

Linear Algebra

Haydn Cheng

Contents

1	Vectors and Vector Spaces	3
1.1	Vectors in \mathbb{R}^n	3
1.2	Vector Spaces	5
1.3	Span and Linear Independence	6
1.4	Basis and Dimension	7
2	Vectors in \mathbb{R}^n and its Geometrical Applications	11
2.1	Scalar Product in \mathbb{R}^n	11
2.2	Vector Product in \mathbb{R}^3	12
3	Linear Maps and Matrices	13
3.1	Maps and Sets	13
3.2	Linear Maps	14
3.3	Coordinates Maps $f : F^n \rightarrow V$	16
3.4	Matrices	16
3.4.1	Linear Maps $f : F^n \rightarrow F^m$	16
3.4.2	Basic Properties	18
3.4.3	Rank of a Matrix	19
3.4.4	Matrix Multiplication	20
3.4.5	Inverse of a Matrix	21
3.4.6	Gaussian Elimination	21
3.5	Linear Maps $f : V \rightarrow W$	23
3.6	Change of Basis	24
4	Determinants	26
4.1	Definition of a Determinant	26
4.2	Properties of a Determinant	28
4.3	Cramer's Rule	31
5	Scalar Products	32
5.1	Real and Hermitian Scalar Products	32
5.2	Orthonormal Basis, Gram-Schmidt Procedure	33

5.3	Adjoint Linear Maps	38
5.4	Unitary Maps	40
6	Eigenvectors and Eigenvalues	47
6.1	Diagonalization of matrices	48
7	Rotational Transformations	52
	Appendices	54
A	Rigorous Proofs	55
A.1	A Sub Vector Space is a Vector Space	55
A.2	Relation between Inverse and Bijective Maps	55
A.3	Inverse of a Composite Map	56
A.4	Properties of a Linear Map (1)	56
A.5	Dimensions Relation of Domain, Kernel and Image	57
A.6	Properties of a Linear Map (2)	58
A.7	Row Rank is Equals to Column Rank	58
A.8	Properties of an Inverse of a Matrix	59
A.9	Rank of a Matrix in Upper Echelon Form	60
A.10	Matrix Rank is Equals to Linear Map Rank	60
B	Extra Materials	61
B.1	Fields	61
B.2	Matrices for Elementary Row Operations	61

Vectors and Vector Spaces

1.1 Vectors in \mathbb{R}^n

Generally, a field F (such as the set of all real numbers \mathbb{R} or complex numbers \mathbb{C} ¹) is an arena within which “regular calculations” involving addition and multiplication of numbers can be carried out.

By F^n we denote the set of all ordered n -tuples of elements from field F

$$F^n = \{(a_1, a_2, \dots, a_n) \mid a_i \in F, i = 1, 2, \dots, n\}. \quad (1.1)$$

As we will see, since F^n satisfies the definition of a vector space, so every n -tuple are already vectors themselves. However, we would usually arrange the elements of the n -tuples into the entries of a n -dimensional column vector as $\mathbf{v} = (a_1 \ a_2 \ \cdots \ a_n)^T$ for easier manipulation.

It has a canonical (standard) basis, called the standard unit vectors, which are the n vectors in F^n where all the entries are zero except for the i^{th} element:

$$\mathbf{e}_i = (0, \dots, 0, \underbrace{1}_{i^{\text{th}} \text{ position}}, 0, \dots, 0) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \leftarrow i^{\text{th}} \text{ position}. \quad (1.2)$$

They are useful because a general vector in \mathbb{R}^n can then be written as a linear combination of the unit vectors

$$\mathbf{v} = v_1 \mathbf{e}_1 + \dots + v_n \mathbf{e}_n = \sum_{i=1}^n v_i \mathbf{e}_i. \quad (1.3)$$

¹More examples of fields are given in section B.1.

It is important to note that the concept of “standard unit vectors” is only applicable to \mathbb{F}^n . However, in later sections, we would introduce another concept (orthogonality) which extends this concept to any vector space.

Example: Google’s search algorithm

Question: Consider an internet with n websites labeled by $k = 1, \dots, n$. The site labeled as k has been referenced to by other sites $L_k \subset 1, \dots, n$. Assign a page rank to each website representing the relative significance of each website.

Solution: A first attempt would be to define the page rank as the number of pages that reference this page, so

$$x_k = n_k. \quad (1.4)$$

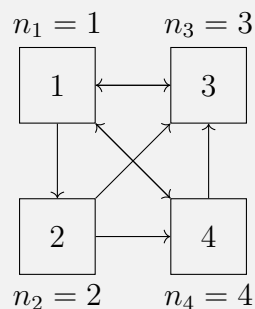
An improved version would be to take into account the page rank of the pages that reference this page, as it is more desirable to be referenced by a page with higher page rank. So

$$x_k = \sum_{j \in L_k} x_j \quad (1.5)$$

would be an improvement to the formula. Furthermore, as a reference is more valuable if a page only references a small number of pages, we may define the page rank as

$$x_k = \sum_{j \in L_k} \frac{x_j}{n_j}. \quad (1.6)$$

A simple example is illustrated below:



As shown in the figure, $n = 4$, $n_1 = 3$, $n_2 = 2$, $n_3 = 1$, $n_4 = 2$, $L_1 = 3, 4$, $L_2 = 1, 3, 4$ and $L_4 = 1, 2$. So the formulas for assigning the page rank becomes

$$\begin{aligned} x_1 &= \frac{x_3}{1} + \frac{x_4}{2} \\ x_2 &= \frac{x_1}{3} \\ x_3 &= \frac{x_1}{3} + \frac{x_2}{2} + \frac{x_4}{2} \\ x_4 &= \frac{x_1}{3} + \frac{x_2}{2}. \end{aligned} \quad (1.7)$$

This system of linear equations can be written in vector and matrix notation, as

$$A\mathbf{x} = \mathbf{x}, \quad (1.8)$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \text{ and } A = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix}. \quad (1.9)$$

This equation describes a so-called “eigenvalue problem”, which we will discuss in a later stage.

1.2 Vector Spaces

We start by defining a vector space and a sub vector space:

Definition 1.2.1 (Definition of a Vector Space). *A vector space V over a field F is a set with two operations:*

1. *vector addition: $(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} + \mathbf{w} \in V$, where $\mathbf{v}, \mathbf{w} \in V$.*
2. *scalar multiplication: $(\alpha, \mathbf{v}) \mapsto \alpha\mathbf{v} \in V$, where $\alpha \in F$ and $\mathbf{v} \in V$.*

And for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $\alpha, \beta \in F$, these operations have to satisfy the following rules:

- (A1) “Associativity”: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
- (A2) “Neutral element”: There exists a “zero vector”, $\mathbf{0} \in V$ so that $\mathbf{0} + \mathbf{v} = \mathbf{v}$.
- (A3) “Inverse element”: There exists an inverse, $-\mathbf{v}$ with $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
- (A4) “Commutativity”: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.
- (A5) $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$.
- (A6) $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$.
- (A7) $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$.
- (A8) $1 \cdot \mathbf{v} = \mathbf{v}$.

The elements $\mathbf{v} \in V$ are called “vectors”, the elements $\alpha \in F$ of the field are called “scalars”.

Definition 1.2.2 (Definition of a Sub Vector Space). *A sub vector space $W \subset V$ is a non-empty subset of a vector space V which is closed under vector addition and scalar multiplication, i.e. ²*

- (B1) $\mathbf{w}_1 + \mathbf{w}_2 \in W \quad \forall \mathbf{w}_1, \mathbf{w}_2 \in W$.
- (B2) $\alpha\mathbf{w} \in W \quad \forall \alpha \in \mathbf{F} \text{ and } \mathbf{w} \in W$.

²The rigorous proof of a sub vector space is a vector space itself is provided in ??.

1.3 Span and Linear Independence

Definition 1.3.1 (Definition of Span). *The span of k vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ in a vector space V over a field F is the set of all linear combinations of all those vectors*

$$\text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) \equiv \left\{ \sum_{i=1}^k \alpha_i \mathbf{v}_i \mid \alpha_i \in F \right\}. \quad (1.10)$$

Definition 1.3.2 (Definition of Linear Independence). *Let V be a vector space over F and $\alpha_1, \dots, \alpha_k \in F$ scalars. A set of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ is called linearly independent if*

$$\sum_{i=1}^k \alpha_i \mathbf{v}_i = \mathbf{0} \implies \alpha_i = 0 \quad \forall i. \quad (1.11)$$

Otherwise, the vectors are called linearly dependent.

