Linear Algebra and Analytic Geometry

Andrew Andreas

September 25, 2024

1 Introduction

Please report any errors to andrew.andreas0@gmail.com

2 Linear Algebra

2.1 Groups

A group $G = (\mathcal{G}, \otimes)$ is a collection of a set, $\mathcal{G}$, and an operation $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ defined by the following properties:

1. **Closure** of $\mathcal{G}$ under $\otimes : \forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
2. **Associativity**: $\forall x, y, z \in \mathcal{G} : (x \otimes y) \otimes z = x \otimes (y \otimes z)$
3. **Neutral Element:** $\forall x \in \mathcal{G}, \exists e \in \mathcal{G} : x \otimes e = x$ and $e \otimes x = x$
4. **Inverse Element:** $\forall x \in \mathcal{G}, \exists \eta \in \mathcal{G} : x \otimes \eta = e$ and $\eta \otimes x = e$

where e is the neutral element and the inverse element is defined with respect to the operation $\otimes$.

If the group is also *commutative* then the group is an **Abelian Group**. Commutativity is a property which states that $\forall x, y \in \mathcal{G}$:

$$x \otimes y = y \otimes x$$

2.2 Vector Spaces

Groups require that the operation is defined for elements within the set $\mathcal{G}$. Vector spaces extend the idea of groups by including an additional operation, an outer operation, which is scalar multiplication.

A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is comprised of a set $\mathcal{V}$ with two operations

$$+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$$
$$\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$$

where

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity:
   - $\forall \lambda \in \mathbb{R}, \mathbf{x}, \mathbf{y} \in \mathcal{V} : \lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$
   - $\forall \psi, \phi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : (\psi + \phi) \cdot \mathbf{x} = \psi\mathbf{x} + \phi\mathbf{x}$
3. Associativity w.r.t. Outer Operation: $\forall \psi, \phi \in \mathbb{R}, \mathbf{x} \in \mathcal{V} : \psi \cdot (\phi \cdot \mathbf{x}) = (\psi \cdot \phi) \cdot \mathbf{x}$
4. Neutral Element w.r.t. Outer Operation: $\forall \mathbf{x} \in \mathcal{V} : 1 \cdot \mathbf{x} = \mathbf{x}$

The elements of $\mathbf{x} \in V$ are *vectors*, the neutral element of $(\mathcal{V}, +)$ is the zero vector $\mathbf{0}$ and the inner operation is called *vector addition*.

The elements $\lambda \in \mathbb{R}$ are called scalars and the outer operation is a *multiplication by scalars*

These operations are linear operations that preserve the structure of the elements of $\mathcal{V}$. In other words, vector addition of two vectors and scalar multiplication are operations which yield another vector.

In order to prove a space is a vector space then we must first prove that the set of vectors $\mathcal{G}$ and the inner operation forms a group. Then it is necessary to prove distributivity with respect to scalar and vector sums, associativity with respect to the outer operation and the existence of a neutral element with respect to the outer operation.

2.3 Vector Subspaces

Let $V = (\mathcal{V}, +, \cdot)$ be a vector space and $\mathcal{U} \subseteq \mathcal{V}$,$\mathcal{U} \neq 0$. Then $U = (\mathcal{U}, +, \cdot)$ is a vector subspace of V if U is a vector space with the vector space operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$

Proving whether a subspace is a valid subspace we need to show the subspace is non-empty and that there is closure with respect to the inner and outer operations.

$\mathcal{U} \neq 0$ means that the set is non-empty and must at least contain the zero element $\mathbf{0}$, though this is a trivial subspace.

2.4 Linear Independence

Given a vector space V and a finite number of vectors $\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\}$ we claim that we can express any vector $\mathbf{v} \in V$ as a linear combination of the vectors $\{\mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_n}\}$

$$\mathbf{v} = \sum_{i=1}^{n} \lambda_i \mathbf{x_i} \in V$$

with $\lambda_i, i = 1, ..., n \in \mathbb{R}$

For a vector space V and the finite set of vectors $\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\}$, if we can non-trivially create a linear combination of these vectors such that

$$\sum_{i=1}^{n} \lambda_i \mathbf{x_i} = \mathbf{0}$$

where $\lambda_i \neq 0$ then the set of vectors $\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\}$ are said to be **linearly dependent**. If only the trivial solution exists such that $\lambda_1 = ... = \lambda_n = 0$ then the vectors are said to be **linearly independent**.

If we can express any single $\mathbf{x_i}$ as a linear combination of the other vectors, then the vectors are linearly dependent

A practical way to check if a set of vectors is linearly dependent is to write the vectors as columns of a matrix and use *Gaussian Elimination* to reduce the matrix to *row-echelon form*. If there are any non-pivot columns, then the set of vectors are linearly dependent and the vectors at the non-pivot positions can be expressed as a linear combination of the pivot columns to their left.

For a vector space V with k linearly independent vectors $\mathbf{b_1}, ..., \mathbf{b_k}$ and m linear combinations

$$\begin{aligned}\mathbf{x_1} &= \sum_{i=1}^{k} \lambda_{i1} \mathbf{b_i} \\ \mathbf{x_2} &= \sum_{i=1}^{k} \lambda_{i2} \mathbf{b_i} \\ &\vdots \\ \mathbf{x_m} &= \sum_{i=1}^{k} \lambda_{im} \mathbf{b_i}\end{aligned}$$

Then for any $\mathbf{x_j}$ we can write this as

$$\mathbf{x_j} = \mathbf{B}\boldsymbol{\lambda}_j, \qquad \boldsymbol{\lambda}_j = \begin{bmatrix} \lambda_{1j} \\ \lambda_{2j} \\ \vdots \\ \lambda_{kj} \end{bmatrix} \qquad j = 1, ..., m$$

If we want to test whether the vectors $\{\mathbf{x_1}, ..., \mathbf{x_m}\}$ are linearly independent, we can check the definition

$$\begin{aligned}\sum_{j=1}^{m} \phi_j \mathbf{x}_j &= \sum_{j=1}^{m} \phi_j \mathbf{B}\boldsymbol{\lambda}_j \\ &= \mathbf{B} \sum_{j=1}^{m} \phi_j \boldsymbol{\lambda}_j\end{aligned}$$

Which means that $\{\mathbf{x_1}, ..., \mathbf{x_m}\}$ are linearly dependent if and only if $\{\boldsymbol{\lambda_1}, ..., \boldsymbol{\lambda_m}\}$ are linearly dependent, which will be true if $m > k$

2.5 Basis and Rank

2.5.1 Basis

Definition. Generating Set Given a vector space $V = (\mathcal{V}, +, \cdot)$ and a set of vectors $\mathcal{A} = \{\mathbf{x_1}, ..., \mathbf{x_k}\} \subseteq \mathcal{V}$. If every vector $\mathbf{v} \in \mathcal{V}$ can be expressed as a linear combination of $\{\mathbf{x_1}, ..., \mathbf{x_k}\}$, then $\mathcal{A}$ is a generating set of V

Definition. Span The set of all linear combinations of $\mathcal{A} = \{\mathbf{x_1}, ..., \mathbf{x_k}\}$ is called the span of $\mathcal{A}$. If $\mathcal{A}$ spans V then we say $V = \text{span}[\mathcal{A}]$.

Definition. Basis A generating set $\mathcal{A}$ is called minimal if there exists no smaller set $\tilde{\mathcal{A}} \subsetneq \mathcal{A} \subseteq \mathcal{V}$ that spans V. Every linearly independent, minimal generating set of V is called a *basis*, $\mathcal{B}$, of V

Every vector space V has a basis $\mathcal{B}$ but this basis is not unique - a vector space can have different bases. All bases of a finite dimensional vector space V have the same *dimension* which is given by the number of basis vectors, $\dim(V)$. Each basis vector can be thought of as an independent direction and so $\dim(V)$ can be thought of as the number of independent directions in this vector space.

Consider a vector space V with spanning vectors $\{\mathbf{x_1}, ..., \mathbf{x_m}\}$, which is the set of all linear combination of the basis vectors $\mathcal{B} = \{\mathbf{b_1}, ..., \mathbf{b_k}\}$ of V. Since $m >> k$ we know that the span is not linearly independent and *is not* a basis of V. To determine a basis from a set of spanning vectors,

- Take the spanning vectors of V and write them as the columns of a matrix $\mathbf{A}$
- Determine the *row-echelon form* of $\mathbf{A}$
- The spanning vectors which are associated with the pivot columns are a basis of V

2.5.2 Rank

The number of linearly independent columns of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is equal to the number of linearly independent rows of $\mathbf{A}$ and is called the **rank**, $\text{rk}(\mathbf{A})$

Some properties of the rank of a matrix are given below:

- $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A^T})$
- The columns of $\mathbf{A}$ span a subspace $U \subseteq \mathbb{R}^m$ with $\dim(U) = \text{rk}(\mathbf{A})$
- The rows of $\mathbf{A}$ form a subspace $W \subseteq \mathbb{R}^n$ with $\dim(W) = \text{rk}(\mathbf{A}^T)$
- For all square matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{A}$ is invertible / regular / non-singular if $\text{rk}(\mathbf{A}) = n$
- For all non-square matrices $\mathbf{C} \in \mathbb{R}^{m \times n}$ and all $\mathbf{b} \in \mathbb{R}^m$, then the linear equation $\mathbf{Cx} = \mathbf{b}$ can only be solved if $\text{rk}(\mathbf{C}) = \text{rk}(\mathbf{C}|\mathbf{b})$ where $\mathbf{C}|\mathbf{b}$ denotes the augmented matrix. This means that $\mathbf{b}$ is in the span of $\mathbf{C}$, or equivalently is in the *column space* of $\mathbf{C}$ so that $\mathbf{b}$ can be represented as a linear combination of the columns of $\mathbf{C}$
- A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to have *full rank* if the it has maximal rank for a matrix of its dimensions, i.e., $\text{rk}(\mathbf{A}) = \min(m, n)$

2.6 Linear Mappings

Definition. Linear Mapping For vector spaces V and W, a mapping $\mathbf{\Phi}$ is called a *linear mapping (or linear transformation / vector homomorphism)* if

$$\forall\ \mathbf{x}, \mathbf{y} \in V, \quad \forall\ \phi, \psi \in \mathbb{R} : \mathbf{\Phi}(\phi\mathbf{x} + \psi\mathbf{y}) = \phi\mathbf{\Phi}(\mathbf{x}) + \psi\mathbf{\Phi}(\mathbf{y})$$

Specifically, a linear mapping $\mathbf{\Phi}$ must satisfy two sufficient conditions:

1. **Additivity**

$$\mathbf{\Phi}(\mathbf{x} + \mathbf{y}) = \mathbf{\Phi}(\mathbf{x}) + \mathbf{\Phi}(\mathbf{y})$$

2. **Homogeneity**

$$\mathbf{\Phi}(\lambda\mathbf{x}) = \lambda\mathbf{\Phi}(\mathbf{x})$$

Vector addition and scalar multiplication preserve the structure of the elements of the set $\mathcal{V}$ of V. Linear maps preserve the structure of the vector space if their application also yields vectors.

