Department of Mathematics, College of Engineering,

Design and Physical Sciences, Brunel University

Mathematics (MMath)

Academic Year 2022 - 2023

The Singular Value Decomposition

By David Blair

Supervisor: Dr Simon Shaw

1 Abstract

In this paper, we explore some of the foundational concepts in linear algebra. We aim to provide intuition behind commonly used propositions and a number of proofs are given for some non-trivial results. After the fundamental concepts are covered, we delve into two types of matrix decomposition leading ultimately to an explorations of the singular value decomposition. This is followed by a practical investigation where some python code has been written to allow the reader to easily apply this decomposition to both grey scale and coloured images. We finish with some recommendations as to further practical work that could be undertaken.

Contents

2 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 2.1 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 [4] 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 [21] is a special type of group with some additional conditions.

Definition 2.2 (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.

3 Linear Independence

Linear independence [5, 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 3.1 (Linear Combination) [5, page 40] 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:

Example: 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 3.2 (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 3.3 (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 3.3 demonstrates that we have a linearly dependent set of vectors.

4 Generating Set and Span

A generating set and the span [5, 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 4.1 (Generating Set and Span) [9] 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 $V = \textit{span}[\mathcal{A}]$.

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 4.1, 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 $\textit{span}[\mathcal{B}] = \mathbb{R}^2$. In this sense, a generating set and the span of a set are two sides of the same coin.

5 Basis

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

Definition 5.1 (Basis) [20] 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 $\mathcal{A}$: *Example*: 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.

6 Row Rank and Column Rank

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

Definition 6.1 For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ where $\boldsymbol{A}_{.i}$ denotes the ith column vector [**rank˙dim˙span**]:

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

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 3.3, $\exists \lambda_3, \lambda_4$ 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 $rk_{\text{col}}(\boldsymbol{A}) = 1$.

Proposition 6.1 Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$, the following statement is always true:

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

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

Proof [16]: Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ be a matrix and let $u = rk_{\text{col}}(\boldsymbol{A})$. 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 3.3, 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 $\boldsymbol{B}$. $\therefore$, 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 $rk_{\text{row}}(\boldsymbol{A}) \leq rk_{\text{col}}(\boldsymbol{A})$ 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 $rk_{\text{row}}(\boldsymbol{K}) \leq rk_{\text{col}}(\boldsymbol{K})$. Because $\boldsymbol{K} = \boldsymbol{A}^\top$, this means that $rk_{\text{col}}(\boldsymbol{A}) \leq rk_{\text{row}}(\boldsymbol{A})$. This means that $rk_{\text{col}}(\boldsymbol{A}) \leq rk_{\text{row}}(\boldsymbol{A}) \leq rk_{\text{col}}(\boldsymbol{A})$ and so proposition 6.1 that the $rk_{\text{col}}(\boldsymbol{A}) = rk_{\text{row}}(\boldsymbol{A})$ is proved.

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

Definition 6.2 (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 $rk(\mathbf{A})$.

7 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 7.1 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 4.1 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) $\forall x, y, z \in \mathcal{G} : (x \bigotimes y) \bigotimes z = x \bigotimes (y \bigotimes z)$:

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

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

   (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 7.1.

We need to introduce two more definitions [3]:

Definition 7.1 (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 7.2 (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 7.2 (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.

8 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 8.1 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 [6].

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

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".

9 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:

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

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

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:

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

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

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:

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

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:

$$\begin{gathered} 2x - 3y = 4 \\ 5x + 7y = 9 \\ \downarrow \\ \begin{bmatrix} 2 & -3 \\ 5 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \end{bmatrix} \\ \boldsymbol{A}\mathbf{v} = \mathbf{B} \\ \mathbf{v} = \boldsymbol{A}^{-1}\mathbf{B} \end{gathered} \tag{39}$$

The inverse of a matrix $\boldsymbol{A}$, denoted as $\boldsymbol{A}^{-1}$, is a matrix that satisfies the equation $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix. For the case of a matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$, the inverse can be calculated using the "inverse matrix formula" [8]:

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

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:

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

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 [5, page 103]:

1. $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$
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 8 on eigenvalues and eigenvectors. We can use the determinant to calculate these values. We can take the definition 8.1 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 9.1 For a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and an eigenvalue $\lambda \in \mathbb{R}$ with corresponding eigenvector $\mathbf{u} \in \mathbb{R}^n$:

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

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:

$$
\begin{aligned}
\mathbf{u} &= \boldsymbol{I}\mathbf{u} && (51)\\
&= ((\boldsymbol{A} - \lambda\boldsymbol{I})^{-1}(\boldsymbol{A} - \lambda\boldsymbol{I}))\mathbf{u} && (52)\\
&= (\boldsymbol{A} - \lambda\boldsymbol{I})^{-1}((\boldsymbol{A} - \lambda\boldsymbol{I})\mathbf{u}) && (53)\\
&= (\boldsymbol{A} - \lambda\boldsymbol{I})^{-1}0 && (54)\\
&= 0 && (55)
\end{aligned}
$$