Corollary 1.3.0.1. *The vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly dependent \iff One vector \mathbf{v}_i can be written as a linear combination of the others.*

Proof. Assume the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly dependent so that the equation $\sum_{i=1}^k \alpha_i \mathbf{v}_i = \mathbf{0}$ has a solution with at least one $\alpha_i \neq 0$, then we can solve for \mathbf{v}_i to get

$$\mathbf{v}_i = -\frac{1}{\alpha_i} \sum_{j \neq i} \alpha_j \mathbf{v}_j, \quad (1.12)$$

where we expressed \mathbf{v}_i as a linear combination of the other vectors. □

Example: Linear Independence of Vectors in \mathbb{R}^3

Question: Verify the following vectors in \mathbb{R}^3 are linearly independent

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \quad (1.13)$$

Solution: Setting the linear combination of vectors to be the zero vector,

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 = \begin{pmatrix} \alpha_3 \\ \alpha_1 + \alpha_2 + \alpha_3 \\ \alpha_1 + 2\alpha_2 - \alpha_3 \end{pmatrix} = \mathbf{0} \quad (1.14)$$

we find that it only has the solution $(\alpha_1, \alpha_2, \alpha_3) = (0, 0, 0)$. Therefore the vectors are linearly independent.

Example: Linear Independence of Solutions to Differential Equations

Question: Prove that the solutions $y(x) = \cos x$ and $y(x) = \sin x$ to the homogeneous, linear second order differential equation

$$\frac{d^2y}{dx^2} = -y \quad (1.15)$$

are linearly independent.

Solution: We start with setting the linear combination of the solutions to 0,

$$\alpha \sin x + \beta \cos x = 0, \quad (1.16)$$

and since the equation has to be satisfied for all x , setting $x = 0$ we learn that $\beta = 0$ and setting $x = \frac{\pi}{2}$ it follows that $\alpha = 0$. Hence $\sin x$ and $\cos x$ are linearly independent.

1.4 Basis and Dimension

Definition 1.4.1 (Definition of a Basis). A set $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ of vectors is called a basis of V if and only if

(C1) $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent.

(C2) $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$

Lemma 1.4.1. If $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis of V , every vector $\mathbf{v} \in V$ can be written as a unique linear combination as

$$\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{v}_i. \quad (1.17)$$

Proof. Firstly we must be able to write \mathbf{v} as a linear combination of the basis, since $\mathbf{v} \in V$ and $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$.

Now assume that \mathbf{v} can be written as two different linear combinations of the basis, i.e.

$$\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{v}_i = \sum_{i=1}^n \beta_i \mathbf{v}_i. \quad (1.18)$$

Taking the difference of these two equations implies

$$\sum_{i=1}^n (\alpha_i - \beta_i) \mathbf{v}_i = 0 \quad (1.19)$$

and from linear independence of the basis, it follows that all $\alpha_i - \beta_i = 0$, so that indeed $\alpha_i = \beta_i$. \square

We would like to call the number of vectors in a basis the dimension of the vector space. First, however, we have to prove that the number of vectors in a basis of a vector space is an invariant.

Lemma 1.4.2 (Exchange Lemma). *Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be a basis of V and $\mathbf{w}_1, \dots, \mathbf{w}_m \in V$ arbitrary vectors. If $m > n$ then $\mathbf{w}_1, \dots, \mathbf{w}_m$ are linearly dependent.*

Proof. Consider the first arbitrary vector \mathbf{w}_1 . If $\mathbf{w}_1 = 0$ then the vectors $\mathbf{w}_1, \dots, \mathbf{w}_m$ are linearly dependent (since the coefficient of \mathbf{w}_1 can be any arbitrary constant when we write down the linear combination of the vectors $\mathbf{w}_1, \dots, \mathbf{w}_m$ to test their linear dependency), so we can assume that $\mathbf{w}_1 \neq 0$.

Since the vectors \mathbf{v}_i form a basis we can write the first arbitrary vector as

$$\mathbf{w}_1 = \sum_{i=1}^n \alpha_i \mathbf{v}_i \quad (1.20)$$

with at least one α_i (say α_1) non-zero (or else \mathbf{w}_1 would be zero). We can therefore solve this equation for \mathbf{v}_1 as

$$\mathbf{v}_1 = \frac{1}{\alpha_1} \left(\mathbf{w}_1 - \sum_{i=2}^n \alpha_i \mathbf{v}_i \right). \quad (1.21)$$

This shows that we use \mathbf{w}_1 to replace \mathbf{v}_1 in basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ such that $V = \text{Span}(\mathbf{w}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$, such that the span is remained unchanged. This is because any vector $\mathbf{u} \in V$ which can be written as a linear combination of the basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ can now be written as a linear combination of the basis $\mathbf{w}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ simply by utilizing eq. (1.21).

This exchange process can be repeated until all \mathbf{v}_i are placed by \mathbf{w}_i and $V = \text{Span}(\mathbf{w}_1, \dots, \mathbf{w}_n)$. Since $m > n$ there is at least one vector \mathbf{w}_{n+1} “left over” which can be written as a linear combination

$$\mathbf{w}_{n+1} = \sum_{i=1}^n \beta_i \mathbf{w}_i \quad (1.22)$$

which shows that the vectors $\mathbf{w}_1, \dots, \mathbf{w}_m$ are linearly dependent. □

Definition 1.4.2 (Definition of Dimension). *For a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of a vector space V over F we call $\dim_F(V) \equiv n$ the dimension of V over F .*

Example: Standard Unit Vectors form a Basis

Question: Prove that the standard unit vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ form a basis over the vector space \mathbb{R}^n and \mathbb{C}^n .

Solution: Set the linear combination of the standard unit vectors to zero,

$$\sum_{i=1}^n \alpha_i \mathbf{e}_i = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = 0 \quad (1.23)$$

which is only true if $\alpha_i = 0 \forall i$.

Thus the standard unit vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ form a basis of \mathbb{R}^n and \mathbb{C}^n seen as vector spaces of the fields \mathbb{R} and \mathbb{C} , respectively, where the dimensions are

$$\dim_{\mathbb{R}}(\mathbb{R}^n) = \dim_{\mathbb{C}}(\mathbb{C}^n) = n. \quad (1.24)$$

However, note that \mathbb{C}^n as a vector space over \mathbb{R} has a basis $\mathbf{e}_1, \dots, \mathbf{e}_n, i\mathbf{e}_1, \dots, i\mathbf{e}_n$ and therefore $\dim_{\mathbb{R}}(\mathbb{C}^n) = 2n$.

Example: Basis of a Polynomial

Question: Prove that the monomials $1, x, x^2, \dots, x^d$ form a basis over a vector space containing all real polynomials of degree d .

Solution: We start with setting

$$\sum_{i=0}^d \alpha_i x^i = 0 \quad (1.25)$$

to check if there is any non-trivial solution. However, by taking the k^{th} derivative with respect to x and then set $x = 0$ (since it has to satisfy for all x), this immediately implies that $\alpha_k = 0 \forall k$ and hence the monomials are linearly independent and form a basis and the vector space has dimension $d + 1$.

Example: Basis of a $n \times m$ matrix

Question: Find a basis and the dimension of the vector space containing all $n \times m$ matrices with real entries.

Solution: The analogy of standard unit vectors for matrices is

$$E_{(ij)} = \begin{pmatrix} 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}^a \quad (1.26)$$

where $i = 1, \dots, n$ and $j = 1, \dots, m$ and the “1” appears in the i^{th} row and the j^{th} column with all other entries zero. Clearly these matrices form a basis, in complete analogy with the standard unit vectors. Therefore the vector space of

$n \times m$ matrices has dimension nm .

^aHere the subscript (i, j) has been used to indicate the row and column which have been changed rather than specific entries of the matrix.

Lemma 1.4.3 (Properties of a Vector Space). *For a vector space V spanned by a finite number of vectors we have:*

- (D1) V has a basis.
- (D2) Every linearly independent set $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ that is not already a basis of V can be completed by some other vectors in V to form a basis.
- (D3) If $n = \dim(V)$, any linearly independent set of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ forms a basis.
- (D4) If $\dim_F(V) = \dim_F(W)$ and $V \subset W$ for two vector spaces V, W then $V = W$.

Proof. (D1) Since $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$, If these vectors are linearly independent we have found a basis. If not, one of the vectors, say v_k can be written as a linear combination of the others and can, hence, be dropped without changing the span, so $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_{k-1})$. This process can be continued until the remaining set of vectors is linearly independent.

(D2) If $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ we are finished. If not there exists a vector $\mathbf{v}_{k+1} \notin \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$. Hence the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}$ must be linearly independent. We can continue adding vectors to the list until it spans the whole space. This process must terminate after a finite number of steps or else we would contradict the exchange lemma 1.4.2, which states that the number of vectors in a basis is an invariant.

(D3) If $\dim(V) = n$ and the linearly independent set $\mathbf{v}_1, \dots, \mathbf{v}_n$ did not span V then from (D2) it could be completed to a basis with more than n elements. However this is a contradiction since according to the exchange lemma 1.4.2 the number of vectors in a basis is an invariant. So the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ must span the space and they form a basis.

(D4) We can choose a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of V . Since $V \subset W$, these basis vectors are linearly independent in W , and since $\dim_F(W) = \dim_F(V)$, they must also form a basis of W , using (D3). Hence $V = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_n) = W$.