Definition. Injective / Surjective / Bijective. Consider a mapping $\Phi : \mathcal{V} \to \mathcal{W}$, where $\mathcal{V}, \mathcal{W}$ are arbitrary sets. Then Φ is called

- **Injective** $\forall\ \mathbf{x}, \mathbf{y} \in \mathcal{V} : \Phi(\mathbf{x}) = \Phi(\mathbf{y}) \Rightarrow \mathbf{x} = \mathbf{y}$
- **Surjective** $\Phi(\mathcal{V}) = \mathcal{W}$
- **Bijective** If it is both *injective* and *surjective*

A surjective mapping means that every element of $\mathcal{W}$ can be reached from $\mathcal{V}$ using the linear map Φ. An injective mapping means that Φ is a one-to-one mapping - every element of $\mathcal{V}$ is mapped to a unique element of $\mathcal{W}$. A bijective mapping has an inverse mapping Ψ such that

$$\Psi \circ \Phi(\mathbf{x}) = \mathbf{x}$$

Further cases of special linear mappings are:

- **Isomorphism** (linear and bijective) $\Phi : V \to W$
- **Endomorphism** (linear) $\Phi : V \to V$
- **Automorphism** (linear and bijective) $\Phi : V \to V$
- **Identity Automorphism** $\text{id}_V : V \to V,\ \mathbf{x} \leftrightarrow \mathbf{x}$

Finite-dimensional vector spaces V and W are *isomorphic* if and only if $\dim(V) = \dim(W)$

This has consequence that matrices of dimension $\mathbb{R}^{m\times n}$ and vectors of dimension $\mathbb{R}^{mn}$ are the same, since there exists a linear bijective mapping from one to the other.

For vector spaces V, W and X assume there exists the following linear mappings

$$\Phi : V \to W$$
$$\Psi : W \to X$$

- The mapping $\Psi \circ \Phi : V \to X$ is also linear
- If $\Phi : V \to W$ is an isomorphism, then $\Phi^{-1} : W \to V$ is also an isomorphism
- If $\Phi : V \to W, \Psi : W \to X$ are linear, then $\Psi + \Phi$ and $\lambda\Psi, \lambda \in \mathbb{R}$ are also linear

2.6.1 Matrix Representations of Linear Mappings

For a vector space $V, B = (\mathbf{b_1}, ..., \mathbf{b_n})$ is an ordered basis of V and $\forall\ \mathbf{x} \in V$ we obtain a unique linear combination

$$\begin{aligned}\mathbf{x} =& \sum_{i=1}^{n} \alpha_i \mathbf{b}_i \\ =& \mathbf{B}\boldsymbol{\alpha}\end{aligned}$$

where $\boldsymbol{\alpha}$ is the coordinate vector with respect to the basis B. Thus we see that a basis defines a coordinate system since the same vector $\mathbf{x}$ can be represented by an equally valid alternate basis, which would necessarily have an alternate coordinate vector $\boldsymbol{\upsilon}$

Definition. Transformation Matrix For vector spaces V and W with corresponding ordered basis $B = (\mathbf{b_1}, ..., \mathbf{b_n})$ and $C = (\mathbf{c_1}, ..., \mathbf{c_m})$, the linear mapping $\mathbf{\Phi} : V \to W$ applied to each $\mathbf{b}_j, j = 1, ..., n$ gives the unique coordinate representation of $\mathbf{\Phi}(\mathbf{b}_j)$ with respect to C

$$\mathbf{\Phi}(\mathbf{b}_j) = \sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i$$

The coefficients α_{ij} of each of the basis vectors $\mathbf{c}_i$ is the coordinate representation of $\mathbf{\Phi}(\mathbf{b}_j)$ with respect to the basis $C \in W$. These coordinates can be expressed as a column vector $\boldsymbol{\alpha}_j \in \mathcal{R}^m$ of the $m \times n$-matrix $\mathbf{A}_\Phi$. This is the transformation matrix for the linear mapping Φ and its elements are given by

$$\mathbf{A}_\Phi(ij) = \alpha_{ij}$$

To see this visually, consider a vector $\mathbf{x} \in V$ which can be expressed as a linear combination of the basis vectors of V, $\mathbf{x} = \sum_{j=1}^{n} k_j \mathbf{b}_j$, where k_j are the coordinate vectors of $\mathbf{x}$ with respect to the basis B. If we want to express this vector in terms of the basis $C \in W \subseteq \mathcal{R}^m$ then we can apply the transformation matrix $\mathbf{A}_\Phi$

$$\begin{aligned}\mathbf{A}_\Phi(\mathbf{x}) =& \mathbf{A}_\Phi(\sum_{j=1}^{n} k_j \mathbf{b}_j) \\ =& \sum_{j=1}^{n} k_j(\mathbf{A}_\Phi \mathbf{b}_j) \\ =& \sum_{j=1}^{n} k_j(\sum_{i=1}^{m} \alpha_{ij}\mathbf{c}_i) \\ =& \sum_{j=1}^{n}\sum_{i=1}^{m} k_j \alpha_{ij}\mathbf{c}_i\end{aligned}$$

2.6.2 Change of Basis

For a linear mapping $\Phi : V \to W$, ordered basis

$$B = (\mathbf{b_1}, ..., \mathbf{b_n}), \qquad \tilde{B} = (\tilde{\mathbf{b}}_\mathbf{1}, ..., \tilde{\mathbf{b}}_\mathbf{n})$$

of V and

$$C = (\mathbf{c_1}, ..., \mathbf{c_m}), \qquad \tilde{C} = (\tilde{\mathbf{c}}_\mathbf{1}, ..., \tilde{\mathbf{c}}_\mathbf{m})$$

of W, and transformation matrix $\mathbf{A}_\Phi \in \mathbb{R}^{m \times n}$ of Φ with respect to B and C, then the corresponding transformation matrix $\tilde{\mathbf{A}}_\Phi \in \mathbb{R}^{m \times n}$ with respect to the bases $\tilde{B}$ and $\tilde{C}$ is given as

$$\tilde{\mathbf{A}}_\Phi = \mathbf{T^{-1}} \mathbf{A_\Phi} \mathbf{S}$$

- $\mathbf{S} \in \mathbb{R}^{n \times n}$ is the id_V transformation matrix that maps the coordinates with respect to $\tilde{B}$ onto coordinates with respect to B
- $\mathbf{T} \in \mathbb{R}^{m \times m}$ is the id_W transformation matrix that maps the coordinates with respect to $\tilde{C}$ onto coordinates with respect to C

Proof The basis vectors $\tilde{\mathbf{b}}_j$ and $\tilde{\mathbf{c}}_j$ can be written as linear combinations of the basis B and C

$$\tilde{\mathbf{b}}_j = \sum_{i=1}^{n} s_{ij} \mathbf{b}_i$$

and

$$\tilde{\mathbf{c}}_k = \sum_{l=1}^{m} t_{lk} \mathbf{c}_l$$

We can also use the linear mapping Φ as applied to $\tilde{\mathbf{b}}_j$ as a linear combination of the basis $\tilde{C}$

$$\begin{aligned}
\Phi(\tilde{\mathbf{b}}_j) &= \sum_{k=1}^{m} \tilde{a}_{kj} \tilde{\mathbf{c}}_k \\
&= \sum_{k=1}^{m} \tilde{a}_{kj} \sum_{l=1}^{m} t_{lk} \mathbf{c}_l \\
&= \sum_{l=1}^{m} \Big(\sum_{k=1}^{m} t_{lk} \tilde{a}_{kj} \Big) \mathbf{c}_l
\end{aligned}$$

and equivalently we can write

$$\begin{aligned}
\Phi(\tilde{\mathbf{b}}_j) &= \Phi\Big(\sum_{i=1}^{n} s_{ij} \mathbf{b}_i \Big) \\
&= \sum_{i=1}^{n} s_{ij} \Phi(\mathbf{b}_j) \\
&= \sum_{i=1}^{n} s_{ij} \sum_{l=1}^{m} a_{li} \mathbf{c}_l \\
&= \sum_{l=1}^{m} \Big(\sum_{i=1}^{n} a_{li} s_{ij} \Big) \mathbf{c}_l
\end{aligned}$$

if we equate the two terms

$$\sum_{k=1}^{m} t_{lk} \tilde{a}_{kj} = \sum_{i=1}^{n} a_{li} s_{ij}$$

writing this in matrix form

$$\mathbf{T} \tilde{\mathbf{A}}_\mathbf{\Phi} = \mathbf{A_\Phi} \mathbf{S}$$

we see the result follows such that

$$\tilde{\mathbf{A}}_{\mathbf{\Phi}} = \mathbf{T}^{-1}\mathbf{A}_{\mathbf{\Phi}}\mathbf{S}$$

Definition. Equivalence Two matrices $\mathbf{A}, \tilde{\mathbf{A}} \in \mathbb{R}^{m\times n}$ are said to be equivalent if there exists regular matrices $\mathbf{S} \in \mathbb{R}^{n\times n}$ and $\mathbf{T} \in \mathbb{R}^{m\times m}$ such that

$$\tilde{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{S}$$

Definition. Similarity Two matrices $\mathbf{A}, \tilde{\mathbf{A}} \in \mathbb{R}^{n\times n}$ are said to be similar if there exists a regular matrix $\mathbf{S} \in \mathbb{R}^{n\times n}$ such that

$$\tilde{\mathbf{A}} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$$

Similar matrices are always equivalent though the converse is not true.

2.7 Image and Kernel

With respect to the linear map Φ where V is the domain and W is the codomain, we define the *Image* and *Kernel* of a linear map.

Definition. Image / Range / Column Space $\mathbf{C}(\mathbf{A})$

$$\text{Im}(\Phi) := \Phi(V) = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V : \Phi(\mathbf{v}) = \mathbf{w}\}$$

The *image* is the set of vectors in $\mathbf{w} \in W$ that can be reached by Φ from any vector in V.

Definition. Kernel / Null Space $\mathbf{N}(\mathbf{A})$

$$\ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{\mathbf{v} \in V : \Phi(\mathbf{v}) = \mathbf{0}_W\}$$

The *kernel* is the set of vectors $\mathbf{v}$ in the domain that Φ maps to the neutral element $\mathbf{0}_W$ (or equivalently the zero vector).

The image is a subspace of W and the kernel is a subspace of V, meaning that they are both non-empty.