However, given that $\mathbf{u} \neq 0$, we have a contradiction and the inverse of $\boldsymbol{A} - \lambda\boldsymbol{I}$ cannot exist. Therefore $\det(\boldsymbol{A} - \boldsymbol{\lambda}\boldsymbol{I}) = 0$. 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 $\det(\boldsymbol{A} - \boldsymbol{\lambda}\boldsymbol{I}) = 0$ then we can calculate our eigenvalues:

$$
\begin{aligned}
\det(\boldsymbol{A} - \boldsymbol{\lambda}\boldsymbol{I}) &= \det\left(\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) && (57)\\
&= \det\left(\begin{bmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{bmatrix}\right) && (58)\\
&= (-\lambda)(-3-\lambda) - (1)(-2) && (59)\\
0 &= \lambda^2 + 3\lambda + 2 && (60)\\
0 &= (\lambda+2)(\lambda+1) \Rightarrow \lambda_1 = -2, \lambda_2 = -1 && (61)
\end{aligned}
$$

Now that we have our eigenvalues, we can use equation 49 to find the corresponding 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 9.2 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 [11]: From definition 8.1 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 3.3 we know that a set of vectors is linearly independent if no non-trivial linear combinations (see definition 3.1) 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 $\boldsymbol{A}$ and simplify:

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

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

$$\begin{aligned} \lambda_{i+1}(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_{i-1}\mathbf{v}_{i-1} + a_i\mathbf{v}_i + a_{i+1}\mathbf{v}_{i+1}) &= \lambda_{i+1}\mathbf{0} \\ a_1\lambda_{i+1}\mathbf{v}_1 + a_2\lambda_{i+1}\mathbf{v}_2 + \ldots + +a_i\lambda_{i+1}\mathbf{v}_i + a_{i+1}\lambda_{i+1}\mathbf{v}_{i+1} &= \mathbf{0} \end{aligned} \tag{71}$$

If we subtract equation 71 from 70:

$$\begin{aligned} &a_1(\lambda_1 - \lambda_{i+1})\mathbf{v}_1 + a_2(\lambda_2 - \lambda_{i+1})\mathbf{v}_2 + \ldots \\ &+a_i(\lambda_i - \lambda_{i+1})\mathbf{v}_i + a_{i+1}(\lambda_{i+1} - \lambda_{i+1})\mathbf{v}_{i+1} = \mathbf{0} \end{aligned} \tag{72}$$

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 9.2 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 9.3 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 [18]:

$$\begin{aligned}
\det(\boldsymbol{A} - \boldsymbol{\lambda I}) &= P(\lambda) && (74)\\
&= (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_{n-1})(\lambda - \lambda_n) && (75)\\
&= (\lambda_1 - \lambda)(\lambda_2 - \lambda)\ldots(\lambda_{n-1} - \lambda)(\lambda_n - \lambda) && (76)
\end{aligned}$$

If we let $\lambda = 0$:

$$\begin{aligned}
\det(\boldsymbol{A} - \boldsymbol{\lambda I}) &= P(\lambda) && (77)\\
\det(\boldsymbol{A}) &= P(0) && (78)\\
&= (\lambda_1 - 0)(\lambda_2 - 0)\ldots(\lambda_{n-1} - 0)(\lambda_n - 0) && (79)\\
&= \prod_{i=1}^{n} \lambda_i && (80)
\end{aligned}$$

And therefore, proposition 9.3 is true. Furthermore, this fact can allow us to demonstrate the connection between the determinant of a matrix and whether it is invertible.

Proposition 9.4 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 9.3 the only way $\det(\boldsymbol{A}) = 0$ is if one of its eigenvalues is 0. From definition 9.1, 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 9.4 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 9.1 (Laplace Expansion). Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. Then, for all $j = 1, \ldots, n$ [5, page 100]:

   (a) Expansion along column j:

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

(b) Expansion along row j:

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n}(-1)^{k+j}a_{jk}\det(\boldsymbol{A_{j,k}}) \tag{82}$$

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.

10 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 2.2) are maintained.

Definition 10.1 A linear transformation between two vector spaces V and W is a map $\Psi : V \to W$ such that the following hold [12]:

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 $\Phi : V \to W$ [5, page 48]:

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 10.2 (Coordinate) [5, page 50] 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):

$$\mathbf{x} = \alpha \mathbf{b}_1 + \ldots + \alpha_n \mathbf{b}_n \tag{83}$$

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

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n \tag{84}$$

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 5.1, we can write each vector in $\mathcal{A}$ as a unique linear combination of the vectors in $\mathcal{B}$. Given scalars c_{ij} where $1 \leq i, j \leq n$:

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

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:

$$\mathbf{a} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \ldots \quad \mathbf{a}_n) \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{85}$$

$$= (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \ldots \quad \mathbf{b}_n)\, \boldsymbol{G} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \tag{86}$$

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 [2]: 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 5.1 and definition 10.2, we want to find $\alpha, \beta \in \mathbb{R}$ such that:

$$\begin{bmatrix} 15 \\ 2 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{87}$$

$$\begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ -5 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \tag{88}$$

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 $\mathbf{w} = (\delta, \epsilon`)$ 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 $y = x$:

$$\mathbf{v} = \alpha \mathbf{d}_1 + \beta \mathbf{d}_2 \tag{89}$$

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

$$= \alpha \boldsymbol{M}\mathbf{d}_1 + \beta \boldsymbol{M}\mathbf{d}_2 \tag{91}$$

$$= \alpha \mathbf{f}_1 + \beta \mathbf{f}_2 \tag{92}$$

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.