□

magic
square

Vectors in \mathbb{R}^n and its Geometrical Applications

2.1 Scalar Product in \mathbb{R}^n

The scalar (dot) product for two n -dimensional column vectors is defined as

$$\mathbf{a} \cdot \mathbf{b} \equiv \sum_{i=1}^n a_i b_i. \quad (2.1)$$

In Einstein summation convention we simply write $\mathbf{a} \cdot \mathbf{b} = a_i b_i$, assuming we sum over any index which appears twice in a given term.

The dot product satisfies a number of obvious properties, namely

1. $\mathbf{a} \cdot \mathbf{b} = a_i b_i = b_i a_i = \mathbf{b} \cdot \mathbf{a}$
2. $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = a_i (b_i + c_i) = a_i b_i + a_i c_i = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$
3. $\mathbf{a} \cdot (\beta \mathbf{b}) = a_i (\beta b_i) = \beta a_i b_i = \beta \mathbf{a} \cdot \mathbf{b}$
4. $\mathbf{a} \cdot \mathbf{a} = \sum_{i=1}^n a_i^2 > 0$ for $\mathbf{a} \neq \mathbf{0}$

The last property allows us to define the length of a vector as

$$|\mathbf{a}| \equiv \sqrt{\mathbf{a} \cdot \mathbf{a}} = \left(\sum_{i=1}^n a_i^2 \right)^{\frac{1}{2}}. \quad (2.2)$$

The dot product satisfies an important inequality

Lemma 2.1.1 (Cauchy-Schwarz inequality). *For any two vectors \mathbf{a} and \mathbf{b} in \mathbb{R}^n we have*

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}| |\mathbf{b}|. \quad (2.3)$$

Proof. Without loss of generality let $|\mathbf{a}| = |\mathbf{b}| = 1$ (if this is not the case then we can always normalize the vectors). Then

$$0 \leq |\mathbf{a} \pm \mathbf{b}|^2 = (\mathbf{a} \pm \mathbf{b}) \cdot (\mathbf{a} \pm \mathbf{b}) = |\mathbf{a}|^2 + 2(\mathbf{a} \cdot \mathbf{b}) + |\mathbf{b}|^2 = 2(1 \pm \mathbf{a} \cdot \mathbf{b}) \implies |\mathbf{a} \cdot \mathbf{b}| \leq 1. \quad (2.4)$$

□

A closely related inequality is the famous

Lemma 2.1.2 (Triangle Inequality). *For any two vectors \mathbf{a} and \mathbf{b} in \mathbb{R}^n we have*

$$|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}| \quad (2.5)$$

Proof. $|\mathbf{a} + \mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 + 2\mathbf{a} \cdot \mathbf{b} \leq |\mathbf{a}|^2 + |\mathbf{b}|^2 + 2|\mathbf{a}||\mathbf{b}| = (|\mathbf{a}| + |\mathbf{b}|)^2.$ □

For two non-zero vectors \mathbf{a} and \mathbf{b} , the Cauchy-Schwarz inequality implies that we can define an unique angle $\theta \in [0, \pi]$ as

$$-1 \leq \cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} \leq 1 \implies \mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \theta. \quad (2.6)$$

We call two vectors orthogonal (or perpendicular) if and only if $\mathbf{a} \cdot \mathbf{b} = 0$, or $\theta = \frac{\pi}{2}$.

2.2 Vector Product in \mathbb{R}^3

Linear Maps and Matrices

3.1 Maps and Sets

Definition 3.1.1 (Definition of a Map). *A map between two sets X and Y assigns to each $x \in X$ a $y \in Y$ which is written as $y = f(x)$ and referred to as the image of x under f . In symbols,*

$$f : X \rightarrow Y, \quad x \mapsto f(x). \quad (3.1)$$

The set X is called the domain of the map f , Y is called the co-domain of f . The set

$$\text{Im}(f) = \{f(x) \mid x \in X\} \subseteq Y \quad (3.2)$$

is called the image of f and consists of all elements of the co-domain which can be obtained as images under f .

Definition 3.1.2 (Terminology for Maps). *Let $f : X \rightarrow Y$ be a map between two sets X and Y . The map f is called*

1. *Injective: if every element of the co-domain is the image of at most one element of the domain. Mathematically,*

$$f(x) = f(\tilde{x}) \implies x = \tilde{x} \quad \forall x, \tilde{x} \in X. \quad (3.3)$$

2. *Surjective: if every element of the co-domain is the image of at least one element of the domain. Mathematically,*

$$\text{Im}(f) = Y \quad (3.4)$$

3. *Bijective: if it is injective and surjective, i.e. if every element of the co-domain is the image of precisely one element of the domain.*

Definition 3.1.3 (Definition of a Composite Map). *For two maps $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ the composite map $g \circ f : X \rightarrow Z$ is defined by*

$$(g \circ f)(x) \equiv g(f(x)) \quad (3.5)$$

From this definition it is easy to show that map composition is associative

$$h \circ (g \circ f) = (h \circ g) \circ f \quad (3.6)$$

since $(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))) = (h \circ g)(f(x)) = ((h \circ g) \circ f)(x)$.

Definition 3.1.4 (Definition of an Identity Map). *A identity map $\text{id}_X : X \rightarrow X$ maps every element in X onto itself, i.e.*

$$\text{id}_X(x) = x \quad \forall x \in X. \quad (3.7)$$

Definition 3.1.5 (Definition of an Inverse Map). *Given a map $f : X \rightarrow Y$, a map $g : Y \rightarrow X$ is called an inverse of f if*

$$(g \circ f) = \text{id}_X \text{ and } (f \circ g) = \text{id}_Y. \quad (3.8)$$

Theorem 3.1.1 (Relation between Inverse and Bijective Maps). *The map $f : X \rightarrow Y$ has an inverse if and only if f is bijective. If the inverse exists it is unique and denoted by $f^{-1} : Y \rightarrow X$.*

Proof. If the map is not surjective, an inverse map could not exist since some y are not in the image of f . If the map is not injective, an inverse map could not exist as well since some $y \in Y$ are the images of more than one element in the domain¹. \square

Theorem 3.1.2 (Inverse of a Composite Map). *If the maps $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ are both bijective, then the composite map $f \circ g : Y \rightarrow Z$ is also bijective and hence by theorem 3.1.1 has an inverse given by².*

$$(f \circ g)^{-1} = g^{-1} \circ f^{-1} \quad (3.9)$$

3.2 Linear Maps

Definition 3.2.1 (Definition of a Linear Map). *A map $f : V \rightarrow W$ between two vector spaces V and W over a field F is called linear if*

$$(E1) \quad f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2)$$

$$(E2) \quad f(\alpha \mathbf{v}) = \alpha f(\mathbf{v})$$

for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$ and for all $\alpha \in F$.

¹The rigorous proof of this theorem can be found in section A.2.

²The theorem is intuitive and trivial but the rigorous proof is given in section A.3.

Definition 3.2.2 (Definition of Kernel). *For a map $f : V \rightarrow W$, the subset of V which consists of all vectors $\mathbf{v} \in V$ mapped to the zero vector, is defined as the kernel of the map $\ker(f)$. Mathematically,*

$$\ker(f) = \{\mathbf{v} \in V \mid f(\mathbf{v}) = \mathbf{0}\} \subset V. \quad (3.10)$$

Lemma 3.2.1 (Properties of a Linear Map (1)). *A linear map $f : V \rightarrow W$ between two vector spaces V and W over F has the following properties:³*

(F1) *The zero vectors are mapped onto each other, so $f(\mathbf{0}) = \mathbf{0}$. Hence $\mathbf{0} \in \ker(f)$.*

(F2) *The kernel of f is a sub vector space of V .*

(F3) *The image of f is a sub vector space of W .*

(F4) *f is surjective $\iff \text{Im}(f) = W \iff \dim \text{Im}(f) = \dim(W)$*

(F5) *f is injective $\iff \ker(f) = \{\mathbf{0}\} \iff \dim \ker(f) = 0$*

(F6) *The scalar multiple αf , where $\alpha \in F$, is linear.*

(F7) *For another linear map $g : V \rightarrow W$, the sum $f + g$ is linear.*

(F8) *For another linear map $g : W \rightarrow U$, the composition $g \circ f$ is linear⁴.*

Definition 3.2.3 (Definition of Rank). *The dimension of the image of a linear map f is called the rank of f . In symbols,*

$$\text{rank}(f) \equiv \dim \text{Im}(f). \quad (3.11)$$

To be concrete, consider a linear map $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ and assume that $\dim \ker(f) = 2$. i.e. the kernel of f is a plane through the origin (since any sub vector space contains the zero vector) in \mathbb{R}^3 . Now consider two vectors $\mathbf{v}_1, \mathbf{v}_2 \in \ker(f) + \mathbf{k}$ which both lie in a plane parallel to $\ker(f)$, shifted by a vector \mathbf{k} . Then we have $\mathbf{v}_1 - \mathbf{v}_2 \in \ker(f)$ so that $f(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$ and, hence, by linearity $f(\mathbf{v}_1) = f(\mathbf{v}_2)$. Therefore, not only do all vectors in the kernel get mapped to the zero vector, but all vectors in a plane parallel to the kernel get mapped to the same non-zero vector. Effectively, the action of the linear map “removes” the two dimensions parallel to the kernel plane and only keeps the remaining dimension perpendicular to it. Hence, the image of this linear map is one-dimensional, which is a line through the origin. This example suggests the following theorem:

Theorem 3.2.2 (Dimension Formula). *For a linear map $f : V \rightarrow W$ we have*

$$\dim \ker(f) + \text{rank}(f) = \dim(V)^5. \quad (3.12)$$

Lemma 3.2.3 (Properties of a Linear Map (2)). *For a linear map $f : V \rightarrow W$ we have:⁶*

³Since the properties are rather intuitive and trivial, the rigorous proof for these properties is provided in section A.4.