Consider a linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ and its associated matrix $\mathbf{A}^{m\times n}$. The following statements are true

- $\text{Im}(\Phi) = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} = \text{span}[\mathbf{a}_1, ..., \mathbf{a}_n]$
- The image is the span of the columns / the *column space* of $\mathbf{A}$ and is therefore a subspace of $\mathbb{R}^m$
- $\text{rk}(\mathbf{A}) = \dim(\text{Im}(\mathbf{A}))$
- The kernel / null space is the general solution to the homogeneous system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{0}$ capturing all possible linear combinations of the elements of $\mathbb{R}^n$
- The kernel is a subspace of $\mathbb{R}^n$

2.7.1 The Null Space

Working with a linear map $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ and its associated matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ the domain $V \in \mathbb{R}^n$ since it takes vectors from $\mathbb{R}^n$ and maps them to the codomain $W \in \mathbb{R}^m$.

The null space is the *general solution* to $\mathbf{A}\mathbf{x} = \mathbf{0}$. This is equivalent to asking what possible linear combinations of the columns of $\mathbf{A}$ yield the zero vector in the codomain. The null space is thus in $\mathbb{R}^n$

$$\begin{aligned}\mathbf{A}\mathbf{x} = &\begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ =&x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n \\ =&\mathbf{0}\end{aligned}$$

One possible solution is the the zero vector $\mathbf{x} = \mathbf{0} \in \mathbb{R}^n$ which is true for all homogeneous system of linear equations. Hence $\mathbf{N}(\mathbf{A})$ is never empty.

Suppose however that the zero vector is not the only solution, which is the case for all non-invertible matrices. To find these vectors we can use *Gaussian elimination* to reduce the matrix to *row echelon form* or its *reduced row echelon form* to determine the dimension of the column space and hence the dimension of the null space. In these reduced forms we can easily see the solutions.

Suppose $\mathbf{A} \in \mathbb{R}^{3\times 4}$ is given by

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 2 & 8 & 10 \\ 3 & 3 & 10 & 13 \end{bmatrix}$$

Which has a row-echelon form $\mathbf{A}_r$

$$\mathbf{A}_r = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with pivot columns c_1 and c_3 meaning these vectors are linearly independent. Since they are linearly independent, no non-zero linear combination of c_1 and c_3 alone can yield the zero vector. To produce the zero vector, we must use the redundant vectors in columns c_2 and c_4 which we call *free variables*. We take these in turn to generate a **special solution** for each free variable. A **special solution** is just one specific solution to $\mathbf{Ax} = \mathbf{0}$. Note however that if $\mathbf{x}$ is a solution, then $b\mathbf{x}$ is also a solution since $\mathbf{A}(b\mathbf{x}) = b\mathbf{Ax} = b\mathbf{0} = \mathbf{0}$. So a special solution must include all points across the line of solutions given by the point $\mathbf{x} \in \mathbb{R}^n$

The complete solution is the set of all possible linear combinations of the **special solutions**. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are each special solutions, then $\mathbf{x_1} + \mathbf{x_2}$ is also a solution since

$$\mathbf{A}(\mathbf{x_1} + \mathbf{x_2}) = \mathbf{Ax_1} + \mathbf{Ax_2} = \mathbf{0} + \mathbf{0}$$

A consequence of this is that we have proved that the null space is in fact a subspace of $\mathbb{R}^4$ (generally of $\mathbb{R}^n$) since it is closed under vector addition and scalar multiplication and contains the zero vector.

To find a special solution, set all other free variables to zero and the target free variable to 1. Let $c_4 = 0$ and $c_2 = 1$ then one possible linear combination which yields the zero vector is

$$-1\underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix}}_{c_1} +1\underbrace{\begin{bmatrix}1\\0\\0\end{bmatrix}}_{c_2} +0\underbrace{\begin{bmatrix}2\\4\\0\end{bmatrix}}_{c_3} +0\underbrace{\begin{bmatrix}3\\4\\0\end{bmatrix}}_{c_4} = \mathbf{0}$$

This yields the special solution

$$x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad x_2 \in \mathbb{R}$$

For the other free variable we follow the same procedure setting $x_2 = 0$ and $x_4 = 1$. The complete solution to $\mathbf{Ax} = \mathbf{0}$ is thus all possible linear combinations which is given by

$$\mathbf{x}_n = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \quad x_2, x_4 \in \mathbb{R}$$

Through Gaussian elimination we have transformed the problem of $\mathbf{Ax} = \mathbf{0}$ by a series of elementary row operations into $\mathbf{A}_r\mathbf{x} = \mathbf{0}$. These transformations can be collected into a singular matrix $\mathbf{E}$ such that $\mathbf{EA} = \mathbf{A}_r$. These row operations leave the column space of the original matrix unchanged and thus allowing us to solve the easier $\mathbf{A}_r\mathbf{x} = \mathbf{0}$ instead

$$\mathbf{Ax} = \mathbf{0}$$
$$\mathbf{EAx} = \mathbf{E0}$$
$$\mathbf{A}_r\mathbf{x} = \mathbf{0}$$

2.7.2 Column Space

The column space of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{C}(\mathbf{A})$, refers to the vector space formed by the column vectors $\mathbf{a}_i \in \mathbb{R}^m$. $\mathbf{C}(\mathbf{A}) \subseteq \mathbb{R}^m$ contains all possible linear combinations of the column vectors and is the span[$\mathbf{A}$]

The system of linear equations $\mathbf{Ax} = \mathbf{b}$ only has a solution when $\mathbf{b}$ lies in the column space of $\mathbf{A}$. Said another way, there is some linear combination of the columns of $\mathbf{A}$ that yields $\mathbf{b}$ and the solution is the vector $\mathbf{x}$. This can be seen by the expansion below

$$\begin{aligned}\mathbf{Ax} =& \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ =& x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n \\ =& \mathbf{b} \in \mathbb{R}^m\end{aligned}$$

The solution to $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{b} \neq \mathbf{0}$ can have a unique solution, no solution or infinitely many solutions. In the case where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a square matrix and singular, we can solve for a unique $\mathbf{x}$ directly

$$\begin{aligned}\mathbf{Ax} =& \mathbf{b} \\ \mathbf{A}^{-1}\mathbf{Ax} =& \mathbf{A}^{-1}\mathbf{b} \\ \mathbf{x} =& \mathbf{A}^{-1}\mathbf{b}\end{aligned}$$

In the more general case where $\mathbf{A}$ is rectangular then the null space will be non-trivial, i.e., it contains more than just the zero vector.

To find the solution we can again use Gaussian elimination on the *augmented matrix* [$\mathbf{Ab}$] to reduce this into row echelon or reduced row echelon form. Taking an example

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix}$$

which has an augmented form and row echelon form

$$\begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 1 & 3 & 1 & 6 & \mathbf{7} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 3 & 0 & 2 & \mathbf{1} \\ 0 & 0 & 1 & 4 & \mathbf{6} \\ 0 & 0 & 0 & 0 & \mathbf{0} \end{bmatrix}$$

We immediately see that the pivot columns are c_1 and c_3.

To get a *particular solution* we set the free variables to zero and can immediately see that $x_1 = 1$ and $x_3 = 6$. Thus a particular solution is

$$\mathbf{x}_p = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix}$$

However there are more solutions than just the particular solution. Since the null space is non-trivial with special solutions

$$\mathbf{x}_n = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

The complete solution is thus $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$ since

$$\begin{aligned}\mathbf{Ax} =& \mathbf{A}\big(\mathbf{x}_p + \mathbf{x}_n\big) \\ =& \mathbf{Ax}_p + \mathbf{Ax}_n \\ =& \mathbf{b} + \mathbf{0}\end{aligned}$$

Since the dimension of the null space is 2 which is equivalent to the number of linearly independent vectors in the null space, the null space forms a plane. The complete solution is then a plane of solutions parallel to the null space

2.7.3 The Rank-Nullity Theorem

For vector spaces V, W and a linear mapping $\Phi : V \to W$ it holds that

$$\dim(\text{Im}(\Phi)) + \dim(\ker(\Phi)) = \dim(V)$$

Direct consequence of the rank-nullity theorem is that

- If $\dim(V) > \dim(\text{Im}(\Phi))$ then the kernel space is non-trivial, i.e., does not contain just the zero vector

For a rectangular matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the rank $\text{rk}(\mathbf{A}) \leq \min(m, n)$. When solving a system of linear equations $\mathbf{Ax} = \mathbf{b}$ have four scenarios:

- When $m = n = \text{rk}(\mathbf{A})$, then there is a **unique solution**
- When $n > m$ and $\text{rk}(\mathbf{A}) = m$, the matrix is short and wide (more features than observations). We will have $n - m$ linearly dependent columns. The null space will be non-empty and have dimension $\dim\big(n - m\big)$. There will be **infinitely many solutions**
- When $m > n$ and $\text{rk}(\mathbf{A}) = n$, the matrix is tall and thin with more observations than features. $\mathbf{Ax} = \mathbf{b}$ has a **unique solution** if $\mathbf{b}$ lies in the column space of $\mathbf{A}$, otherwise there is **no solution**
- When $\text{rk}(\mathbf{A}) < m$ and $n > \text{rk}(\mathbf{A})$, i.e., the matrix is not full rank. There are **no solutions or infinitely many solutions**

2.8 The Determinant

The determinant $\det(\mathbf{A})$ or $|\mathbf{A}|$ of a square matrix is a mapping to a scalar. Notation for the determinant is given below

$$\det(\mathbf{A}) = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

Theorem 1. A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$. In the 2×2 case with matrix

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ has inverse } \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

where $\det(\mathbf{A}) = ad - bc$. Hence the inverse only exists when the matrix has non zero determinant.

2.9 Computing the Determinant

There are several methods that can be used to compute the determinant of a $n \times n$ square matrix

1. Take the product of the n pivot columns
2. Using the Laplace expansion (or cofactor formula)

Theorem 2. (Laplace expansion) For a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, then for $j = 1, 2, \dots, n$:

1. *Expansion along column j*

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

2. *Expansion along row j*

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

where $\det A'_{k,j}$ is the sub-matrix of A when we isolate row k and column j

2.10 Properties of Determinants

The determinant has the following properties which can be built up from the first three properties

1. The determinant of the $n \times n$ identity matrix is 1

2. The determinant changes sign when two rows are exchanged

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = cb - da = -\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

   An even number of row exchanges results in a sign $+1$ and an odd number of row exchanges results in a -1 sign. This means that the determinant of any permutation matrix $\mathbf{P}$ can be computed by considering the number of row exchanges from the identity matrix $\mathbf{I}$

3. The determinant is a linear function of its rows or columns. If any single row or column is multiplied by a constant t then the determinant is also scaled by t

$$\begin{vmatrix} tc & td \\ a & b \end{vmatrix} = t\begin{vmatrix} c & d \\ a & b \end{vmatrix}$$

   and

$$\begin{vmatrix} tc & d \\ ta & b \end{vmatrix} = t\begin{vmatrix} c & d \\ a & b \end{vmatrix}$$

   If first rows of two matrices are added

$$\begin{vmatrix} a+a' & b+b' \\ c & c \end{vmatrix} = \begin{vmatrix} a & b \\ c & c \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & c \end{vmatrix}$$

4. If two rows of $\mathbf{A}$ are equal then $\det(\mathbf{A}) = 0$. If we consider rule 2 and conduct a row exchange then $\det(\mathbf{A})$ should change sign, $D = -\det(\mathbf{A})$ but since the matrix is unchanged $D = \det(\mathbf{A})$. For $D = -D$, the only possible scalar which will satisfy this is 0

5. Subtracting a multiple of one row from another leaves $\det(\mathbf{A})$ unchanged (row $2 - \lambda \times$ row 1)

$$\begin{aligned} \begin{vmatrix} a & b \\ c-\lambda a & d-\lambda b \end{vmatrix} =& a(d-\lambda b) - b(c-\lambda a) \\ =& ad - bc + \lambda(ab - ab) \\ =& ad - bc \end{aligned}$$

   The significance here is that the determinant is not changed in the elimination process which transforms $\mathbf{A} \to \mathbf{U}$, so $\det(\mathbf{A}) = \det(\mathbf{U})$.