11 Matrix Decompositions

We will now move onto the first of two matrix decompositions, the eigenvalue decomposition or eigendecomposition. You can think of matrix decompositions as a form of factorisation, breaking one matrix into the product of several others.

Theorem 11.1 (Eigendecomposition).[5, page 116] A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ can be factored into

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$$

where $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{D}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{A}$, if and only if the eigenvectors of $\boldsymbol{A}$ form a basis of $\mathbb{R}^n$.

Proof: The proof of its existence follows from definition 8.1. Let $\boldsymbol{D}$ be a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P} \in$

$\mathbb{R}^{n \times n}$ where the ith column is the $\mathbf{p}_i$ eigenvector of $\boldsymbol{A}$. For all i:

$$\boldsymbol{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i \tag{93}$$

$$\boldsymbol{AP} = \boldsymbol{PD} \tag{94}$$

$$\boldsymbol{A} = \boldsymbol{PDP}^{-1} \tag{95}$$

Example: We will use the matrix from equation 56 which had the following eigenvalues and eigenvectors:

$$\lambda_1 = -1 : \mathbf{p}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \lambda_2 = -2 : \mathbf{p}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \tag{96}$$

First we want to construct the $\boldsymbol{P}$ and $\boldsymbol{D}$ matrices:

$$\boldsymbol{P} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \quad \boldsymbol{D} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \tag{97}$$

And then to calculate the inverse of $\boldsymbol{P}$:

$$\boldsymbol{P}^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}^{-1} = -1 \begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \tag{98}$$

From which we can perform the decomposition:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \tag{99}$$

As we've described previously, linear transformations can be represented using matrices. Equally, all matrices represent a specific linear transformation. We can use this to provide some intuition for the eigendecomposition. The

matrix $\boldsymbol{P}$ first takes the standard basis vectors onto the eigenbasis. This is a basis in $\mathbb{R}^2$ formed from the eigenvectors of the matrix $\boldsymbol{A}$. Since $\boldsymbol{D}$ is a diagonal matrix, its transformation is equivalent to scaling each eigenbasis by their corresponding eigenvalues. Finally, $\boldsymbol{P}^{-1}$ performs the opposite transformation to the change of basis to the eigenbasis and results in the standard basis vectors being scaled by $\boldsymbol{A}$. Another way to think of this decomposition is that it is the matrix representation of the equation $\boldsymbol{A}\mathbf{p}_i = \lambda_i \mathbf{p}_i$ for all eigenvector eigenvalue pairs of the matrix $\boldsymbol{A}$.
The reason we've reintroduced this decomposition is because it allows us to demonstrate the existence of what this dissertation is about, the singular value decomposition:

Theorem 11.2 *(SVD Theorem)* [5, page 119] Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be a rectangular matrix of rank $r \in [0, min(m, n)]$. The SVD of $\boldsymbol{A}$ is a decomposition of the form:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

with an orthogonal matrix $U \in \mathbb{R}^{m\times m}$ with column vectors $\mathbf{u}_i$, $i = 1, \ldots, m$, and an orthogonal matrix $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ with column vectors $\mathbf{v}_j$, $j = 1, \ldots, n$. Moreover, $\boldsymbol{\Sigma}$ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i \leq 0$ and $\Sigma_{ij} = 0$, $i \neq j$.

We can think of the singular value decomposition as a method that allows us to approximate a large matrix, essentially capturing the important information. In order to demonstrate its existence, we can build on the eigenvalue decomposition discussed previously.

The singular value decomposition always exists for a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ whereas the eigendecomposition is only defined for square matrices $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. However, we can use the eigendecomposition to demonstrate the existence of the singular value decomposition (summarised in [7] but originating from [19]). We will consider real matrices however the SVD is defined equally well on matrices of complex values.

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with rank r. The matrix $\boldsymbol{A}^\top \boldsymbol{A}$ is symmetric and positive definite (PSD) [17]. Therefore, we can diagonalise the matrix with an eigendecomposition:

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{v}\boldsymbol{\Lambda}\boldsymbol{V}^\top = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^\top = \sum_{i=1}^{n} (\sigma_i)^2 \mathbf{v}_i \mathbf{v}_i^\top \tag{100}$$

We have used a new quantity, namely σ_i (the singular values) as the square root of the i-th eigenvalue. This is fine because PSD matrices have non-negative eigenvalues. We know that from the definition of an eigenvalue-eigenvector pair:

$$\boldsymbol{A}^\top \boldsymbol{A} \mathbf{v}_i = (\sigma_i)^2 \mathbf{v}_i \tag{101}$$