⁴From (F7) and (F8), since the scalar multiple of a linear map and the sum of two linear maps is again a linear map, the set of linear maps does indeed form a vector space itself, denoted $\text{Hom}(V, W)$, called the homomorphisms from V to W .

⁵Since this formula is relatively intuitive and imaginable, the rigorous proof is given in section A.5.

⁶The rigorous proof is given in section A.6 due to its simplicity and imaginability.

(G1) f is bijective $\implies \dim(V) = \dim(W)$.

(G2) If f is invertible then the inverse $f^{-1} : W \rightarrow V$ is also a linear map.

3.3 Coordinates Maps $f : F^n \rightarrow V$

To map a n -dimensional column vector $\alpha \in F^n$ with components α_i to a n -dimensional vector $\mathbf{v} \in V$ over F with basis $\mathbf{v}_1, \dots, \mathbf{v}_n$, we can define the coordinate map $\varphi : F^n \rightarrow V$ as

$$\varphi(\alpha) = \sum_{i=1}^n \alpha_i \mathbf{v}_i. \quad (3.13)$$

Verifying this definition indeed satisfies the criterias for a linear map:

$$\begin{aligned} \varphi(\alpha_1 + \alpha_2) &= \sum_{i=1}^n (\alpha_{1i} + \alpha_{2i}) \mathbf{v}_i = \sum_{i=1}^n \alpha_{1i} \mathbf{v}_i + \sum_{i=1}^n \alpha_{2i} \mathbf{v}_i = \varphi(\alpha_1) + \varphi(\alpha_2) \\ \varphi(a\alpha) &= \sum_{i=1}^n (a\alpha_i) \mathbf{v}_i = a \sum_{i=1}^n \alpha_i \mathbf{v}_i = a\varphi(\alpha). \end{aligned} \quad (3.14)$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ forms a basis it is clear that $\text{Im}(\varphi) = V$ and, hence φ is surjective. On the other hand, since we have proved that a vector can be uniquely described by a basis in lemma 1.4.1, φ is injective. Thus, φ is bijective. From theorem 3.1.1, we know that φ has an inverse $\varphi^{-1} : V \rightarrow F^n$. The inverse map assigns to a vector $\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{v}_i \in V$ its coordinate vector $\alpha \in F^n$

$$\varphi^{-1}(\mathbf{v}) = \varphi^{-1} \left(\sum_{i=1}^n \alpha_i \mathbf{v}_i \right) = \alpha. \quad (3.15)$$

A linear and bijective map between two vector spaces is also referred to as a (vector space) isomorphism and two vector spaces related by such a map are called isomorphic. What the above discussion shows is that every n -dimensional vector space V over F is isomorphic to F^n by means of a coordinate map φ .

3.4 Matrices

3.4.1 Linear Maps $f : F^n \rightarrow F^m$

To map a n -dimensional column vector $\mathbf{v} \in F^n$ onto another m -dimensional column vector $\mathbf{w} \in F^m$, we can define a $m \times n$ matrix as

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \quad (3.16)$$

with entries $a_{ij} \in F$. The column vectors consisting of the i^{th} row and the j^{th} column are denoted as \mathbf{A}_i and \mathbf{A}^j respectively.

Define the matrix multiplication of a $m \times n$ matrix on a n -dimensional vector as

$$A\mathbf{v} \equiv \begin{pmatrix} a_{11}v_1 + \cdots + a_{1n}v_n \\ \vdots \\ a_{m1}v_1 + \cdots + a_{mn}v_n \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 \cdot \mathbf{v} \\ \vdots \\ \mathbf{A}_m \cdot \mathbf{v} \end{pmatrix} = \sum_{i=1}^m v_i \mathbf{A}^i \quad (3.17)$$

Using index notation,

$$(A\mathbf{v})_i = \sum_{j=1}^n a_{ij}v_j = a_{ij}v_j, \quad (3.18)$$

where a sum over j is implied by the Einstein summation convention as it appeared twice in the same term, the free index, i , on the contrary only appeared once, which labels which component of the vector is being described.

Using this notation it is straightforward to verify if matrix multiplication defined this way does indeed represent a linear map using the criterias in definition 3.2.1:

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2)_i &= a_{ij}(v_{1j} + v_{2j}) = a_{ij}v_{1j} + a_{ij}v_{2j} = f(\mathbf{v}_1)_i + f(\mathbf{v}_2)_i, \\ f(\alpha\mathbf{v})_i &= a_{ij}(\alpha v_j) = \alpha(a_{ij}v_j) = \alpha f(\mathbf{v})_i. \end{aligned} \quad (3.19)$$

In above, we have proved that every matrix multiplication represents a linear map action. However, we have not shown the converse is true, *i.e.* whether every linear map action can be represented by a matrix multiplication.

To prove this, we start with the standard unit vectors \mathbf{e}_j of F^n and $\tilde{\mathbf{e}}_i$ of F^m . The images, $f(\mathbf{e}_j) \in F^m$, of the standard unit vectors of F^n can be written as

$$f(\mathbf{e}_j) = \sum_{i=1}^m a_{ij}\tilde{\mathbf{e}}_i \quad (3.20)$$

for some suitable set of coefficients a_{ij} . Now consider an arbitrary vector $\mathbf{v} \in F^n$ as a linear combination $\mathbf{v} = \sum_{j=1}^n v_j \mathbf{e}_j$. The image of this vector under f is

$$f(\mathbf{v}) = f\left(\sum_{j=1}^n v_j \mathbf{e}_j\right) = \sum_{j=1}^n v_j f(\mathbf{e}_j) = \sum_{j=1}^n v_j \sum_{i=1}^m a_{ij}\tilde{\mathbf{e}}_i = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}v_j\right) \tilde{\mathbf{e}}_i. \quad (3.21)$$

Hence for the i^{th} component of this image we have

$$(f(\mathbf{v}))_i = \sum_{j=1}^n a_{ij}v_j = (A\mathbf{v})_i, \quad (3.22)$$

⁷This changing of order of summation is similar to what has been done in section A.7.

where we used eq. (3.18) to establish the second equality.

Lemma 3.4.1. *Every linear map $f : F^n \rightarrow F^m$ can be written in terms of an $m \times n$ matrix A , such that $f(\mathbf{v}) = A\mathbf{v} \forall \mathbf{v} \in F^n$. If $f(\mathbf{e}_j) = \sum_{i=1}^m a_{ij}\tilde{\mathbf{e}}_i$ for the standard unit vectors \mathbf{e}_i of F^n and $\tilde{\mathbf{e}}_i$ of F^m , then a_{ij} are the entries of A .*

Example: Cross Product as Matrix

Question: Prove that cross product is a linear map and find the matrix corresponding to it.

Solution: Consider a fixed vector $\mathbf{n} \in \mathbb{R}^3$ and a map $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined by $f(\mathbf{v}) = \mathbf{n} \times \mathbf{v}$. From the properties of the vector product, it is easy to show that this map is linear as it satisfies the linearity conditions (E1) and (E2)

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2) &= \mathbf{n} \times (\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{n} \times \mathbf{v}_1 + \mathbf{n} \times \mathbf{v}_2 = f(\mathbf{v}_1) + f(\mathbf{v}_2) \\ f(\alpha\mathbf{v}) &= \mathbf{n} \times (\alpha\mathbf{v}) = \alpha\mathbf{n} \times \mathbf{v} = \alpha f(\mathbf{v}). \end{aligned} \quad (3.23)$$

Hence we know from lemma 3.4.1 that f can be described by a 3×3 matrix A which can be worked out by studying the action of f on the standard unit vectors \mathbf{e}_i . If $\mathbf{n} = (n_1 \ n_2 \ n_3)^T$ we find by explicit computation that

$$\begin{aligned} f(\mathbf{e}_1) &= \mathbf{n} \times \mathbf{e}_1 = n_3\mathbf{e}_2 - n_2\mathbf{e}_3 \\ f(\mathbf{e}_2) &= \mathbf{n} \times \mathbf{e}_2 = -n_3\mathbf{e}_1 + n_1\mathbf{e}_3 \\ f(\mathbf{e}_3) &= \mathbf{n} \times \mathbf{e}_3 = n_2\mathbf{e}_1 - n_1\mathbf{e}_2. \end{aligned} \quad (3.24)$$

From lemma 3.4.1 we know that the coefficients which appear in the expression for $f(\mathbf{e}_j)$ form the j^{th} column of the matrix A . Hence the desired matrix is

$$A = \begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix}, \quad (3.25)$$

and we have $f(\mathbf{v}) = \mathbf{n} \times \mathbf{v} = A\mathbf{v}$ for all vectors $\mathbf{v} \in \mathbb{R}^3$.