6. A matrix with zero rows has $\det(\mathbf{A}) = 0$. Consider adding row 2 to row 1

$$\begin{vmatrix} 0 & 0 \\ c & d \end{vmatrix} = \begin{vmatrix} c & d \\ c & d \end{vmatrix} = 0$$

7. If $\mathbf{A}$ is triangular then $\det(\mathbf{A}) = \prod_{i=1}^{n} a_{ii}$. To see this consider a matrix in triangular form and reduce this to diagonal form

$$\begin{vmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{vmatrix} = a_{11}a_{22}...a_{nn}$$

   This is equivalent to an identity matrix $\mathbf{I}_n$ where each column has been scaled by a_{ii}. Utilizing the linearity of the determinant and rule 1 that $\det(\mathbf{I}) = 1$, we can factor out a_{11} from the first column, a_{22} from the second column and so on

8. The $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
   A proof of this follows the outline of proving that if $\det(\mathbf{A})$ and $|\mathbf{AB}|/|\mathbf{B}|$ have the same properties $(1, 2, 3)$ then they must be the same.

   Property 1 If $\mathbf{A} = \mathbf{I}$ then $|\mathbf{AB}|/|\mathbf{B}| = |\mathbf{B}|/|\mathbf{B}| = 1$

   Property 2 A row change of $\mathbf{A}$ results in a row change in $\mathbf{AB}$ so both $\det(\mathbf{A})$ and $|\mathbf{AB}||\mathbf{B}|$ both change sign

   Property 3 Multiplication of a row of $\mathbf{A}$ by a scalar scales the determinant. This scaling also scales $|\mathbf{AB}|$ by the same amount. If we add row 1 of $\mathbf{A}'$ to $\mathbf{A}$ then the determinant is a sum of determinants. If we consider adding row 1 of $\mathbf{A}'\mathbf{B}$ to $\mathbf{AB}$ then again the determinants add. When divided by $|\mathbf{B}|$, the ratios add

9. The transpose $\mathbf{A}^T$ has the same determinant as $\mathbf{A}$
   $\mathbf{A}$ can be factorized into its $\mathbf{LU}$ decomposition. Using the rule for products then $\det(\mathbf{LU}) = \det(\mathbf{L})\det(\mathbf{U})$

   $$\det(\mathbf{P})\det(\mathbf{A}) = \det(\mathbf{L})\det(\mathbf{U}) \quad \text{and} \quad \det(\mathbf{A}^T)\det(\mathbf{P}^T) = \det(\mathbf{U}^T)\det(\mathbf{L}^T)$$

   $\mathbf{L}$ and $\mathbf{L}^T$ have $1's$ across the diagonals and therefore $\det(\mathbf{L}) = \det(\mathbf{L^T}) = 1$. $\mathbf{U}$ and $\mathbf{U}^T$ have the same diagonal entries and thus have the same determinant. Since $\mathbf{P}^T\mathbf{P} = \mathbf{I}$ then $\det(\mathbf{P}^T\mathbf{P}) = \det(\mathbf{P}^T)\det(\mathbf{P}) = \det(\mathbf{I}) = 1$ meaning that $\det(\mathbf{P}^T)$ and $\det(\mathbf{P})$ must both equal 1 or -1.

   This then means that

   $$|\mathbf{A}| = |\mathbf{A}^T|$$

3 Analytic Geometry

3.1 Norms

Definition. Norm A norm on a vector space V is a function

$$\begin{aligned} ||\cdot|| : V &\to \mathbb{R} \\ \mathbf{x} &\to ||\mathbf{x}|| \end{aligned}$$

which assigns each vector its *length* $||\mathbf{x}|| \in \mathbb{R}$ such that for all $\lambda \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in V$ the following statements hold:

- *Absolutely Homogeneous* $||\lambda\mathbf{x}|| = |\lambda|||\mathbf{x}||$
- *Triangle Inequality* $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$
- *Positive Definite* $||\mathbf{x}|| \geq 0$ and $||\mathbf{x}|| = 0 \iff \mathbf{x} = \mathbf{0}$

3.2 Inner Products

Definition. Inner Product Let V be a vector space and $\Omega : V \times V \to \mathbb{R}$ be a *symmetric, positive definite bilinear mapping*, then Ω is an inner product.

Definition. Bilinear Mapping A bilinear mapping Ω is a mapping of two arguments that is linear in both arguments such that for a vector space V and vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V,\ \lambda, \phi \in \mathbb{R}$

$$\begin{aligned} \Omega(\lambda\mathbf{x} + \phi\mathbf{y}, \mathbf{z}) &= \lambda\Omega(\mathbf{xz}) + \phi\Omega(\mathbf{yz}) \\ \Omega(\mathbf{x}, \phi\mathbf{y} + \lambda\mathbf{z}) &= \phi\Omega(\mathbf{xz}) + \lambda\Omega(\mathbf{xy}) \end{aligned}$$

Definition. Symmetric Ω is called symmetric if the ordering of the arguments does not matter, such that $\Omega(\mathbf{x}, \mathbf{y}) = \Omega(\mathbf{y}, \mathbf{x})\ \forall\ \mathbf{x}, \mathbf{y} \in V$

Definition. Positive Definite Ω is called positive definite if

$$\forall\ \mathbf{x} \in V \backslash \{\mathbf{0}\} : \Omega(\mathbf{x}, \mathbf{x}) > 0, \quad \Omega(\mathbf{0}, \mathbf{0}) = \mathbf{0}$$

A vector space V and an inner product $\langle \cdot, \cdot \rangle$ define an inner product space $(V, \langle \cdot, \cdot \rangle)$

3.2.1 Symmetric, Positive Definite Matrices

Definition. Symmetric Positive Definite Matrix A symmetric matrix $\mathbf{A} \in \mathbb{R}^n$ is said to be positive definite if

$$\forall\ \mathbf{x} \in V \backslash \{\mathbf{0}\} : \mathbf{x^T A x} > 0$$

The weaker requirement of $\mathbf{x^T A x} \geq 0$ results in a *symmetric positive semi-definite matrix*

Such matrices are defined via the *inner product*. Consider an inner product space $(V, \langle \cdot, \cdot \rangle)$ with an ordered basis $B = (\mathbf{b_1}, ..., \mathbf{b_n})$ and the arbitrary vectors $\mathbf{x}, \mathbf{y} \in V$

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{y} \rangle =& \left\langle \sum_{i=1}^{n} \lambda_i \mathbf{b}_i, \sum_{j=1}^{n} \psi_j \mathbf{b}_j \right\rangle \\
=& \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \langle \mathbf{b}_i, \mathbf{b}_j \rangle \psi_j \\
=& \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i A_{ij} \psi_j \\
=& \hat{\mathbf{x}}^\mathbf{T} \mathbf{A} \hat{\mathbf{y}}
\end{aligned}$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the coordinate vectors of $\mathbf{x}$ and $\mathbf{y}$ with respect to the basis B

If a matrix $\mathbf{A} \in \mathbb{R}^n$ is symmetric and positive definite then the following properties hold:

- The null space of $\mathbf{A}$ consists of just the $\mathbf{0}$ vector since $\mathbf{x^T A x} > 0 \Rightarrow \mathbf{A x} \neq \mathbf{0}$ if $\mathbf{x} \neq 0$
- The diagonal elements of $\mathbf{A}$, A_{ii}, are positive since $A_{ii} = \mathbf{e}_i^T \mathbf{A e}_i$

3.3 Lengths and Distances

Any inner product induces a norm

$$||\mathbf{x}|| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

however the converse statement is not true.

For any inner product space $(V, \langle \cdot, \cdot \rangle)$ the induced norm $||\cdot||$ satisfies the *Cauchy-Schwarz inequality*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq ||\mathbf{x}|| ||\mathbf{y}||$$

Definition. Distance For an inner product space $(V, \langle \cdot, \cdot \rangle), d(\mathbf{x}, \mathbf{y})$ is called the distance between $\mathbf{x}$ and $\mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in V$

$$d(\mathbf{x}, \mathbf{y}) := ||\mathbf{x} - \mathbf{y}|| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$$

Definition. Metric The mapping

$$\begin{aligned}
d : V \times V &\to \mathbb{R} \\
(\mathbf{x}, \mathbf{y}) &\mapsto d(\mathbf{x}, \mathbf{y})
\end{aligned}$$

is called a *metric* and satisfies the the following

- $d(\mathbf{x}, \mathbf{x}) = 0$
- d is positive definite for any distinct points
- d is symmetric
- *Triangle inequality*: $d(\mathbf{x}, \mathbf{z}) > d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})\ \forall\ \mathbf{x}, \mathbf{y}, \mathbf{z} \in V$

3.4 Angles and Orthogonality

Inner products also define the geometry of a vector space by defining the angle ω between two vectors. If we consider the *Cauchy-Schwarz inequality* again we create two inequalities which we can combine into one

$$-1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}||||\mathbf{y}||} \text{ and } \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}||||\mathbf{y}||} \leq 1$$

so that

$$-1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}||||\mathbf{y}||} \leq 1$$

The range of this function is $[-1, 1]$ which is equivalent to the range of the cosine function so that between $[0, \pi]$ for some ω

$$\cos \omega = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{||\mathbf{x}||||\mathbf{y}||}$$

where ω is the angle between the two vectors $\mathbf{x}$ and $\mathbf{y}$ and allows us to characterize the concept of orthogonality

Definition. Orthogonality Two vectors are said to be orthogonal if and only if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

Definition. Orthogonal Matrix A square matrix $\mathbf{A} \in \mathbb{R}^n$ is orthogonal if and only if its **columns and rows** are *orthonormal* such that

$$\mathbf{A}\mathbf{A}^\mathbf{T} = \mathbf{I} = \mathbf{A}^\mathbf{T}\mathbf{A} \quad \implies \mathbf{A}^{-1} = \mathbf{A}^\mathbf{T}$$