Lets construct a new vector $\mathbf{u}_i$ such that:

$$\mathbf{u}_i = \frac{\boldsymbol{A}\mathbf{v}_i}{\sigma_i} \tag{102}$$

By construction, $\mathbf{u}_i$ is a unit vector of $\boldsymbol{A}\boldsymbol{A}^\top$. Let $\boldsymbol{V}$ be a matrix such that $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ where the i-th column of $\mathbf{v}_i$ because $\boldsymbol{A}\boldsymbol{A}^\top \in \mathbb{R}^{n \times n}$. Let $\boldsymbol{U}$ be a matrix such that $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ where the i-th column $\mathbf{u}_i$ because $\boldsymbol{A}\mathbf{v}_i$ is an m-vector. Finally, let $\boldsymbol{\Sigma}$ be a diagonal matrix whose i-th element is σ_i. We can express the relationships we've discussed so far in matrix form as:

$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{V}\boldsymbol{\Sigma}^{-1} \tag{103}$$

$$\boldsymbol{U}\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{V} \tag{104}$$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \tag{105}$$

where we use the fact $\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{I}$ and $\boldsymbol{\Sigma}^\top$ is a diagonal matrix where the i-th value is the reciprocal of σ_i. This means that given the eigendecomposition, we can demonstrate the existence of the singular value decomposition.

Example: Consider the following matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$:

$$\boldsymbol{A} = \begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix} \tag{106}$$

We first need to calculate the eigenvalues and eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$:

$$\lambda_1 = 40 : \mathbf{p}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \lambda_2 = 10 : \mathbf{p}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{107}$$

And then construct the eigendecomposition:

$$\boldsymbol{A}^\top\boldsymbol{A} = \begin{bmatrix} 25 & -15 \\ -15 & 25 \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 40 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{108}$$

The $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ can then be calculated as the $\boldsymbol{V} = \boldsymbol{P}$ and $\boldsymbol{\Sigma}$ is the element-wise square root of $\boldsymbol{D}$:

$$\boldsymbol{V} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} = \begin{bmatrix} \sqrt{40} & 0 \\ 0 & \sqrt{10} \end{bmatrix} \tag{109}$$

After which the columns of $\boldsymbol{U}$ can be calculated as follows:

$$\mathbf{u}_1 = \frac{1}{\sigma_1}\boldsymbol{A}\mathbf{v}_1 = \frac{1}{\sqrt{40}}\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix} \tag{110}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2}\boldsymbol{A}\mathbf{v}_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} 4 & 0 \\ 3 & -5 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{bmatrix} \tag{111}$$

$$\boldsymbol{U} = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix} \tag{112}$$

Leading to the construction of the singular value decomposition:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{bmatrix}\begin{bmatrix} \sqrt{40} & 0 \\ 0 & \sqrt{10} \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{113}$$

An important property of the singular value decomposition is that it allows us to decompose a matrix of any size as the sum of rank-1 matrices. The more rank-1 matrices we use in the summation, the closer the matrix comes to being the original matrix. Its because of this that we have a way to approximate a high rank matrix with one of a lower rank [10].

$$\boldsymbol{A}(k) = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i \tag{114}$$

An interesting component of the equation above is that each rank-1 matrix is weighted by its corresponding singular value. This means that the more rank-1 matrices we include in this approximation, the lower the effect of each additional rank-1 matrix. A great way to visualise what is going on here is by considering a greyscale image as a matrix. consider the following image: We are going to use python for visualisations and calculations. We first need to import a number of libraries [13] [14] [15]:

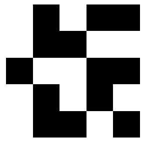


Figure 1: Greyscale test image

```
import matplotlib.pyplot as plt
import numpy as np
from PIL import Image
```

Using the Python Imaging Library (PIL) We can create a matrix of the greyscale values of this image:

```
img = Image.open("pixel_art.png")
arr = np.asarray(img.getdata(3))
arr = arr.reshape(5, 5)
```

Which will produce the following matrix:

$$\begin{bmatrix} 0 & 255 & 0 & 255 & 255 \\ 0 & 255 & 255 & 0 & 0 \\ 255 & 0 & 0 & 255 & 255 \\ 0 & 255 & 0 & 255 & 0 \\ 0 & 255 & 255 & 0 & 255 \end{bmatrix} \tag{115}$$

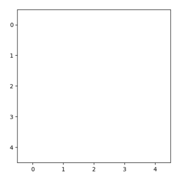
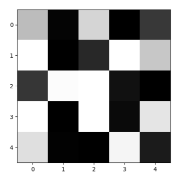
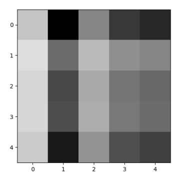
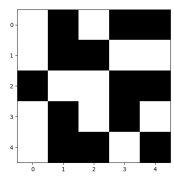
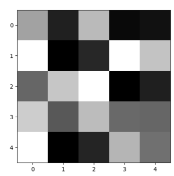
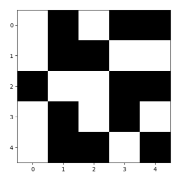
And now we can perform the Singular Value Decomposition on this matrix:

```
U, s, V = np.linalg.svd(arr, full_matrices=True)
```