Everything is much more elegant in index notation where

$$A_{ij} = [f(\mathbf{e}_j)]_i = [\mathbf{n} \times \mathbf{e}_j]_i = \epsilon_{ikl}n_k [\mathbf{e}_j]_l = \epsilon_{ikl}n_k \delta_{jl} = \epsilon_{ikj}n_k \quad (3.26)$$

which is in agreement with the explicit form of matrix A .

3.4.2 Basic Properties

A very specific matrix is the identity (or unit) matrix $\mathbb{I}_n : F^n \rightarrow F^n$ given by

$$\mathbb{I}_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \text{ or } (\mathbb{I}_n)_{ij} = \delta_{ij} \implies (\mathbb{I}\mathbf{v})_i = \delta_{ij}v_j = v_i. \quad (3.27)$$

So, seen as linear map, the identity matrix corresponds to the identity map.

More generally, a diagonal matrix is a matrix D with non-zero entries only along the diagonal, so $D_{ij} = 0 \forall i \neq j$. It can be written as

$$D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \equiv \text{diag}(d_1, \dots, d_n). \quad (3.28)$$

A $m \times n$ matrix A is called square or quadratic if the number of rows is equals to the number of column, so $m = n$.

The complex conjugate $A^* : F^m \rightarrow F^n$ of a matrix $A : F^m \rightarrow F^n$ is simply the matrix whose entries are the complex conjugates of the entries in A , so in component form, $A_{ij}^* = (A_{ij})^*$.

The transpose of an $n \times m$ matrix $A : F^m \rightarrow F^n$ is a $m \times n$ matrix $A^T : F^n \rightarrow F^m$ obtained by exchanging the rows and columns of A . In component form, this means $(A^T)_{ij} = A_{ji}$.

A square matrix A is said to be symmetric if $A = A^T$, or $A_{ij} = A_{ji}$ for all entries, and is called anti-symmetric if $A = -A^T$, or $A_{ij} = -A_{ji}$ for all entries (note that all diagonal entries A_{ii} of an anti-symmetric matrix vanish).

A combination of the previous two operations is the hermitian conjugate of a $n \times m$ matrix $A : F^m \rightarrow F^n$ which is defined as a $m \times n$ matrix $A^\dagger : F^n \rightarrow F^m$ obtained by taking the complex conjugate of the transpose of A , *i.e.* $A^\dagger \equiv (A^T)^*$. A quadratic matrix is said to be hermitian if the matrix is invariant under hermitian conjugation, *i.e.* if $A = A^\dagger$, and anti-hermitian if $A = -A^\dagger$.

Trivially, $(A + B)^\dagger = A^\dagger + B^\dagger$ and $(\alpha A)^\dagger = \alpha^* A^\dagger$.

3.4.3 Rank of a Matrix

From eq. (3.17) we see that the image of \mathbf{v} is given by a linear combination of the column vectors \mathbf{A}^i with coefficients equal to the component of \mathbf{v} . Thus

Thus, the image of \mathbf{v} is given by a linear combination of the column vectors \mathbf{A}^j with coefficients equal to the components of \mathbf{v} . This observation tells us that

$$\text{Im}(A) = \text{Span}(\mathbf{A}^1, \dots, \mathbf{A}^m), \quad (3.29)$$

so the image of the matrix is spanned by its column vectors.

We have earlier define the rank of a linear map as the dimension of its image, so $\text{rank}(f) = \dim \text{Im}(f)$. Treating matrix multiplication as the action of a linear map, we have

$$\begin{aligned} \text{rank}(A) &= \dim \text{Im}(A) = \dim \text{Span}(\mathbf{A}^1, \dots, \mathbf{A}^m) \\ &= \text{maximal number of linear independent column vectors of } A. \end{aligned} \quad (3.30)$$

For obvious reasons this is also sometimes called the column rank of the matrix A . This terminology suggests we can also define the row rank of the matrix A as the maximal

number of linearly independent row vectors of A . Having two types of ranks available for a matrix seems awkward but fortunately have

Theorem 3.4.2. *Row and column rank are equal for any matrix*⁸.

Example: Kernel and Image of A Matrix

Question: Consider the 3×3 matrix

$$A = \begin{pmatrix} -1 & 4 & 3 \\ 2 & -3 & -1 \\ 3 & 2 & 5 \end{pmatrix}, \quad (3.31)$$

which defines a map $A : \mathbb{R}^3 \rightarrow \mathbb{R}^3$.

Find the kernel and the image of this matrix.

Solution: It is clear that the first two columns of this matrix are linearly independent (they are not multiples of each other) and that the third column is the sum of the first two. Hence, the rank of this matrix is two. This means that the dimension of its image is two while, from eq. (3.12), the dimension of its kernel is one. To find the kernel of A explicitly we have to solve $A\mathbf{v} = \mathbf{0}$. With $\mathbf{v} = (x \ y \ z)^T$ this leads to $x = y = -z$. The image of A is, in general, spanned by the column vectors, but since $\mathbf{A}^3 = \mathbf{A}^1 + \mathbf{A}^2$, it is already spanned by the first two columns. In conclusion, we have

$$\ker(A) = \text{Span} \left(\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right) \text{ and } \text{Im}(A) = \text{Span} \left(\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix} \right). \quad (3.32)$$

3.4.4 Matrix Multiplication

We have proved earlier in (F8) of lemma 3.2.1 that the composition of linear maps is again linear. Since all linear maps between column vectors are matrices. Hence, the composition of two matrices must again be a matrix.

To work this out more explicitly, we start with an $m \times n$ matrix A and an $r \times m$ matrix B which generate linear maps according to chain $F^n \xrightarrow{A} F^m \xrightarrow{B} F^r$. We would like to determine the matrix C which describes the composite map $B \circ A : F^n \rightarrow F^r$. By straightforward computation we find

$$\begin{aligned} (B(A\mathbf{v}))_i &= \sum_{j=1}^n B_{ij}(A\mathbf{v})_j = \sum_{j=1}^n B_{ij} \left(\sum_{k=1}^n A_{jk}\mathbf{v}_k \right) = \sum_{k=1}^n \left(\left(\sum_{j=1}^n B_{ij}A_{jk} \right) \mathbf{v}_k \right) = \sum_{k=1}^n C_{ik}v_k = (C\mathbf{v})_i \\ &\implies C_{ik} = \sum_{j=1}^n B_{ij}A_{jk} = \mathbf{B}_i \cdot \mathbf{A}^k. \end{aligned} \quad (3.33)$$

⁸The proof of this non-trivial theorem is not given here due to its length and relative insignificance, but provided in section A.7.

Note that the two free indices in C_{ik} means that there are two components of C that we are interested in and the indices are describing.

Matrix multiplication is associative, as can be shown by

$$\begin{aligned}(A(BC))_{ij} &= A_{ik}(BC)_{kj} = A_{ik}B_{kl}C_{lj} = (AB)_{il}C_{lj} = ((AB)C)_{ij} \\ &\implies A(BC) = (AB)C.\end{aligned}\tag{3.34}$$

A useful relationship involving multiplication and transposition of matrices is

$$\begin{aligned}((AB)^T)_{ij} &= (AB)_{ji} = A_{jk}B_{ki} = B_{ki}A_{jk} = (B^T)_{ik}(A^T)_{kj} = (B^TA^T)_{ij} \\ &\implies (AB)^T = B^TA^T.\end{aligned}\tag{3.35}$$

For complex case, since $(AB)^* = A^*B^*$, we have $(AB)^\dagger = B^\dagger A^\dagger$.

Using matrix terminology, the dot product of two m -dimensional column vectors \mathbf{v} and \mathbf{w} can be written as

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w}.\tag{3.36}$$

3.4.5 Inverse of a Matrix

For a matrix $A : F^n \rightarrow F^m$ representing a linear map $f : V \rightarrow W$ to have an inverse, it must be bijective as proved in theorem 3.1.1. Thus from (F5) and (G1) we have $\dim \ker(f) = 0$ and $\dim(V) = \dim(W)$, which implies $\text{rank}(f) = n = m$ using eq. (3.12)

Lemma 3.4.3 (Properties of Matrix Inverse). *A quadratic $n \times n$ matrix $A : F^n \rightarrow F^n$ is invertible if and only if its rank is maximal, i.e. $\text{rank}(A) = n$. If A, B are two invertible $n \times n$ matrices we have⁹*

(H1) *The inverse matrix, denoted A^{-1} , is the unique matrix satisfying $AA^{-1} = A^{-1}A = \mathbb{I}_n$.*

(H2) $(AB)^{-1} = B^{-1}A^{-1}$

(H3) A^{-1} is invertible and $(A^{-1})^{-1} = A$

(H4) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$

3.4.6 Gaussian Elimination

Definition 3.4.1. *The following manipulations of a matrix are called elementary row¹⁰ operations.*

(I1) *Exchange two rows.*

(I2) *Add a multiple of one row to another.*

(I3) *Multiply a row with a non-zero scalar.*

The kind of matrix we want to reduce to is called (upper) echelon form¹¹, where the

⁹The proof is given in section A.8 as always.