If the columns and rows are *orthonormal* it means that they are orthogonal unit vectors such that $\{\mathbf{a}_1, ..., \mathbf{a}_n\}$

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad \text{and} \quad ||\mathbf{a}_i|| = 1$$

Orthogonal matrices provide the benefit that their linear transformations preserve the length of a vector and also preserved the angle between two vectors. Consider the euclidean norm of the vector $\mathbf{Ax}$

$$\begin{aligned} ||\mathbf{Ax}||^2 =& (\mathbf{Ax})^T(\mathbf{Ax}) \\ =& \mathbf{x}^\mathbf{T}\mathbf{A}^\mathbf{T}\mathbf{Ax} \\ =& \mathbf{x}^\mathbf{T}\mathbf{x} \\ =& ||\mathbf{x}||^2 \end{aligned}$$

and for two vectors $\mathbf{Ax}$ and $\mathbf{Ay}$

$$\begin{aligned} \cos \omega =& \frac{(\mathbf{Ax})^T(\mathbf{Ay})}{||\mathbf{Ax}||||\mathbf{Ay}||} \\ =& \frac{\mathbf{x}^T\mathbf{A}^T\mathbf{Ay}}{\sqrt{\mathbf{x}^\mathbf{T}\mathbf{A}^\mathbf{T}\mathbf{A}\mathbf{x}\mathbf{y}^\mathbf{T}\mathbf{A}^\mathbf{T}\mathbf{A}\mathbf{y}}} \\ =& \frac{\mathbf{x}^T\mathbf{y}}{||\mathbf{x}||||\mathbf{y}||} \end{aligned}$$

3.5 Orthonormal Basis

Definition. Pairwise Orthogonal If a set of vectors are pairwise orthogonal then distinct vectors are orthogonal. For a sect of vectors $\{\mathbf{e}_1, ..., \mathbf{e}_m\}$

$$\langle \mathbf{e}_i, \mathbf{e}_k \rangle = \begin{cases} 0, & i \neq k \\ x, & i = k,\ x > 0 \end{cases}$$

Definition. Orthonormal Basis For an n-dimensional vector space V with basis $B = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$, if the basis vectors are pairwise orthogonal and each vector has unit length then this is an orthonormal basis

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

If B is pairwise orthogonal and $\langle \mathbf{b}_i, \mathbf{b}_i \rangle \neq 1$, then B is an *orthogonal basis*

3.6 Orthogonal Complement

Consider a D-dimensional subspace V and an M-dimensional subspace $U \subseteq V$, then the *orthogonal complement* $U^\perp$ is a $(D - M)$-dimensional subspace of V which contains all vectors in V that are orthogonal to the vectors in U.

The intersection of U and $U^\perp$ is the zero vector: $U \cap U^\perp = \{\mathbf{0}\}$. This means that the inner product space $H = (K, \langle \cdot, \cdot \rangle)$ where $K = (\kappa, +, \cdot)$ and $\kappa = \{\mathbf{x} : \mathbf{x} \in U \cup U^\perp\}$ has the same span as $V \Rightarrow \text{span}[H] = \text{span}[V]$,

As a consequence, every vector $\mathbf{x} \in V$ can be written as a linear combination of the basis vectors of $U, B_U = (\mathbf{u}_1, ..., \mathbf{u}_M)$ and $U^\perp, B_{U^\perp} = (\boldsymbol{\upsilon}_1, ..., \boldsymbol{\upsilon}_{D-M})$

$$\mathbf{x} = \sum_{i=1}^{M} \lambda_i \mathbf{u}_i + \sum_{j=1}^{D-M} \gamma_j \boldsymbol{\upsilon}_j, \quad \lambda_i, \gamma_j \in \mathbb{R}$$

3.7 Orthogonal Projections

Definition. Projection For a vector space V and a subspace $U \subseteq V$ then a projection, π is a linear mapping $\pi : V \to U$ if

$$\pi^2 = \pi \circ \pi = \pi$$

representing this linear transformation as a matrix $\mathbf{P}_\pi$, the definition is equivalent to $\mathbf{P}_\pi^2 = \mathbf{P}_\pi$

3.7.1 Projection onto One-Dimensional Subspaces

Consider a line that passes through the origin which can be represented by the vector $\mathbf{b} \in \mathbb{R}^n$ which is a basis vector of the subspace U where $U \subseteq \mathbb{R}^n$. For some vector $\mathbf{x} \in \mathbf{R}^n$, if we want to project this onto the subspace U then we want our projection $\pi_U(\mathbf{x})$ to do so with minimal loss of information and this is achieved through an orthogonal projection.

To see this, consider that we would like to minimize the distance between $\mathbf{x}$ and the point to which it is projected $\pi_U(\mathbf{x})$ which is given by

$$||\mathbf{x} - \pi_{\mathbf{U}}(\mathbf{x})||$$

From geometry we know that the point on $\mathbf{b}$ that minimizes the length of the vector $\mathbf{x} - \pi_U(\mathbf{x})$ is the point $\lambda\mathbf{b}$ such that $\mathbf{b} \perp \mathbf{x} - \lambda\mathbf{b}$

$$\langle \mathbf{b}, \mathbf{x} - \lambda\mathbf{b} \rangle = 0 \iff \lambda = \frac{\langle \mathbf{b}, \mathbf{x} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle}$$

such that if the inner product is the dot product then

$$\lambda = \frac{\mathbf{b^T x}}{\mathbf{b^T b}}$$

Then the projection $\pi_U(\mathbf{x})$ is given by

$$\pi_U(\mathbf{x}) = \lambda\mathbf{b} = \frac{\mathbf{b^T x}}{\mathbf{b^T b}}\mathbf{b}$$

Since the projection can be represented as a matrix then we see the projection matrix $\pi_U(\mathbf{x}) = \mathbf{P}_\pi \mathbf{x}$ we immediately see that

$$\mathbf{P}\pi = \frac{\mathbf{b}\mathbf{b}^\mathbf{T}}{\mathbf{b}^\mathbf{T}\mathbf{b}}$$

The projection matrix $\mathbf{P}\pi$ projects any vector $\mathbf{x}$ onto the line through the origin $\mathbf{b}$

Note $(\mathbf{b}\mathbf{b}^T)\mathbf{x} = \mathbf{b}(\mathbf{b}^T\mathbf{x})$ since

$$\mathbf{b}\mathbf{b}^T = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \dots b_n \end{pmatrix} = \begin{pmatrix} b_1b_1 & b_1b_2 & \dots b_1b_n \\ b_2b_1 & b_2b_2 & \dots b_2b_n \\ \vdots & \vdots & \vdots \\ b_nb_1 & b_nb_2 & \dots b_nb_n \end{pmatrix}$$

Then

$$\mathbf{b}\mathbf{b}^T\mathbf{x} = \begin{pmatrix} b_1b_1 & b_1b_2 & \dots b_1b_n \\ b_2b_1 & b_2b_2 & \dots b_2b_n \\ \vdots & \vdots & \vdots \\ b_nb_1 & b_nb_2 & \dots b_nb_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1(\mathbf{b}^T\mathbf{x}) \\ b_2(\mathbf{b}^T\mathbf{x}) \\ \vdots \\ b_n(\mathbf{b}^T\mathbf{x}) \end{pmatrix} = \mathbf{b}(\mathbf{b}^T\mathbf{x})$$

3.7.2 Projection onto General Subspaces

For a subspace $U \subseteq \mathbb{R}^n$ with $\dim(U) = m, m \geq 1$ with basis $(\mathbf{b}_1, ..., \mathbf{b}_m)$ then any projection $\pi_U(\mathbf{x})$ is a linear combination of the basis vectors of U

$$\begin{aligned} \pi_U(\mathbf{x}) =& \sum_{j=1}^{m} \lambda_j \mathbf{b}_j \\ =& \mathbf{B}\boldsymbol{\lambda}, \quad \mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_m] \in \mathbb{R}^{n \times m} \end{aligned}$$

We know then that the vector $\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}$ must be pairwise orthogonal to every basis vector $(\mathbf{b}_1, ..., \mathbf{b}_m)$ such that

$$\begin{aligned} \langle \mathbf{b}_1, \mathbf{x} - \mathbf{B}\boldsymbol{\lambda} \rangle &= 0 \\ \langle \mathbf{b}_2, \mathbf{x} - \mathbf{B}\boldsymbol{\lambda} \rangle &= 0 \\ &\vdots \\ \langle \mathbf{b}_n, \mathbf{x} - \mathbf{B}\boldsymbol{\lambda} \rangle &= 0 \end{aligned}$$

where if we choose the dot product for the inner product we can rewrite this as

$$\mathbf{B}^T(\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = \mathbf{0} \iff \boldsymbol{\lambda} = (\mathbf{B}^\mathbf{T}\mathbf{B})^{-1}\mathbf{B}^\mathbf{T}\mathbf{x}$$

meaning that the projection matrix

$$\begin{aligned} \pi_U(\mathbf{x}) =& \mathbf{B}\boldsymbol{\lambda} \\ =& \mathbf{B}(\mathbf{B}^\mathbf{T}\mathbf{B})^{-1}\mathbf{B}^\mathbf{T}\mathbf{x} \\ =& (\mathbf{B}^\mathbf{T}\mathbf{B})^{-1}\mathbf{B}\mathbf{B}^\mathbf{T}\mathbf{x} \end{aligned}$$

since $\pi_U(\mathbf{x}) = \mathbf{P}\mathbf{x}$ we see that

$$\mathbf{P} = (\mathbf{B}^\mathbf{T}\mathbf{B})^{-1}\mathbf{B}\mathbf{B}^\mathbf{T}$$

Note that if the basis of $U, (\mathbf{b}_1, ..., \mathbf{b}_m)$ is an orthonormal basis, then $\mathbf{B}^T\mathbf{B} = I$ the projection matrix is

$$\mathbf{P} = \mathbf{B}\mathbf{B}^T$$

and

$$\boldsymbol{\lambda} = \mathbf{B}^T\mathbf{x}$$

3.8 Gram-Schmidt Orthogonalization

The Gram-Schmidt method allows us to construct an n-dimensional orthonormal basis, $(\boldsymbol{\upsilon}_1, ..., \boldsymbol{\upsilon}_n)$, from any n-dimensional basis $(\mathbf{b}_1, ..., \mathbf{b}_n)$ for a vector space $U \subseteq \mathbb{R}^n$