Using equation 114 we can create a number of approximations each with increasing rank. Note that when plotting the image, matplotlib inverts the colour scale so we need to subtract every value from 255.

```
r = 1 #r = 5 produces the original image
aprx = 255 - np.round(U[:, :r] * s[:r] * np.matrix(V[:r, :]))
plt.imshow(aprx, cmap = "gray", vmin = 0, vmax = 255)
plt.show()
```

Below we've generated several approximations of the image:

Components	Approximation	Components	Approximation
r = 0		r = 3	
r = 1		r = 4	
r = 2		r = 5	

When r = 1, we get a single rank-1 component of the image. You can see that each row or column is some linear combination of every other row or column respectively. That's why you get that sort of grid pattern appearing. Since the SVD can decompose a matrix into the sum of rank-1 matrices, variations of this grid like pattern, when layered on top of each other, can give us any matrix we want.

This process we've described is fine for a greyscale image, but what if we wanted to apply this to coloured images? Firstly, we need to create a test image which we will use:

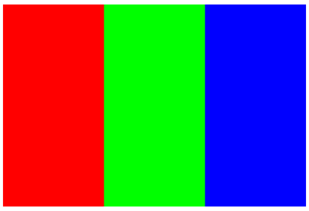


Figure 2: Test RGB image for SVD

Each pixel in a coloured image is represented as a tuple of three integer values indicating the intensity of the red, green and blue components. This image is 200 x 300 pixels in size. Its split into three 200 x 100 segments with each part taking the maximum intensity of its respective colour. The colour of each pixel in the red section is (255, 0, 0) , (0, 255, 0) in the green section and (0, 0, 255) in the blue section. We can break this down into three images where each pixel is the intensity of the respective colour:

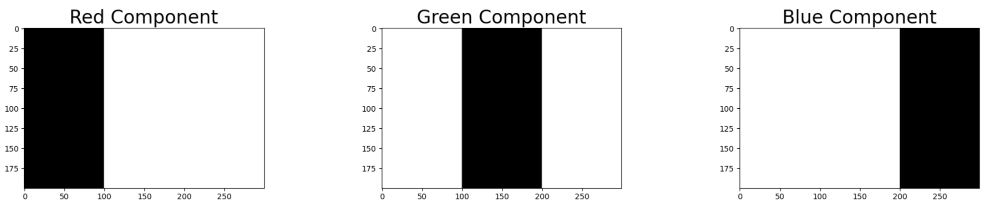


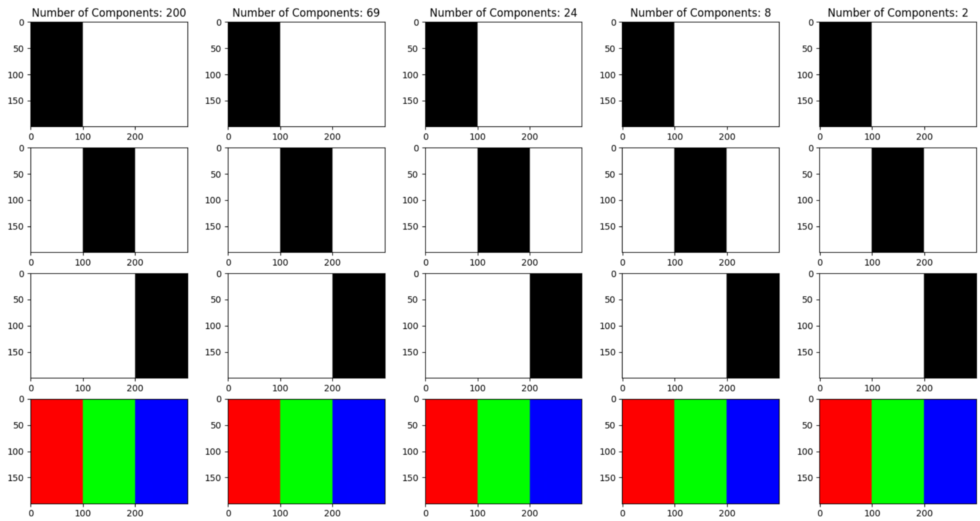
Figure 3: RGB components of the test image

Note that each component is now a greyscale image because each pixel is only taking values between 1 and 255 inclusively. We've written some python code that can do a lot with regards to deconstructing, decomposition and plotting various approximations of any greyscale and RGB image which can be found at `https://github.com/DavidBlairs/Advanced_Project/tree/main/Practical`. We won't describe how the functions work but we will include what code is needed to achieve each image.

Using the following code, we can plot 5 approximations of our RGB test image.

```
test = image_svd("test_image.png")
test.plot_n_approximations_rgb(5)
plt.show()
```

The first three rows break down each colour component into increasing approximations of the matrix for that colour component. These approximations are then layered back onto each other to produce the image in the bottom row.