¹⁰Analogous definitions hold for elementary column operations.

¹¹A neat fact about elementary row operations is that they can be generated by multiplying with

magic squares and coding theory

entries a_{ij_i} are non-zero for all $i = 1, \dots, r$, all other entries above are arbitrary and all entries below are zero which has the form

$$A = \begin{pmatrix} \dots & a_{1j_1} & & & * \\ & & a_{2j_2} & & \ddots \\ & & & \ddots & \\ & & & & a_{rj_r} & \dots \\ 0 & & & & & \vdots \end{pmatrix} \quad (3.37)$$

Lemma 3.4.4. *The first r row vectors in this matrix are linearly independent¹³.*

This fact allow us to easily read off the rank of a matrix, which is not affected by the elementary row operations, as

$$\text{rank}(A) = r = \text{the number of steps in upper echelon form} \quad (3.38)$$

To devise an algorithm to compute the inverse of a matrix, using elementary row operations, we observe that every quadratic, invertible $n \times n$ matrix A can be converted into identity matrix \mathbb{I}_n by a sequence of row operations. Schematically, this works as follows:

$$A \xrightarrow{\text{echelon form}} \begin{pmatrix} a'_{11} & & * \\ & a'_{22} & \\ & & \ddots \\ 0 & & & a'_{nn} \end{pmatrix} \xrightarrow{(I1, I2)} \begin{pmatrix} a'_{11} & & 0 \\ & a'_{22} & \\ & & \ddots \\ 0 & & & a'_{nn} \end{pmatrix} \xrightarrow{(I3)} \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \mathbb{I}_n \quad (3.39)$$

In the first step, we bring A into upper echelon form by the algorithm already discussed. At this point we can read off the rank of the matrix. If $\text{rank}(A) < n$ the inverse does not exist and we can stop. On the other hand, if $\text{rank}(A) = n$ then all diagonal entries a'_{ii} in the upper echelon form must be non-zero.

Therefore in the second step we can make all entries above the diagonal zero. In this way we arrive at a diagonal matrix with diagonal entries $a'_{ii} \neq 0$ which in the final step can be converted into the identity matrix by multiplying every row with a suitable scalar.

This process also means that we can find a set of matrices P_1, \dots, P_k generating appropriate elementary row operations such that

$$\mathbb{I}_n = P_k \dots P_1 A \implies A^{-1} = P_1 \dots P_k \mathbb{I}_n. \quad (3.40)$$

These equations imply an explicit (though tedious) algorithm to compute the inverse of a square matrix. We first convert A into the identity matrix \mathbb{I}_n using elementary

certain matrices from the left. In other words, to perform a row operation on a matrix A , we can always find a suitable matrix P such that the row operation is generated by $A \rightarrow PA$ ¹². This means we can bring a matrix A into upper echelon form by matrix multiplications $P_k \dots P_1 A$.

¹³The proof of this lemma is provided in section A.9.

row operations as described above, and then simply carry out the same operations by multiplying the corresponding matrix in parallel.

3.5 Linear Maps $f : V \rightarrow W$

Start with a linear map $f : V \rightarrow W$ over F with dimensions n and m , respectively.

We introduce a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of V and a basis $\mathbf{w}_1, \dots, \mathbf{w}_m$ of W , then we have $f(\mathbf{v}_j) = \sum_{i=1}^m a_{ij} \mathbf{w}_i$ since the images of the basis vectors in V are in W .

An arbitrary vector $\mathbf{v} \in V$ and $\mathbf{w} = f(\mathbf{v}) \in W$ can be written as $\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{v}_i$ with coordinate vectors $\boldsymbol{\alpha} = (\alpha_1 \dots \alpha_n)^T$ and $\mathbf{w} = \sum_{j=1}^m \beta_j \mathbf{w}_j$ with coordinate vectors $\boldsymbol{\beta} = (\beta_1 \dots \beta_m)^T$, respectively.

We can introduce coordinate maps $\varphi : F^n \rightarrow V$ and $\psi : F^m \rightarrow W$ relative to each basis which act as $\varphi(\boldsymbol{\alpha}) = \sum_{i=1}^n \alpha_i \mathbf{v}_i$ and $\psi(\boldsymbol{\beta}) = \sum_{j=1}^m \beta_j \mathbf{w}_j$.

The situation so far can be summarized by the diagram

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \varphi \uparrow & & \uparrow \psi \\ F^n & \xrightarrow{A=?} & F^m \end{array} \quad (3.41)$$

Abstractly, the matrix A is given by

$$A = \psi^{-1} \circ f \circ \varphi. \quad (3.42)$$

From lemma 3.4.1 we know that we can work out the components of a matrix by letting it act on the standard unit vectors in F^n , so

$$\begin{aligned} A\mathbf{e}_j &= \psi^{-1} \circ f \circ \varphi(\mathbf{e}_j) = \psi^{-1} \circ f(\mathbf{v}_j) = \psi^{-1} \left(\sum_{i=1}^m a_{ij} \mathbf{w}_i \right) \\ &= \sum_{i=1}^m a_{ij} \psi^{-1}(\mathbf{w}_i) = \sum_{i=1}^m a_{ij} \tilde{\mathbf{e}}_i. \end{aligned} \quad (3.43)$$

where $\tilde{\mathbf{e}}_i$ are the standard unit vectors in F^m . Comparing with lemma 3.4.1 it follows that a_{ij} are the entries of the desired matrix A .

Thus we extend lemma 3.4.1 to general basis vectors $\mathbf{v}_i \in V$ (not just $\mathbf{e}_i \in F^n$).

Lemma 3.5.1. *Let $f : V \rightarrow W$ be a linear map, $\mathbf{v}_1, \dots, \mathbf{v}_n$ a basis of V and $\mathbf{w}_1, \dots, \mathbf{w}_m$ a basis of W . The entries a_{ij} of the $m \times n$ matrix A which describes this linear map relative to this choice of basis can be read off from the images of the basis vectors as*

$$f(\mathbf{v}_j) = \sum_{i=1}^m a_{ij} \mathbf{w}_i. \quad (3.44)$$

We have $\text{rank}(A) = \text{rank}(f)$ and, in particular, A is invertible if and only if f is¹⁴.

Example:

Question:

Solution: For example, consider the linear map $B : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by the matrix

$$B = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}, \quad (3.45)$$

For simplicity, we choose the same basis for the domain and the co-domain, namely $\mathbf{v}_1 = \mathbf{w}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{v}_2 = \mathbf{w}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Then, the images of the basis vectors under B are

$$B\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -1\mathbf{w}_1 - 2\mathbf{w}_2 \text{ and } B\mathbf{v}_2 = \begin{pmatrix} -1 \\ -2 \end{pmatrix} = -1\mathbf{w}_1 + 0\mathbf{w}_2. \quad (3.46)$$

By arranging the coefficients from $B\mathbf{v}_1$ into the first column of a matrix the coefficients from $B\mathbf{v}_2$ into the second column we get the matrix

$$B' = \begin{pmatrix} -1 & -1 \\ -2 & 0 \end{pmatrix}, \quad (3.47)$$

which is the matrix representing the linear map B relative to the basis $\mathbf{v}_1, \mathbf{v}_2$.

To get what this exactly means, write an arbitrary 2-dimensional vector in \mathbb{R}^2 as

$$\begin{pmatrix} x \\ y \end{pmatrix} = x'\mathbf{v}_1 + y'\mathbf{v}_2 = \begin{pmatrix} x' - y' \\ 2x' + y' \end{pmatrix} \quad (3.48)$$

so that a vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is described relative to the basis $\mathbf{w}_1, \mathbf{w}_2$ by the coordinate vector $\begin{pmatrix} x' \\ y' \end{pmatrix}$.

3.6 Change of Basis

We consider a linear map $f : V \rightarrow V$ from a vector space to itself. The two sets of basis vectors, coordinate maps and representing matrices are then denoted by

basis of V	coordinate map	coordinate vector	representing matrix	
$\mathbf{v}_1, \dots, \mathbf{v}_n$	$\varphi(\alpha) = \sum_{i=1}^n \alpha_i \mathbf{v}_i$	$\alpha = (\alpha_1, \dots, \alpha_n)^T$	$A = \varphi^{-1} \circ f \circ \varphi$	(3.49)
$\mathbf{v}'_1, \dots, \mathbf{v}'_n$	$\varphi'(\alpha') = \sum_{i=1}^n \alpha'_i \mathbf{v}'_i$	$\alpha' = (\alpha'_1, \dots, \alpha'_n)^T$	$A' = \varphi'^{-1} \circ f \circ \varphi'$	

¹⁴The proof of this claim is provided in ??.

