Linear Algebra and Analytic Geometry

Andrew Andreas

August 1, 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 $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ where the following:

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 of scalar multiplication.

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

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

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 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.

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 subspace 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 argue 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 this indicates that the set of vectors are linearly dependent and that 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 basis. All basis 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, W, a mapping is called a *linear mapping (or linear transformation / vector homomorphism)* if

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

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

Definition. Injective / Surjective / Bijective. 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** $\mathrm{id}_V : V \to V,\ \mathbf{x} \leftrightarrow \mathbf{x}$

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 $\mathbb{R}^{mn}$ the same, since there exists a linear bijective mapping from one to the other.

For vector spaces V, W, X

- The linear mappings

  $$\begin{aligned}\Phi :& V \to W\\ \Psi :& W \to X\end{aligned}$$

  the mapping $\Phi \circ \Psi : V \to X$ is also linear
- If $\Phi : V \to W$ is an isomorphism, then $\Phi^{-1} : W \to V$ is also an isomorphism
- If $\Phi : V \to W, \Psi : V \to W$ 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 $\boldsymbol{v}$

Definition. Transformation Matrix For a vector space V, W with corresponding ordered basis $B = (\mathbf{b_1}, ..., \mathbf{b_n}), C = (\mathbf{c_1}, ..., \mathbf{c_m})$, the linear mapping $\mathbf{\Phi} : V \to W$ applied to each $\mathbf{b}_j, j = 1, ..., n$ gives the unique 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 $m \times n$-matrix $\mathbf{A}_\Phi$ whose elements are given by

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

is the transformation matrix for the linear mapping Φ.

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

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

- $\mathbf{S} \in \mathbb{R}^{n\times n}$ 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

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

we see the result follows such that

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

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

Definition. Kernel / Null Space

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

Definition. Image / Range

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

With respect to the linear map Φ V is the domain and W is the codomain.

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

The image is a subspace of W and the null space are vector 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

- $\mathrm{Im}(\Phi) = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} = \mathrm{span}[\mathbf{a}_1, ..., \mathbf{a}_n]$
- The image is the span of the columns / the *column space* of $\mathbf{A}$ and is therefore a subspace of $\mathbb{R}^m$
- $\mathrm{rk}(\mathbf{A}) = \dim(\mathrm{Im}(\mathbf{A}))$
- 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 $\mathbb{R}^n$

Definition. The Rank-Nullity Theorem For vector spaces V, W and a linear mapping $\Phi : V \to W$ it holds that

$$\dim(\mathrm{Im}(\Phi)) + \dim(\mathrm{rk}(\Phi)) = \dim(V)$$

Direct consequence of the rank-nullity theorem is that

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

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 $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V, \lambda, \phi \in \mathbb{R}$

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

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

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

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

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 is positive definite
- 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 allow 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

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

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

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

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

$$\mathbf{a}_i^T \mathbf{a}_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

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 $\mathbf{Ax}$

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

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

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

3.5 Orthonormal Basis

Definition. Orthonormal Basis For an n-dimensional vector space V and basis $B = \{\mathbf{b}_1, ..., \mathbf{b}_n\}$ of V, then if

$$
\begin{aligned}
\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0, \text{ for } i \neq j \\
\langle \mathbf{b}_i, \mathbf{b}_i \rangle = 1
\end{aligned}
$$

for all $i, j = 1, ..., n$, then B is an orthonormal basis. If the B is only pairwise orthogonal and $\langle \mathbf{b}_i, \mathbf{b}_i \rangle \neq 1$, then the B is an *orthogonal basis*

3.6 Orthogonal Complement

Consider an 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 $H = (K, \langle\cdot,\cdot\rangle)$ has the same span as $V \Rightarrow \text{span}[H] = \text{span}[V]$, where $K = (\kappa, +, \cdot)$ is the vector space and $\kappa = \{\mathbf{x} : \mathbf{x} \in U \cap U^\perp\}$.

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 $U^\perp, B_{U^\perp} = (\boldsymbol{v}_1, ..., \boldsymbol{v}_{D-M})$

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

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

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

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

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

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

meaning that the projection matrix

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

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

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

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

where

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

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