One question we might want to ask about the number of components is whether we can quantify the effect of increasing the number of components with respect to how much variance of the original image is captured in every component. An important aspect of the singular value decomposition is that the singular values themselves can tell us the relative importance of each component. This is because the summand in 114 is weighted by each singular value. We can formalise the variance explained by each component in the following formula [5, page 131]:

$$R_i^2 = \frac{s_i^2}{\sum_j s_j^2} \tag{116}$$

The code we've written can plot this information for a certain number of singular values and for each RGB component.

```
test = image_svd("test_image.png")
test.plot_singular_rgb(10)
plt.show()
```

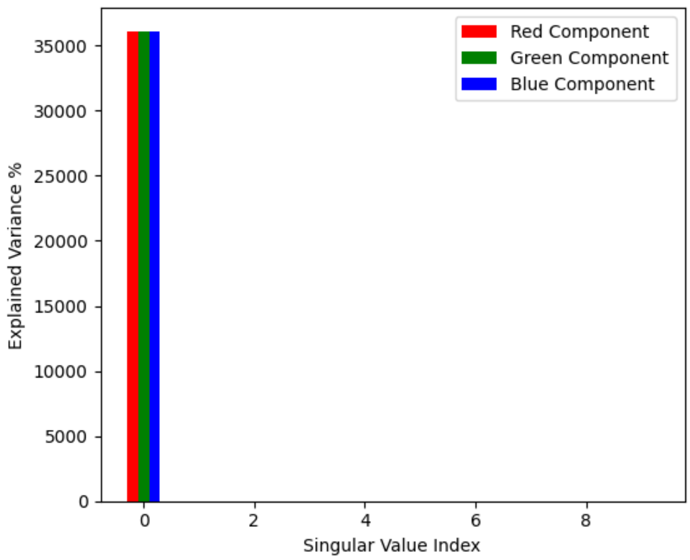


Figure 4: Variance Explained for each colour component

This bar chart is for our RGB test image. You may notice that 100% of the variance of the original image is captured in just one component. This makes sense since the test image only has a rank of 1 so all the information is contained in just the first component. Now that we've shown how our code works for a test image, we can start to experiment with some more interesting images. Consider figure 5:

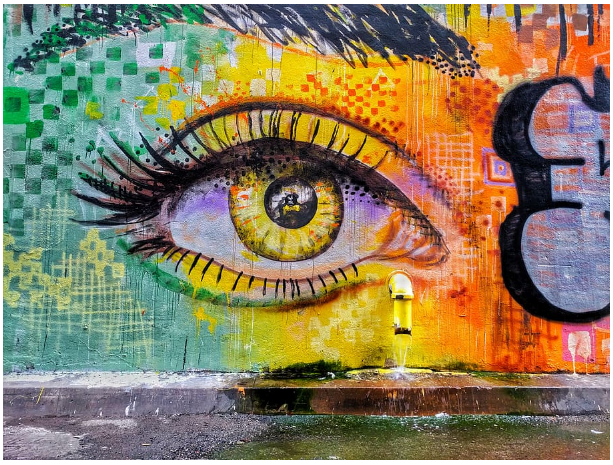


Figure 5: Colourful Test Image

Lets first take this image and look at the variance explained by each component (figure 6).

```
test = image_svd("colourful_eye.png")
test.plot_singular_rgb(10)
plt.show()
```

The scree plot tells us that approximately 80% of the variance in the original image is captured in the first component. If we remove the first component, we can see what is happening as we add more principle components (see figure 7).

Plotting various approximations of this image, we can see how much information is lost as the approximations get worse (see figure 8).

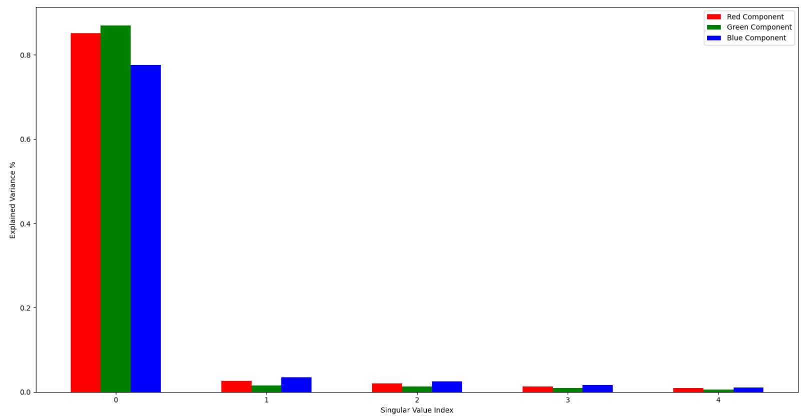


Figure 6: Scree plot for colourful image.

