0.8.1 (Laplace Expansion). Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Then, for all $j = 1, \ldots, n$:

(a) Expansion along column j:

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A_{k,j}}) \tag{80}$$

(b) Expansion along row j:

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A_{j,k}}) \tag{81}$$

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

0.9 Linear Transformations

Linear transformations are concerned with going from one vector space to another. During the transformation, the properties of a vector space (defined in 0.1.2) are maintained.

Definition 0.9.1 A linear transformation between two vector spaces V and W is a map $\Psi : V \to W$ such that the following hold:

1. $\Psi(\mathbf{v}_1 + \mathbf{v}_2) = \Psi(\mathbf{v}_1) + \Psi(a\mathbf{v}_2)$ for any vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V.

2. $\Psi(\alpha\mathbf{v}) = \alpha\Psi(\mathbf{v})$ for any scalar $\alpha \in \mathbb{R}$.

After transformation is applied, straight lines in the first vector space will remain as straight lines in the final vector space. It is also the case that $\Psi(0) = 0$. There are several different types of linear transformations (or linear mappings) but we will restrict our focus to a small few. Given two vector spaces V and W and a linear transformation $\Phi : V \to W$:

1. (Injective) If $\forall \mathbf{x}, \mathbf{y} \in V : \Phi(\mathbf{x}) = \Phi(\mathbf{y}) \implies \mathbf{x} = \mathbf{y}$. What this means is that distinct elements in V are mapped onto distinct elements in W. It is not necessarily the case that every element in W can be reached from V.

2. (Surjective) If $\Phi(V) = W$. What this means is that every element in W can be reached from V. This doesn't need to be a unique relationship.

3. (Bijective) If Φ is bijective, then it is injective and surjective. "Bi" represents that elements come in pairs; one from V and one from W such that there is a unique relationship between elements from one vector space to the other. It is also the case that every element in W can be reached from V.

Given these definitions, we can describe the type of transformation that we are going to focus on: isomorphic transformations. This means that the transformation is linear (i.e. straight lines are maintained and zero is unchanged) and bijective. A transformation allows us to go from one coordinate system to another. It is important to clarify what a coordinate actually is.

Definition 0.9.2 (Coordinate) Consider a vector space V and an ordered basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of V. For any $\mathbf{x} \in V$ we obtain a unique representation (linear combination):

$$\mathbf{x} = \alpha \mathbf{b}_1 + \ldots + \alpha_n \mathbf{b}_n \tag{82}$$

of $\mathbf{x}$ with respect to B. Then $\alpha_1, \ldots, \alpha_n$ are the coordinates of $\mathbf{x}$ with respect to B, and the vector

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \tag{83}$$

is the coordinate vector / coordinate representation of $\mathbf{x}$ with respect to the ordered basis B.

In the case of $\mathbb{R}^2$, we have discussed previously how every vector $\mathbf{v}$ can be represented as linear combinations of $\mathbf{e}_1 = (1,0)^\top$ and $\mathbf{e}_2 = (0,1)^\top$ such that $\forall \mathbf{v} \in \mathbb{R}^2$, there exists $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\mathbf{v} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2$. In this case (and in the case of all ordered sets of basis vectors) the coordinate would be (α_1, α_2). This is why you often see coordinates in applied mathematics represented as $\alpha_1 \mathbf{i} + \alpha_2 \mathbf{j}$. Lets suppose we had a vector space V and two sets of vectors $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$. By definition 0.4.1, we can write each vector in $\mathcal{A}$ as a unique linear combination of the vectors in $\mathcal{B}$. Given scalars c_{ij} where $1 \le i, j \le n$:

$$
\begin{aligned}
\mathbf{a}_1 &= c_{1,1}\mathbf{b}_1 + c_{1,2}\mathbf{b}_{1,2} + \ldots + c_{1,n}\mathbf{u}_n \\
& \quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\mathbf{a}_n &= c_{n,1}\mathbf{b}_1 + c_{n,2}\mathbf{b}_{1,2} + \ldots + c_{n,n}\mathbf{u}_n
\end{aligned}
$$

Lets introduce a matrix $\boldsymbol{G} = (c_{ij})_{ij}$ which we will call the **change-of-basis matrix**. Any vector $\mathbf{v}$ can now be represented in the following way:

$$
\mathbf{a} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{84}
$$

$$
= \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_n \end{pmatrix} \boldsymbol{G} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{85}
$$

In other words, to construct a matrix which allows us to transform any coordinate constructed from the basis vectors in $\mathcal{B}$ to those in $\mathcal{A}$, we can use

the coordinates of the basis vectors themselves to define the transformation. Furthermore, This transformation is injective.

Example: Consider a vector space V and a vector $\mathbf{v} = 15\mathbf{e}_1 - 5\mathbf{e}_2$ where $\mathbf{e}_1 = (1,0)^\top$ and $\mathbf{e}_2 = (0,1)^\top$ form a basis for V. We want to take the coordinate $(15,-5)$ and represent it using two new basis vectors of V, $\mathbf{d}_1 = (1,1)^\top$ and $\mathbf{d}_2 = (3,-2)^\top$. From definition 0.4.1 and definition 0.9.2, we want to find $\alpha, \beta \in \mathbb{R}$ such that:

$$\begin{bmatrix} 15 \\ 2 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{86}$$

$$\begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ -5 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \tag{87}$$

Therefore, the coordinate $(15,-5)$ with respect to the basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ is equivalent to $(3,4)$ with respect to the basis vectors $\mathbf{d}_1$ and $\mathbf{d}_2$.

In this example, an important concept appears. Linear transformations can be represented as a matrix. More precisely, this is how they're defined. Suppose we wanted to define a linear transformation as a matrix that represented a reflection in the line $y = x$. Let V be a vector space and $\mathbf{v} = (\alpha, \beta)$ be a vector in that vector space. V has basis vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ meaning that $\mathbf{v} = \alpha\mathbf{d}_1 + \beta\mathbf{d}_2$. Let W be a vector space and $\mathbf{w} = (\delta, \epsilon^\prime)$ be a vector in that vector space. W has basis vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ meaning that $\mathbf{v} = \delta\mathbf{f}_1 + \epsilon\mathbf{f}_2$. Let

$\boldsymbol{A}$ be the matrix that defines the reflection in the line $y = x$:

$$\mathbf{v} = \alpha\mathbf{d}_1 + \beta\mathbf{d}_2 \tag{88}$$

$$\boldsymbol{M}\mathbf{v} = \boldsymbol{M}[\alpha\mathbf{d}_1 + \beta\mathbf{d}_2] \tag{89}$$

$$= \alpha\boldsymbol{M}\mathbf{d}_1 + \beta\boldsymbol{M}\mathbf{d}_2 \tag{90}$$

$$= \alpha\mathbf{f}_1 + \beta\mathbf{f}_2 \tag{91}$$

What this demonstrates is that if we consider how the basis vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ for V are changed after the described transformation, we can construct the matrix $\boldsymbol{M}$ with each column vector being the ordered basis vectors $\mathbf{f}_1$ and $\mathbf{f}_2$. This then allows us to perform the reflection on any coordinate in V.

0.10 Matrix Decompositions

28

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.