We would like to then find a relationship between A and A' , *i.e.* between the representing matrices for f relative to the unprimed and the primed basis, so

$$\begin{aligned} A' &= \varphi'^{-1} \circ f \circ \varphi' = \varphi'^{-1} \circ \varphi \circ \varphi^{-1} \circ f \circ \varphi \circ \varphi'^{-1} \circ \varphi' \\ &= \underbrace{\varphi'^{-1} \circ \varphi}_P \circ \underbrace{\varphi^{-1} \circ f \circ \varphi}_A \circ \underbrace{\varphi^{-1} \circ \varphi'}_{P^{-1}} = PAP^{-1}. \end{aligned} \quad (3.50)$$

The interpretation of $P = \varphi'^{-1} \circ \varphi$ can be seen from

$$\alpha' = \varphi'^{-1}(\mathbf{v}) = \varphi'^{-1} \circ \varphi(\alpha) = P\alpha. \quad (3.51)$$

Hence, P converts unprimed coordinate vectors α into the corresponding primed coordinate vector α' . Note that there is a intuitive explanation to eq. (3.50): When acting the equation on a primed coordinate vector α' , the first thing we obtain on the RHS is $P^{-1}\alpha'$, which is the corresponding unprimed coordinate vector on which the matrix A on the RHS can sensibly act, converting it into another unprimed coordinate vector. The final action of P converts this back into a primed coordinate vector. Altogether, this is the action of the matrix A' on α' .

To compute P , we can simply use lemma 3.5.1.

Example: Basis Transformation of a Matrix

Question: Relative to the unprimed basis $\mathbf{v}_1 = \mathbf{e}_1$ and $\mathbf{v}_2 = \mathbf{e}_2$ of standard unit vectors, a linear map $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is described by the matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Determine the matrix A' which describes the same linear map relative to the basis $\mathbf{v}'_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\mathbf{v}'_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Solution: The first way is to use lemma 3.5.1 and compute the images of the basis vectors in the domain

$$A \quad (3.52)$$

4.1 Definition of a Determinant

Definition 4.1.1. A determinant maps n vectors $\mathbf{a}_1, \dots, \mathbf{a}_n \in F^n$ to a number, denoted $\det(\mathbf{a}_1, \dots, \mathbf{a}_n) \in F$, such that the following properties are satisfied:

$$(J1) \det(\dots, \alpha \mathbf{a} + \beta \mathbf{b}, \dots) = \alpha \det(\dots, \mathbf{a}, \dots) + \beta \det(\dots, \mathbf{b}, \dots)$$

This means the determinant is linear in each argument.

$$(J2) \det(\dots, \mathbf{a}, \dots, \mathbf{b}, \dots) = -\det(\dots, \mathbf{b}, \dots, \mathbf{a}, \dots)$$

This means the determinant is completely anti-symmetric.

$$(J3) \det(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1$$

The determinant of the standard unit vectors is one.

The determinant of an $n \times n$ matrix A is defined as the determinant of its column vectors, so $\det(A) := \det(\mathbf{A}^1, \dots, \mathbf{A}^n)$.

An easy but important conclusion from these properties is that the determinant with two same arguments must vanish, so from (J2) it follows that $\det(\dots, \mathbf{a}, \dots, \mathbf{a}, \dots) = -\det(\dots, \mathbf{a}, \dots, \mathbf{a}, \dots)$, which means that $\det(\dots, \mathbf{a}, \dots, \mathbf{a}, \dots) = 0$.

Before we can find an explicit form of a determinant, we have to introduce the concept of permutation, which is an operation which changes the order of a certain set of n objects. Mathematically, the set of all permutations of n objects is given by

$$S_n \equiv \{ \sigma : \{1, \dots, n\} \rightarrow \{1, \dots, n\} \mid \sigma \text{ is bijective} \}, \quad (4.1)$$

and this set has $n!$ elements. A useful notation for a permutation mapping $1 \rightarrow \sigma(1), \dots, n \rightarrow \sigma(n)$ is¹

$$\sigma = \begin{pmatrix} 1 & \dots & n \\ \sigma(1) & \dots & \sigma(n) \end{pmatrix}. \quad (4.2)$$

The specific permutations which only swap two numbers and leave all other numbers unchanged are called transpositions. A basic and important fact about permutations, is

¹Note, despite the similar notation, this is not a matrix

that every permutation can be written as a composition of transpositions, so any $\sigma \in S_n$ can be written as $\sigma = \tau_1 \circ \dots \circ \tau_k$, where $\tau_1, \dots, \tau_k \in S_n$ are transpositions. It can be shown also that the number of transpositions required is always either even or odd for a given transposition. It therefore makes sense to define the sign of permutation as

$$\text{sgn}(\sigma) \equiv (-1)^k = \begin{cases} +1 & : \text{“even permutation”} \\ -1 & : \text{“odd permutation”} \end{cases} . \quad (4.3)$$

Some direct consequences from this definitions are $\text{sgn}(\sigma_1 \circ \sigma_2) = \text{sgn}(\sigma_1)\text{sgn}(\sigma_2)$ and $1 = \text{sgn}(\sigma \circ \sigma^{-1}) = \text{sgn}(\sigma)\text{sgn}(\sigma^{-1}) \implies \text{sgn}(\sigma^{-1}) = \text{sgn}(\sigma)$.

To derive an explicit formula for the determinant, we start with an $n \times n$ matrix A with entries a_{ij} whose column vectors we write as linear combinations of the standard unit vectors:

$$\mathbf{A}^i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix} = \sum_j a_{ji} \mathbf{e}_j. \quad (4.4)$$

We find

$$\begin{aligned} \det(\mathbf{A}) &= \det(\mathbf{A}^1, \dots, \mathbf{A}^n) = \det\left(\sum_{j_1=1}^n a_{j_1 1} \mathbf{e}_{j_1}, \dots, \sum_{j_n=1}^n a_{j_n n} \mathbf{e}_{j_n}\right) \\ &\stackrel{(J1)}{=} \sum_{j_1, \dots, j_n} a_{j_1 1} \dots a_{j_n n} \det(\mathbf{e}_{j_1}, \dots, \mathbf{e}_{j_n}) \stackrel{j_a = \sigma(a)}{=} \sum_{\sigma \in S_n} a_{\sigma(1)1} \dots a_{\sigma(n)n} \det(\mathbf{e}_{\sigma(1)}, \dots, \mathbf{e}_{\sigma(n)}) \\ &\stackrel{(J2)}{=} \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n} \det(\mathbf{e}_1, \dots, \mathbf{e}_n) \stackrel{(J3)}{=} \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n} \end{aligned} \quad (4.5)$$

where in the second step of the second line we replace \sum_{j_1, \dots, j_n} as $\sum_{\sigma \in S_n}$ since when we sum over all j_i for $i = 1, \dots, n$, consider the term when $(j_1, \dots, j_k, \dots, j_n) = (2, \dots, 6, \dots, 1)$, since it is a possible permutation of $(1, \dots, n)$, it is in S_n . For the cases such as $(j_1, \dots, j_k, \dots, j_n) = (5, \dots, 5, \dots, 2)$, where since it is not a possible permutation of $(1, \dots, n)$, it is not included in S_n , the term is zero since the determinant of any set of vectors consisting of two identical vectors is always zero as shown before.

Using the n -dimensional generalization of the Levi-Civita symbol, defined by

$$\epsilon_{i_1 \dots i_n} = \begin{cases} +1 & \text{if } i_1, \dots, i_n \text{ is an even permutation of } 1, \dots, n \\ -1 & \text{if } i_1, \dots, i_n \text{ is an odd permutation of } 1, \dots, n \\ 0 & \text{otherwise} \end{cases} , \quad (4.6)$$

the determinant of a matrix A can be written by

$$\det(A) = \epsilon_{i_1 \dots i_n} a_{i_1 1} \dots a_{i_n n}, \quad (4.7)$$

with a sum over the n indices i_1, \dots, i_n is implied.

In three dimensions we find

$$\begin{aligned} \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} &= \epsilon_{ijk} a_i b_j c_k \\ &= a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_2 b_1 c_3 - a_3 b_2 c_1 - a_1 b_3 c_2 \\ &= \langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \end{aligned} \quad (4.8)$$

An interesting class of matrices for which the determinant is simple consists of upper (or lower) triangular matrices, *i.e.* matrices with all entries below (or above) the diagonal vanishing. In this case

$$\det \begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix} = a_1 \dots a_n \quad (4.9)$$

so the determinant is simply the produce of the diagonal elements.