The method proceeds as follows

- Set $\boldsymbol{\upsilon}_1 = \mathbf{b}_1$
- $\boldsymbol{\upsilon}_k = \mathbf{b}_k - \pi_{\text{span}_{[\boldsymbol{\upsilon}_i, ..., \boldsymbol{\upsilon}_{k-1}]}}(\mathbf{b}_k),\ k = 1, ..., n$

where

$$\begin{aligned}\pi_{\text{span}_{[\boldsymbol{\upsilon}_i, ..., \boldsymbol{\upsilon}_{k-1}]}}(\mathbf{b}_k) =& (\mathbf{U}_{k-1}^T \mathbf{U}_{k-1})^{-1} \mathbf{U}_{k-1} \mathbf{U}_{k-1}^T \mathbf{b}_k \\ =& \mathbf{U}_{k-1} \mathbf{U}_{k-1}^T \mathbf{b}_k, \quad \mathbf{U}_{k-1} = \begin{bmatrix} \boldsymbol{\upsilon}_1 & \boldsymbol{\upsilon}_2 & \cdots & \boldsymbol{\upsilon}_{k-1} \end{bmatrix} \in \mathbb{R}^{n \times (k-1)}\end{aligned}$$

where since the $k-1$ basis vectors are orthonormal, then $(\mathbf{U}_{k-1}^T \mathbf{U}_{k-1})^{-1} = I$

4 Characteristics of a Matrix

There are two separate values that can summarize a matrix, the determinant, which we discussed earlier and its trace, whose properties we will briefly review

4.1 The Trace

The Trace of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is defined as

$$\text{tr}(\mathbf{A}) := \sum_{i=1}^{n} a_{ii}$$

The trace of a square matrix $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$, has the following properties

- $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
- $\text{tr}(\alpha \mathbf{A}) = \alpha \text{tr}(\mathbf{A}), \alpha \in \mathbb{R}$
- $\text{tr}(\mathbf{I_n}) = n$
- $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$

4.2 Similar Matrices

Two matrices $\mathbf{A} \in \mathbf{R}^{n \times n}$ and $\mathbf{B} \in \mathbf{R}^{n \times n}$ are similar if there exists an invertible matrix $\mathbf{P} \in \mathbf{R}^{n \times n}$ such that

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

The matrices $\mathbf{A}$ and $\mathbf{P}^{-1}\mathbf{BP}$ represent the same linear transformation but with respect to different basis. If $\mathbf{A}$ represents a linear transformation with respect to the canonical basis $C \in \mathbf{R}^n$, then the $\mathbf{P}^{-1}\mathbf{BP}$ represents the same transformation with respect to a new basis $\tilde{C} \in \mathbf{R}^n$

1. $\mathbf{P}$ results in a change of basis from the standard canonical basis C to $\tilde{C}$
2. The matrix $\mathbf{B}$, which operates in the same vector space as $\mathbf{A}$, makes the same transformation but with respect to the basis $\tilde{C}$
3. $\mathbf{P}^{-1}$ then performs a change of basis back to the canonical basis C from $\tilde{C}$

4.3 Diagonal Matrices

Diagonal matrices are matrices whose non-diagonal elements are zero and are thus of the form

$$\begin{pmatrix} k_1 & 0 & \dots & 0 \\ 0 & k_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & k_n \end{pmatrix}$$

Diagonal matrices result in a scaling of the column vectors of of another matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix} = \begin{bmatrix} k_1 a_{11} & k_2 a_{12} \\ k_1 a_{12} & k_2 a_{22} \end{bmatrix} = (k_1 \mathbf{a}_1 \mid k_2 \mathbf{a}_2)$$

A matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is said to be **diagonalizable** if it is similar to a diagonal matrix.

$$\mathbf{A} = \mathbf{P}^{-1} \mathbf{\Lambda} \mathbf{P}$$

where $\mathbf{\Lambda}$ is a diagonal matrix

5 Eigenvalues and Eigenvectors

Eigenvalue Equation For a square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$, then $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ and $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}$ is the corresponding eigenvector if

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

The geometric idea here is that an eigenvector is a special vector that under the transformation by $\mathbf{A}$ points in the same direction.

λ is an eigenvalue of $\mathbf{A} \in \mathbb{R}^{n \times n}$ if and only if it is the root of the characteristic polynomial, $P_A(\lambda)$, of $\mathbf{A}$. The characteristic polynomial, $P_A(\lambda)$, of a matrix $\mathbf{A}$ is defined below

$$P_A(\lambda) := \det(\mathbf{A} - \lambda \mathbf{I})$$

where $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ since the matrix is not full rank because there exists a non-trivial solution (i.e., $\mathbf{x} \neq 0$) to the problem

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$$

This means that the eigenvector, $\mathbf{x}$, which is a solution to the above equation for a given λ is in the null space of $\mathbf{A} - \lambda \mathbf{I}$ since it is the solution to $(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$

Eigenspace For a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, the Eigenspace is the set of all eigenvectors associated with a specific eigenvalue, λ and it spans a subspace of $\mathbb{R}^n$. The eigenspace is denoted E_λ.

5.1 Properties

1. $\mathbf{A}$ and $\mathbf{A}^T$ have the same eigenvalues since $\det(\mathbf{A}) = \det(\mathbf{A}^T)$, though not necessarily the same eigenvectors. To see this, let $\mathbf{B} = (\mathbf{A} - \lambda \mathbf{I})$ and note that the transpose is a linear operator, then

$$\begin{aligned} \det\left(\mathbf{B}^T\right) &= \det\left((\mathbf{A} - \lambda \mathbf{I})^T\right) \\ &= \det\left(\mathbf{A}^T - \lambda \mathbf{I}^T\right) \\ &= \det\left(\mathbf{A}^T - \lambda \mathbf{I}\right) \\ &= \det(\mathbf{B}) \end{aligned}$$

   which means they have the same roots of the characteristic polynomial and therefore the same eigenvalues

2. The eigenvectors associated with distinct eigenvalues are *linearly independent*. Suppose that $\mathbf{x}_1$ and $\mathbf{x}_2$ are eigenvectors of the matrix $\mathbf{A}$ with eigenvalues λ_1 and λ_2. If the vectors are linearly independent then the only solution to this equation is trivial, i.e., $c_1 = c_2 = 0$

$$c_1\mathbf{x_1} + c_2\mathbf{x_2} = \mathbf{0}$$

We generate two equations, the first by left multiplying by $\mathbf{A}$ and the second by multiplying by λ_2 resulting in

$$c_1\lambda_1\mathbf{x_1} + c_2\lambda_2\mathbf{x_2} = \mathbf{0} \tag{1}$$
$$c_1\lambda_2\mathbf{x_1} + c_2\lambda_2\mathbf{x_2} = \mathbf{0} \tag{2}$$

If we take (1) − (2) then

$$c_1\mathbf{x}_1\big(\lambda_1 - \lambda_2\big) \leftrightarrow c_1 = 0$$

This time multiplying by λ_1

$$c_1\lambda_1\mathbf{x_1} + c_2\lambda_1\mathbf{x_2} = \mathbf{0} \tag{3}$$

If we take (1) − (3) then

$$c_2\mathbf{x}_2\big(\lambda_2 - \lambda_1\big) \Leftrightarrow c_2 = 0$$

Since $c_1 = c_2 = 0$ then we have that $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent.

3. Similar matrices possess the same eigenvalues - this means that a linear mapping Φ has eigenvalues that are independent of choice of basis

   Consider the matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times n}$ which are similar and thus $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P} \iff \mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$. If λ is an eigenvalue of $\mathbf{A}$

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

   Substituting in for $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$

$$\begin{aligned}\mathbf{P}^{-1}\mathbf{B}\mathbf{P}\mathbf{v} &= \lambda\mathbf{v}\\ \mathbf{B}\mathbf{P}\mathbf{v} &= \lambda\mathbf{P}\mathbf{v}\end{aligned}$$

   We see that λ is also an eigenvalue of $\mathbf{B}$ with eigenvector $\mathbf{P}\mathbf{v}$. The eigenvectors $\mathbf{P}\mathbf{v}$ and $\mathbf{v}$ are the same vector but under different coordinate representations

4. The product of the n eigenvalues equals the determinant

5. The sum of the n eigenvalues equals the *trace*

The eigenvectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ of a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with n distinct eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ are linearly independent and thus form a basis of $\mathbb{R}^n$.

Algebraic Multiplicity is the number of times a given eigenvalue appears as the root of the characteristic polynomial

Geometric Multiplicity is the number of linearly independent eigenvectors in the eigenspace of the eigenvalue λ_i

An eigenvalue always has at least one associated eigenvector but if the algebraic multiplicity > 1, then it may have more. Note that the geometric multiplicity $\leq$ algebraic multiplicity.

A matrix is not diagonalizable (not similar to a diagonal matrix) if it has less than n linearly independent eigenvectors (and hence cannot form $\mathbf{P}$). This occurs when the there is a repeated root in the characteristic polynomial but the geometric multiplicity $<$ algebraic multiplicity. This means that there is less than n linearly independent eigenvectors and hence this matrix is not diagonalizable.

5.2 Spectral Theorem

Every symmetric matrix $\mathbf{S}$ has the factorization $\mathbf{Q\Lambda Q}^T$ with real eigenvalues in $\mathbf{\Lambda}$ and orthonormal eigenvectors in the columns of $\mathbf{Q}$. The eigenvectors are *orthogonal* and can be *normalized* to be *orthonormal*. Since all eigenvectors are orthogonal (and hence linearly independent) then symmetric matrices are diagonalizable

$$\mathbf{S} = \mathbf{P\Lambda P}^{-1}$$

and

$$\mathbf{S}^T = \left(\mathbf{P}^{-1}\right)^T \mathbf{\Lambda P}^T$$

Since $\mathbf{S} = \mathbf{S^T}$ then we have that $\mathbf{P\Lambda P}^{-1} = \left(\mathbf{P}^{-1}\right)^T \mathbf{\Lambda P}^T \iff \mathbf{P}^T = \mathbf{P}^{-1}$

To denote the special nature of these orthogonal matrices we give them a unique symbol $\mathbf{Q}$. So a symmetric matrix can be diagonalized to

$$\mathbf{S} = \mathbf{Q\Lambda Q}^T \quad \text{where } \mathbf{Q}^{-1} = \mathbf{Q}^T$$

A proof of these two statements are given in turn

1. **Claim:** Symmetric matrices have real eigenvalues

   Suppose that

   $$\mathbf{Sx} = \lambda\mathbf{x}$$

   and at this point we are in the most general case where the eigenvalues λ can be complex and the eigenvectors $\mathbf{x}$ can have complex components.