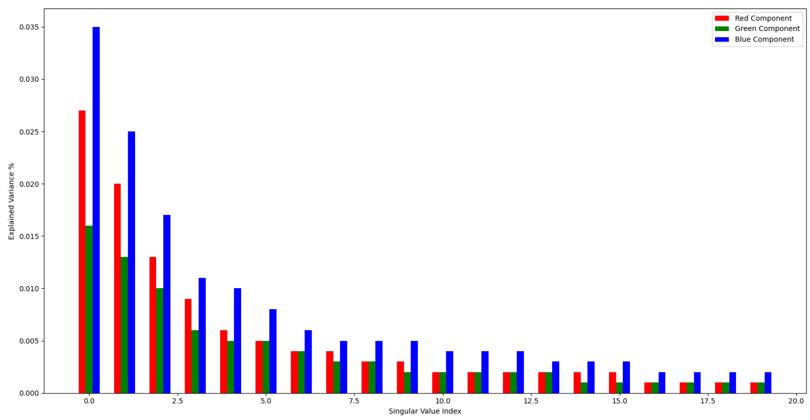


Figure 7: Scree plot for colourful image (PC1 removed).

```
test = image_svd("r_place.png")
test.plot_n_approximations_rgb(5)
plt.show()
```

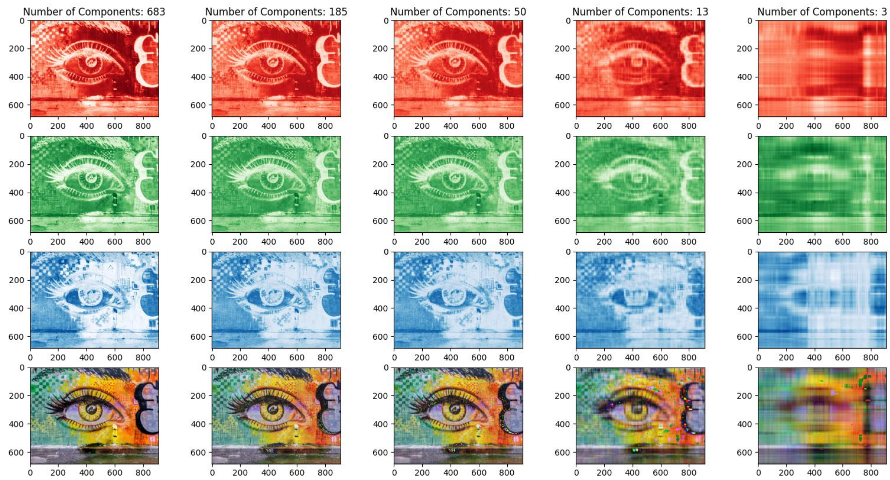


Figure 8: Various approximations of the colourful image.

12 Conclusion and Recommendations

In this dissertation we've discussed some of the main concepts in linear algebra required to explain a proof of the existence of the singular value decomposition for all real matrices. We started by discussing what a vector space is (page 4) to provide an axiomatic basis for further discussions. This led into an important concept known as a linearly independence (definition 3.3). Following this, we explored generating sets and the span (section 4) of a set of vectors, both being two sides of the same coin. This gave us a foundation for exploring a concept known as a basis (section 5). We moved onto the row rank and column rank of a matrix (section 6), providing an example to demonstrate why $rk_{\text{col}}(\boldsymbol{A}) = \dim(\text{span}[\{\boldsymbol{A}_{\cdot 1}, \ldots, \boldsymbol{A}_{\cdot m}\}])$ (page 11) for any matrix $\boldsymbol{A} \in \mathbb{R}^{n \times m}$. We then went on to our first proof that $rk_{\text{row}}(\boldsymbol{A}) = rk_{\text{col}}(\boldsymbol{A})$ for all $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ (proposition 6.1) which meant we could define a new definition, the rank of a matrix (page 12). From here, we went into the row space and column space (section 7) of a matrix and at the same time, proved that the span of a set of vectors forms a vector space itself (proposition 7.1). At this point we introduced eigenvalue eigenvector pairs (section 8) and determinants as a method to calculate them along with a proof of why this is the case. We proved why eigenvectors are linearly independent (proposition 9.2) and why the product of eigenvalues is the determinant (proposition 9.3). This was to give the determinant a less arbitrary feeling. Following this, we proved that $\boldsymbol{A}^{-1}$ exists $\iff$ $\det(\boldsymbol{A}) \neq 0$ using a proof by contradiction (proposition 9.4). From here, we went into linear transformations (section 10) to provide an intuitive way to explain the next and main topic of the dissertation, matrix decompositions. We explained the eigenvalue decomposition along with a proof of its existence for any matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ (page 30). This allows us to prove the existence of the singular value decomposition for any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ (page 32). Now that the linear algebra was laid out, we went into more practical examples of using the singular value decomposition to construct approximations of greyscale images followed by coloured images. A set of python functions was created to allow the reader to experiment with this matrix decomposition on images (page 35).

While our original goal was to build up to the singular value decomposition with a more systematized and concrete approach, it soon became apparent this wouldn't be entirely possible. You can think of mathematics as a brick wall, building up layers of reasoning that are reliant on the previous layers with axiomatic statements at the foundations. While this is somewhat true, it doesn't help with regards to how you may explain each of these concepts. There's numerous paths that you can take to get to the same proposition and if we were to include every single brick or deduction, this dissertation would be far too long so determining what to include and what to overlook was very difficult and subjective.