   The complex conjugate of λ and $\mathbf{x}$ is $\overline{\lambda}$ and $\overline{\mathbf{x}}$, respectively. The complex conjugate of $\mathbf{Sx}$ is

   $$\mathbf{S}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

   If we take the dot product of $\overline{\mathbf{x}}$ with $\mathbf{Sx}$ and then the dot product of $\mathbf{S}\overline{\mathbf{x}}$ with $\mathbf{x}$ we get

   $$\overline{\mathbf{x}}^T\mathbf{Sx} = \overline{\mathbf{x}}^T\lambda\mathbf{x} \quad \text{and} \quad \overline{\mathbf{x}}^T\mathbf{Sx} = \overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x}$$

   We notice that we now have two expressions for $\overline{\mathbf{x}}^T\mathbf{Sx}$ and so we can write

   $$\overline{\mathbf{x}}^T\lambda\mathbf{x} = \overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x} \iff \lambda = \overline{\lambda} \iff a + bi = a - bi \iff b = 0$$

   which proves that all eigenvalues are real

2. **Claim:** Real symmetric matrices have eigenvectors which are perpendicular

   Suppose that $\mathbf{Sx} = \lambda_1\mathbf{x}$ and $\mathbf{Sy} = \lambda_2\mathbf{y}$ and that $\lambda_1 \neq \lambda_2$

   If we take the dot product of $\mathbf{y}$ with $\mathbf{Sx}$ and the dot product of $\mathbf{x}$ with $\mathbf{Sy}$ we get

   $$\left(\mathbf{Sx}\right)^T\mathbf{y} = \mathbf{x}^T\mathbf{Sy} = \lambda_1\mathbf{x}^T\mathbf{y}$$

   and

   $$\mathbf{x}^T\left(\mathbf{Sy}\right) = \mathbf{x}^T\mathbf{Sy} = \lambda_2\mathbf{x}^T\mathbf{y}$$

   We notice the two expressions on the LHS are equivalent and therefore that both of the RHS must also be equivalent

   $$\lambda_1\mathbf{x}^T\mathbf{y} = \lambda_2\mathbf{x}^T\mathbf{y}$$

   Since $\lambda_1 \neq \lambda_2 \iff \mathbf{x}^T\mathbf{y} = 0$. This proves that any two eigenvectors of $\mathbf{S}$ with distinct eigenvalues must be orthogonal

5.3 Positive Definite Matrices

Positive Definite A square symmetric matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ is said to be positive definite if all its eigenvalues $\lambda_i > 0, i = 1, ..., n$.

If we consider the eigenvalue equation $\mathbf{Sx} = \lambda\mathbf{x}$ with $\lambda > 0$ and take its inner product with $\mathbf{x}$ then this forms the definition of a symmetric positive definite matrix. $\mathbf{S}$ is positive definite if

$$\mathbf{x}^T\mathbf{Sx} > 0, \forall \mathbf{x} \neq \mathbf{0}$$

To see this, use the definition of $\mathbf{Sx}$

$$\mathbf{x}^T\mathbf{Sx} = \lambda\mathbf{x}^T\mathbf{x}$$

where $\mathbf{x}^T\mathbf{x} = ||\mathbf{x}||^2 > 0$ for any $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}$. Since $\lambda > 0$, the positiveness of the quadratic form relies on the eigenvalues being real and positive. Necessary conditions for positive definiteness are

- A matrix $\mathbf{S}$ is full rank so that $\nexists \mathbf{x}, \mathbf{x} \neq 0 \mid \mathbf{Sx} = \mathbf{0}$. If the matrix was not full rank there would exist an $\mathbf{x} \neq \mathbf{0} \mid \mathbf{Sx} = \mathbf{0} \iff \mathbf{x}^T\mathbf{Sx} = 0$
- The matrix is symmetric meaning that all its eigenvalues are real and that it can be diagonalized with an orthonormal basis

Any vector $\mathbf{x} \in \mathbb{R}^n$ can be expressed as a linear combination of the eigenvectors, since they form an orthogonal basis of $\mathbb{R}^n$. To see this fact, use the diagonal form of $\mathbf{S} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ in the quadratic form

$$\mathbf{x}^T\mathbf{Sx} = \mathbf{x}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{x} = \mathbf{y}^T\mathbf{Sy}$$

Then $\mathbf{x}^T\mathbf{Q} = \mathbf{y}$ is a linear combination of the row space of $\mathbf{Q}$ (which is also an orthonormal basis). This expresses the coordinates of the vector $\mathbf{x}$ in the basis defined by the row space of $\mathbf{Q}$. If we write this as

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n$$

where c_i denotes the coordinates of $\mathbf{x}$ in the basis of the row space of $\mathbf{Q}^T$. Then if we substitute this into

$$\begin{aligned}
\mathbf{y}^T\mathbf{Sy} &= \big(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n\big)^T\mathbf{S}\big(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n\big) \\
&= \sum_{i=1}^{n} c_i^2\mathbf{v}_i^T\mathbf{S}\mathbf{v}_i + \sum_{i \neq j} c_ic_j\mathbf{v}_i^T\mathbf{S}\mathbf{v}_j \\
&= \sum_{i=1}^{n} \lambda_ic_i^2\mathbf{v}_i^T\mathbf{v}_i + \sum_{i \neq j} \lambda_jc_ic_j \underbrace{\mathbf{v}_i^T\mathbf{v}_j}_{=0} \\
&= \sum_{i=1}^{n} \lambda_ic_i^2\mathbf{v}_i^T\mathbf{v}_i
\end{aligned}$$

where the last term shows that since $\lambda_i > 0$, $||\mathbf{v}_i||^2 > 0$ and $c_i^2 \geq 0, i = 1, ..., n$ with at least one $c_i \neq 0$ then we have positive definiteness for any $\mathbf{x}$.

5.3.1 Sylvester's Criterion

Sylvester's criterion is a necessary and sufficient condition for positive definiteness. It states that a $n \times n$ symmetric matrix $\mathbf{S}$ is positive definite if and only if all the principle minors are positive. A principle minor is the determinant of the $k \times k,\ k \leq n$ submatrices of $\mathbf{S}$.

The reverse of the argument, that positive definite matrices have positive principle minors $\mathbf{S}_k$, is easy to see. If $\mathbf{S}$ is positive definite then $\mathbf{x}^T\mathbf{Sx} > 0,\ \forall \mathbf{x} \neq \mathbf{0}$, which includes all $\mathbf{x}$ of the form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad k < n$$

Since $\mathbf{x}^T\mathbf{S}\mathbf{x} = \mathbf{x}_k^T\mathbf{S}_k\mathbf{x}_k > 0 \iff \det\mathbf{S}_k > 0, \forall k$.

To prove this claim in the forward direction we can use induction and remember that the determinant is the product of the eigenvalues. Suppose that Sylvester's criterion holds for $\mathbf{S}_n$ meaning that $\det\mathbf{S}_i > 0, i = 1, ..., n$. We need to show then that $\mathbf{S}_{n+1}$ is positive definite if and only if the principle minor of $\mathbf{S}_{n+1}$ is positive.

The general form for an $(n+1) \times (n+1)$ symmetric matrix is

$$\mathbf{S}_{n+1} = \begin{pmatrix} \mathbf{S}_n & \mathbf{v} \\ \mathbf{v}^T & d \end{pmatrix}$$

where $\mathbf{S}_n$ is a symmetric positive definite matrix by induction, $\mathbf{v}$ is a vector and $d > 0$. $\det\mathbf{S}_{n+1}$ is the leading principle minor of size $(n+1) \times (n+1)$ and is equal to the product of its pivots

$$\det\mathbf{S}_{n+1} = \det\mathbf{S}_n \cdot \left(d - \mathbf{v}^T\mathbf{S}_n^{-1}\mathbf{v}\right)$$

Since we know $\det\mathbf{S}_n > 0$ by the induction hypothesis, then for $\det\mathbf{S}_{n+1} > 0$ we must have $d - \mathbf{v}^T\mathbf{S}_n^{-1}\mathbf{v} > 0$. This ensures that $\mathbf{S}_{n+1}$ is positive definite and that all the leading principle minors of $\mathbf{S}_{n+1}$ are positive, satisfying Sylvester's criterion.

5.4 Positive Semidefinite Matrices

Positive semidefiniteness is a weaker inequality that requires all eigenvalues of $\mathbf{S}$ are non-negative. Formally it states that for any $\mathbf{x}$ then

$$\mathbf{x}^T\mathbf{S}\mathbf{x} \geq 0$$

therefore all positive definite matrices are positive semidefinite. Necessary conditions for a matrix to be positive semidefinite are as follows

- All eigenvalues are non-negative $\lambda_i \geq 0, \forall i = 1, ..., n$
- The energy of the system $\mathbf{x}^T\mathbf{S}\mathbf{x} \geq 0$
- Sylvester's criterion yields leading principal minors $\det\mathbf{S}_{k\times k} \geq 0, k = 1, ..., n$ that are all non-negative

Given a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ we can construct a symmetric, positive semidefinite matrix $\mathbf{S} = \mathbf{A}^T\mathbf{A} \in \mathbb{R}^{m\times m}$

$$\mathbf{S} = \begin{pmatrix} — & \mathbf{a}_1^T & — \\ — & \mathbf{a}_2^T & — \\ — & \vdots & — \\ — & \mathbf{a}_m^T & — \end{pmatrix} \begin{pmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_m \\ | & | & & | \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1^T\mathbf{a}_1 & \mathbf{a}_1^T\mathbf{a}_2 & ... & \mathbf{a}_1^T\mathbf{a}_m \\ \mathbf{a}_2^T\mathbf{a}_1 & \mathbf{a}_2^T\mathbf{a}_2 & ... & \mathbf{a}_2^T\mathbf{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_m^T\mathbf{a}_1 & ... & ... & \mathbf{a}_m^T\mathbf{a}_m \end{pmatrix}$$

where we can also refer to $\mathbf{S}$ as the correlation matrix as it computes a dot product between the component vectors of the matrix, $\mathbf{A}_i,\ i = 1, .., m$. We can see that $\mathbf{S}$ is symmetric since the dot product is commutative, i.e., the entry $\mathbf{S}_{ij} = \mathbf{a}_i^T\mathbf{a}_j = \mathbf{a}_j^T\mathbf{a}_i = \mathbf{S}_{ji}$. However, the usual definition follows

$$\begin{aligned} \mathbf{S} &= \mathbf{A}^T\mathbf{A} \\ &= (\mathbf{A}\mathbf{A}^T)^T \\ &= (\mathbf{A})^T(\mathbf{A}^T)^T \\ &= \mathbf{A}^T\mathbf{A} \\ &= \mathbf{S}^T \end{aligned}$$

5.5 Test for Local Extrema

The hessian matrix is a symmetric matrix that contains the second order partial derivatives of a scalar-valued function. Since it is symmetric it has real eigenvalues.

Extending the idea of the second derivative test to the multivariate case, the second order partial derivatives characterise the rate of change of the first order partial derivative in each direction. For a local minima we require that in every direction we traverse, the function values increase around this point.

Using the 2×2 hessian in the xy plane as an example

$$\mathbf{H} = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

So we require that the leading principle minors $\mathbf{H}_1 = f_{xx} > 0$ and $\mathbf{H}_2 = f_{xx}f_{yy} - f_{xy}^2 > 0$. If we move along the x axis then the function increases - that is satisfied by the first principle minor. The second term tests movement anywhere in the xy plane and asserts that a necessary condition for this to be a minimum is that any movement along this plane must lead to an increase in function value, which is what we would expect. Hence a necessary and sufficient condition for a local minima is that the Hessian $\mathbf{H}$ be positive definite at the critical point.

Let $U \subseteq \mathbb{R}^n$ be open and $f : U \to \mathbb{R}$ a function of class C^2. Suppose that $\mathbf{a} \in U$ is a critical point of f

- If the Hessian $\mathbf{H}_f(\mathbf{a})$ is positive definite, then f has a **local minimum** at $\mathbf{a}$
- If the Hessian $\mathbf{H}_f(\mathbf{a})$ is negative definite, then f has a **local maximum** at $\mathbf{a}$
- If the determinant of the Hessian $\det \mathbf{H}_f(\mathbf{a}) \neq 0$ and $\mathbf{H}_f(\mathbf{a})$ is neither positive or negative definite, then $\mathbf{a}$ is a **saddle point**

5.6 Eigendecomposition and Diagonalization

For a matrix $\mathbf{A}$ that is *similar* to a diagonal matrix $\mathbf{D}$ we have

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

If we left multiply $\mathbf{D}$ by $\mathbf{P}$ and assume that the elements of $\mathbf{D}$ are scalars $\lambda_1, \lambda_2, ..., \lambda_n$ and $\mathbf{P} = (\mathbf{p}_1 \mid \mathbf{p}_2 \mid \ldots \mid \mathbf{p}_n)$, then

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$$
$$(\mathbf{A}\mathbf{p}_1 \mid \mathbf{A}\mathbf{p}_2 \mid \ldots \mid \mathbf{A}\mathbf{p}_n) = (\lambda_1\mathbf{p}_1 \mid \lambda_1\mathbf{p}_2 \mid \ldots \mid \lambda_n\mathbf{p}_n)$$

which can be rewritten into the familiar form to show that the vectors, $\mathbf{p}_i, i = 1, .., n$ are the eigenvectors that correspond to the eigenvalues $\lambda_i, i = 1, ..., n$ of the matrix $\mathbf{A}$

$$\mathbf{A}\mathbf{p}_1 = \lambda_1\mathbf{p}_1$$
$$\vdots$$
$$\mathbf{A}\mathbf{p}_n = \lambda_n\mathbf{p}_n$$

Theorem 3. (Eigendecomposition) A square matrix, $\mathbf{A} \in \mathbb{R}^{n \times n}$ can be decomposed into

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

if and only if the eigenvectors of $\mathbf{A}$ form a basis of $\mathbb{R}^{n \times n}$

Theorem 4. Symmetric matrices always have a diagonal form which, following on from above, means that the eigenvectors of a symmetric matrix always form a basis of $\mathbb{R}^{n \times n}$

This is a direct consequence of the **Spectral Theorem** which goes further and states that an orthonormal basis exists for the eigenvectors of $\mathbf{A}$. As a result of the properties of an orthonormal matrix, i.e., $\mathbf{P}^T = \mathbf{P}^{-1}$, the eigendecomposition can be written as

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

which provides the computational benefit of not having to compute the inverse matrix of $\mathbf{P}$.

5.7 The Singular Value Decomposition

The singular value decomposition, **SVD**, is a matrix decomposition that exists for all matrices, not just square matrices. For a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ then its **SVD** is

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where $\mathbf{\Sigma} \in \mathbb{R}^{n\times m}$ with $\Sigma_{ii} = \sigma_i \geq 0$ and $\Sigma_{ij} = 0, i \neq j$ and both $\mathbf{U} \subseteq \mathbb{R}^{n\times n}$ and $\mathbf{V} \subseteq \mathbb{R}^{m\times m}$ are orthonormal matrices.

The columns of $\mathbf{U}$, denoted $\mathbf{u}_i, i = 1, ..., n$ are called the left singular vectors and the columns of $\mathbf{V}$, denoted $\mathbf{v}_i, i = 1, ..., m$ are called the right singular vectors.

For a rectangular matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, recall that this matrix represents a linear mapping, Φ from a vector space $V \in \mathbb{R}^m$ to the vector space $W \in \mathbb{R}^n$

$$\Phi : V \longrightarrow W$$

Each of these vector spaces have their own standard bases, B and C. This means that the $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ decomposition performs the same linear mapping as $\mathbf{A}$ but does so sequentially.

1. The matrix $\mathbf{V}$ performs a basis change from the canonical basis, $B \subseteq \mathbb{R}^m$ to another basis $\tilde{B} \subseteq \mathbb{R}^m$. The basic vectors are the eigenvectors of the matrix $\mathbf{A}\mathbf{A}^T$
2. Having performed the change of basis to $\tilde{B}$, the matrix $\mathbf{\Sigma}$ performs two actions
   - Adds ($n > m$) or deletes dimensions ($n < m$)
   - Scales the coordinate vectors by the singular values, σ_i
   - This is the transformation matrix of the linear mapping $\Phi : V \longrightarrow W$ with basis $\tilde{B} \longrightarrow \tilde{C}$
3. The matrix $\mathbf{U}$ performs a basis change from the basis $\tilde{C} \longrightarrow C$ in the codomain $\mathbb{R}^n$

To compute the component matrices of the singular value decomposition, we will start with the right singular matrix, $\mathbf{V}^T$. We can construct a symmetric matrix from any matrix $\mathbf{A}$ by computing the dot product with itself $\mathbf{A}^T\mathbf{A}$. Since this matrix is symmetric, it is diagonalizable, has real eigenvalues and its eigenvectors are orthogonal and can form an orthonormal basis

$$\begin{aligned}\mathbf{A}^T\mathbf{A} =& \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ =& \mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ =& \mathbf{V}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{V}^T\end{aligned}$$

where since $\mathbf{U}$ is an orthonormal matrix, then $\mathbf{U}^{-1} = \mathbf{U}^T$, so that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$

If we compare this result to that of the eigendecomposition of $\mathbf{A}^T\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$, then we can directly see that

$$\begin{aligned}\mathbf{V}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{V}^T =& \mathbf{P}\mathbf{D}\mathbf{P}^T \\ \mathbf{V} =& \mathbf{P} \\ \mathbf{\Sigma}^T\mathbf{\Sigma} =& \mathbf{D} \\ \sigma_i^2 =& \lambda_i\end{aligned}$$

which tells us that the right singular vectors are the eigenvectors of $\mathbf{A}^T\mathbf{A}$ and that the singular values of $\mathbf{A}, \sigma_i$ are the square root of the eigenvalues of $\mathbf{A}^T\mathbf{A}, \lambda_i$

Turning our attention to the left singular vectors, $\mathbf{U}$ we follow a similar procedure in that we form another symmetric matrix, $\mathbf{A}\mathbf{A}^T$ yielding

$$\begin{aligned}\mathbf{A}\mathbf{A}^T =& \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\right)^T \\ =& \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T \\ =& \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T\end{aligned}$$

where if we again compare this to the eigendecomposition of $\mathbf{A}\mathbf{A}^T = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^T$, we can see that

$$\begin{aligned}\mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^T &= \mathbf{S}\mathbf{\Lambda}\mathbf{S}^T \\ \mathbf{U} &= \mathbf{S} \\ \mathbf{\Sigma}\mathbf{\Sigma}^T &= \mathbf{\Lambda} \\ \sigma_i^2 &= \lambda_i\end{aligned}$$

So the left singular vectors $\mathbf{U}$ are the eigenvectors of the matrix $\mathbf{A}\mathbf{A}^T$ and the eigenvalues of $\mathbf{A}\mathbf{A}^T$ are the square of the singular values of $\mathbf{A}$

Note that since the singular values of $\mathbf{A}$ are equal to the square root of the eigenvalues of both $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$, then it is true that the eigenvalues of these matrices must be the same.

5.7.1 Computing the left singular vectors U

An alternative way to compute the left singular vectors $\mathbf{u}_i$, other than via the characteristic polynomial of $\mathbf{A}\mathbf{A}^T$ is to consider the eigenvalue equation for the matrix $\mathbf{A}^T\mathbf{A}$ and its eigenvector $\mathbf{v}_i$

$$\begin{aligned}\mathbf{A}^T\mathbf{A}\mathbf{v}_i &= \lambda_i\mathbf{v}_i \\ \mathbf{A}\mathbf{A}^T\big(\mathbf{A}\mathbf{v}_i\big) &= \lambda_i\big(\mathbf{A}\mathbf{v}_i\big)\end{aligned}$$

which is the eigenvalue equation for $\mathbf{A}\mathbf{A}^T$ and it tells us that $\mathbf{A}\mathbf{v}_i$, the image of the right singular vector under $\mathbf{A}$, is the i^{th} eigenvector. So to construct the orthonormal basis required for the left singular vector $\mathbf{U}$, then we need to normalize these vectors

$$\mathbf{u}_i = \frac{\mathbf{A}\mathbf{v}_i}{||\mathbf{A}\mathbf{v}_i||} = \frac{\mathbf{A}\mathbf{v}_i}{\sqrt{\mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v}_i}} = \frac{\mathbf{A}\mathbf{v}_i}{\sqrt{\lambda_i}\sqrt{\mathbf{v}_i^T\mathbf{v}_i}} = \frac{1}{\sqrt{\lambda_i}}\mathbf{A}\mathbf{v}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$$

5.7.2 Intuition of the SVD

For a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ where $n \gg m$, the SVD is given by

$$\begin{aligned}\mathbf{A} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \\ &= \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \end{bmatrix} \begin{bmatrix} — & \mathbf{v}_1^T & — \\ — & \mathbf{v}_2^T & — \\ — & \vdots & — \end{bmatrix} \\ &= \begin{bmatrix} | & | & & | \\ \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \dots & 0 \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & \mathbf{v}_1^T & — \\ — & \mathbf{v}_2^T & — \\ — & \vdots & — \end{bmatrix} \\ &= \sigma_1\mathbf{u_1}\mathbf{v_1}^T + \sigma_2\mathbf{u_2}\mathbf{v_2}^T + ... + \sigma_m\mathbf{u_m}\mathbf{v_m}^T\end{aligned}$$

where despite the fact we have n left singular vectors, because $\mathbf{A}$ is a $\mathbf{R}^{n\times m}$ matrix where $n \gg m$, this means that it can be at most rank m and hence can have at most m distinct singular values.

So we have a linear combination of the vectors of $\mathbf{U}$ and $\mathbf{V}$ which are scaled by the singular values and they can be interpreted as a sum of rank 1 matrices that increasingly improve the estimation of $\mathbf{A}$. We can see this because if we consider the dimensions of $\mathbf{u}_i$ and $\mathbf{v}_i^T$, this is an *outer product* between an $n \times 1$ column vector and a $1 \times n$ row vector which yields a $n \times n$ matrix. The reason this matrix is of rank 1 is because the entire matrix is entirely dependent on a single row and column vector.