It is certainly true that most example dissertations provided were more focused on performing some investigation with their conclusions being the findings of this investigation. We decided to focus more on a general overview of key concepts in linear algebra. Many of the proofs provided were quite difficult to understand intuitively and finding new and interesting ways to explain certain topics was what was thought would give this approach some value.

There are numerous avenues that could be taken from this work. With regards to improving the mathematical communication. This could be by using graphs, specifically networks to illustrate the connection between certain areas. We would of also wanted to develop graphics to explain a lot of the theory with regards to basis and generating sets. Exploring how many of these theories apply when working with complex matrices may be an interesting path to take. On the topic of more practical avenues of further research, it would be interesting to find a way to quantify the difference between the original image and the constructed image to see if this correlates with the number of components used in the construction. Another example we didn't explore was the applications of the singular value decomposition to machine learning. the SVD can be used as a form of dimensionality reduction for high dimensional datasets while also capturing the maximum variance of the original dataset for a given number of dimensions. An interesting approach would be to use the MNIST dataset and investigate how the accuracy of the model is affected if dimensionality reduction is applied to the training set. We could of also tried to write code to perform a form of lossy compression on images, determining a suitable number of components required for an effective form of compression.

References

[1] Georg Cantor. “Beiträge zur Begründung der transfiniten Mengenlehre”. In: *Mathematische Annalen* 46.4 (1895), pp. 481–512.

[2] *Change of basis — Chapter 13, Essence of linear algebra — youtube.com*. `https://www.youtube.com/watch?v=P2LTAUO1TdA`. [Accessed 20-Apr-2023].

[3] *Column Space and Row Space of a Matrix — people.math.carleton.ca*. `https://people.math.carleton.ca/~kcheung/math/notes/MATH1107/wk09/09_column_space_row_space.html`. [Accessed 20-Apr-2023].

[4] *Definition of a Group — mathstats.uncg.edu*. `https://mathstats.uncg.edu/sites/pauli/112/HTML/section-61.html`. [Accessed 20-Apr-2023].

[5] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020, p. 36.

[6] *Eigenvectors and eigenvalues — Chapter 14, Essence of linear algebra — youtube.com*. `https://www.youtube.com/watch?v=PFDu9oVAE-g`. [Accessed 20-Apr-2023].

[7] Gregory Gundersen. *Proof of the Singular Value Decomposition — gregorygundersen.com*. `https://gregorygundersen.com/blog/2018/12/20/svd-proof/`. [Accessed 14-Mar-2023].

[8] *Inverse of 2x2 Matrix - Formula, Shortcut, Adjoint of 2x2 — cuemath.com*. `https://www.cuemath.com/algebra/inverse-of-2x2-matrix/`. [Accessed 20-Apr-2023].

[9] *Lecture 6 from University of Kentucky, Knoxville.* `https://web.math.utk.edu/~mengesha/teaching/Math2025/`. [Accessed 20-Apr-2023].

[10] *Lecture 9 from The Modern Algorithmic Toolbox (CS168), Spring 2023 — web.stanford.edu.* `https://web.stanford.edu/class/cs168/`. [Accessed 20-Apr-2023].

[11] *Linear independence of eigenvectors — statlect.com.* `https://www.statlect.com/matrix-algebra/linear-independence-of-eigenvectors`. [Accessed 20-Apr-2023].

[12] *Linear Transformation – from Wolfram MathWorld — mathworld.wolfram.com.* `https://mathworld.wolfram.com/LinearTransformation.html`. [Accessed 20-Apr-2023].

[13] *matplotlib — pypi.org.* `https://pypi.org/project/matplotlib/`. [Accessed 20-Apr-2023].

[14] *NumPy — numpy.org.* `https://numpy.org/`. [Accessed 20-Apr-2023].

[15] *Pillow — pillow.readthedocs.io.* `https://pillow.readthedocs.io/en/stable/`. [Accessed 20-Apr-2023].

[16] *Rank of a matrix — statlect.com.* `https://www.statlect.com/matrix-algebra/rank-of-a-matrix`. [Accessed 20-Apr-2023].

[17] Hans Schwerdtfeger. *Introduction to linear algebra and the theory of matrices.* P. Noordhoff, 1961.

[18] *Show that the determinant of A is equal to the product of its eigenvalues — math.stackexchange.com.* `https://math.stackexchange.com/questions/507641/show-that-the-determinant-of-a-is-equal-to-the-product-of-its-eigenvalues`. [Accessed 20-Apr-2023].

[19] Gilbert Strang et al. *Introduction to linear algebra*. Vol. 3. Wellesley-Cambridge Press Wellesley, MA, 1993.

[20] *Vector Basis – from Wolfram MathWorld — mathworld.wolfram.com*. `https://mathworld.wolfram.com/VectorBasis.html`. [Accessed 20-Apr-2023].

[21] *Vector Space – from Wolfram MathWorld — mathworld.wolfram.com*. `https://mathworld.wolfram.com/VectorSpace.html`. [Accessed 20-Apr-2023].