4.2 Properties of a Determinant

We now investigate some important properties of determinants We begin with the following lemma:

Lemma 4.2.1. *The determinant of a matrix and its transpose are the same, so $\det(A) = \det(A^T)$.*

Proof. By setting $j_a = \sigma(a)$, for a permutation $\sigma \in S_n$, we can re-write a term in the sum for the determinant in eq. (4.5) as

$$A_{\sigma(1)1} \dots A_{\sigma(n)n} = A_{j_1 \sigma^{-1}(j_1)} \dots A_{j_n \sigma^{-1}(j_n)} = A_{1 \sigma^{-1}(1)} \dots A_{n \sigma^{-1}(n)}. \quad (4.10)$$

From this observation, the determinant can be written as

$$\begin{aligned} \det(A) &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) A_{1 \sigma^{-1}(1)} \dots A_{n \sigma^{-1}(n)} = \sum_{\sigma^{-1} \in S_n} \text{sgn}(\sigma^{-1}) A_{1 \sigma^{-1}(1)} \dots A_{n \sigma^{-1}(n)} \\ &\stackrel{\rho = \sigma^{-1}}{=} \sum_{\rho \in S_n} \text{sgn}(\rho) (A^T)_{\rho(1)1} \dots (A^T)_{\rho(n)n} = \det(A^T). \end{aligned} \quad (4.11)$$

□

Theorem 4.2.2. $\det(AB) = \det(A) \det(B)$, for any two $n \times n$ matrices A, B .

Proof. Since $(AB)_{ij} = \sum_k A_{ik} B_{kj}$, so the j^{th} column of AB is

$$(AB)^j = \sum_k B_{kj} \mathbf{A}^k, \quad (4.12)$$

where \mathbf{A}^k are the columns of A . Hence

$$\begin{aligned} \det(AB) &= \det((AB)^1, \dots, (AB)^n) = \det\left(\sum_{k_1} B_{k_1 1} \mathbf{A}^{k_1}, \dots, \sum_{k_n} B_{k_n n} \mathbf{A}^{k_n}\right) \\ &\stackrel{(J1)}{=} \sum_{k_1, \dots, k_n} B_{k_1 1} \dots B_{k_n n} \det(\mathbf{A}^{k_1}, \dots, \mathbf{A}^{k_n}) \\ &\stackrel{k_a = \sigma(a)}{=} \sum_{\sigma \in S_n} B_{\sigma(1)1} \dots B_{\sigma(n)n} \det(\mathbf{A}^{\sigma(1)}, \dots, \mathbf{A}^{\sigma(n)}) \\ &\stackrel{(J2)}{=} \underbrace{\sum_{\sigma \in S_n} \text{sgn}(\sigma) B_{\sigma(1)1} \dots B_{\sigma(n)n}}_{\det(B)} \underbrace{\det(\mathbf{A}^1, \dots, \mathbf{A}^n)}_{\det(A)} = \det(A) \det(B). \end{aligned} \quad (4.13)$$

□

Corollary 4.2.2.1. *For an $n \times n$ matrix A we have:*

$$A \text{ is bijective (i.e. } A \text{ has an inverse)} \iff \det(A) \neq 0 \quad (4.14)$$

If A is invertible then $\det(A^{-1}) = (\det(A))^{-1}$.

Proof. If A is bijective it has an inverse A^{-1} and $1 = \det(\mathbb{I}_n) = \det(AA^{-1}) = \det(A) \det(A^{-1})$. This implies that $\det(A) \neq 0$ and that $\det(A^{-1}) = (\det(A))^{-1}$ which is the second part of our assertion. □

This implies that the determinant remains unchanged under basis transformations and as a result the determinant is a genuine property of a linear map, since

$$\det(PAP^{-1}) = \det(P) \det(A) \det(P)^{-1} = \det(A). \quad (4.15)$$

Our next goal is to find a recursive method to calculate the determinant, essentially by writing the determinant of a matrix in terms of determinants of sub-matrices. For an $n \times n$ matrix A , we define the associated $n \times n$ matrix as

$$\tilde{A}_{(i,j)} = \begin{pmatrix} & 0 & \leftarrow j^{\text{th}} \text{ col.} \\ & \vdots & \\ & 0 & \\ 0 & \dots & 0 & 1 & 0 & 0 & 0 \\ & 0 & \\ & \vdots & \\ & 0 & \end{pmatrix} \leftarrow i^{\text{th}} \text{ row} \quad (4.16)$$

The co-factor matrix is then defined as a $n \times n$ matrix C with entries

$$C_{ij} \equiv \det(\tilde{A}_{(i,j)}) = (-1)^{i+j} \det(A_{(i,j)}), \quad (4.17)$$

where $A_{(i,j)}$ is a $(n-1) \times (n-1)$ matrix where the i^{th} row and the j^{th} column is deleted. The $(-1)^{i+j}$ factor comes from the process of permuting the rows and column of $\tilde{A}_{(i,j)}$ before deleting the i^{th} row and the j^{th} column such that the determinant would remain unchanged after the deletion.

Lemma 4.2.3. *For an $n \times n$ matrix A with associated co-factor matrix C , we have*

$$C^T A = \det(A) \mathbb{I}. \quad (4.18)$$

Proof. This follows from the definition of the co-factor matrix, more or less by direct calculation.

$$\begin{aligned} (C^T A)_{ij} &= \sum_k (C^T)_{ik} A_{kj} = \sum_k A_{kj} C_{ki} = \sum_k A_{kj} \det(\tilde{A}_{(k,i)}) \\ &= \sum_k A_{kj} \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{e}_k, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) \\ &= \det\left(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \sum_k A_{kj} \mathbf{e}_k, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n\right) \\ &= \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{A}^j, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) \stackrel{(5.1)}{=} \delta_{ij} \det(A) = (\det(A) I_n)_{ij} \end{aligned} \quad (4.19)$$

□

An immediate conclusion from lemma 4.2.3 is

$$\det(A) = (C^T A)_{jj} = \sum_i (C^T)_{ji} A_{ij} = \sum_i C_{ij} A_{ij} = \sum_i (-1)^{i+j} A_{ij} \det(A_{(i,j)}). \quad (4.20)$$

This is known as the Laplace expansion of the determinant. Note that we can choose any column j and compute the determinant of A by summing over the entries i in this column times the determinants of the corresponding sub-matrices $A_{(i,j)}$.

Also, lemma 4.2.3 is a new method to compute the inverse of a matrix. If A is invertible, then from corollary 4.2.2.1 $\det(A) \neq 0$ and we can divide by $\det(A)$ to get

$$A^{-1} = \frac{1}{\det(A)} C^T. \quad (4.21)$$

For a large matrix, it is usually easier to first convert the matrix into upper triangular form, since exchanging rows only change the sign of a determinant, multiplying a row with a non-zero scalar only change the determinant by the same factor, and adding a multiple of one row to another leaves the determinant unchanged, since this is equivalent of adding a multiple of one column to another because of $\det(A^T) = \det(A)$, thus, $\det(\mathbf{A}^1, \dots, \mathbf{A}^j, \dots, \mathbf{A}^k + \lambda \mathbf{A}^j, \dots, \mathbf{A}^n) = \det(\mathbf{A}^1, \dots, \mathbf{A}^n) + \lambda \det(\mathbf{A}^1, \dots, \mathbf{A}^j, \dots, \mathbf{A}^j, \dots, \mathbf{A}^n) = \det(\mathbf{A}^1, \dots, \mathbf{A}^n)$.

4.3 Cramer's Rule

Here we introduce Cramer's Rule, which uses determinants to solve systems of linear equations $A\mathbf{x} = \mathbf{b}$ for the cases of quadratic and invertible $n \times n$ matrices A and where $\mathbf{x} = (x_1 \dots x_n)^T$. We first define the matrices

$$B_{(i)} \equiv (\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{b}, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n), \quad (4.22)$$

Then we find

$$\begin{aligned} \det(B_{(i)}) &= \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{b}, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) = \det\left(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \sum_j x_j \mathbf{A}^j, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n\right) \\ &= \sum_j x_j \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{A}^j, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) = x_i \det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{A}^i, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n) \\ &= x_i \det(A), \end{aligned} \quad (4.23)$$

where we have used ?? to express $\mathbf{b} = \sum_j x_j \mathbf{A}^j$.

Thus

$$x_i = \frac{\det(B_{(i)})}{\det(A)} = \frac{\det(\mathbf{A}^1, \dots, \mathbf{A}^{i-1}, \mathbf{b}, \mathbf{A}^{i+1}, \dots, \mathbf{A}^n)}{\det(A)}. \quad (4.24)$$

Scalar Products

We have introduced the standard scalar product on \mathbb{R}^n (the dot product). Here we study its generalizations to arbitrary real and complex vector spaces.

5.1 Real and Hermitian Scalar Products

Definition 5.1.1. A real (hermitian) scalar product on a vector space V over $F = \mathbb{R}$ ($F = \mathbb{C}$) is a map $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ (\mathbb{C}) satisfying

(K1) $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$, for a real scalar product, $F = \mathbb{R}$
 $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle^*$, for a hermitian scalar product, $F = \mathbb{C}$

(K2) $\langle \mathbf{v}, \alpha \mathbf{u} + \beta \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{u} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle$

(K3) $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ if $\mathbf{v} \neq 0$

for all vectors $\mathbf{v}, \mathbf{u}, \mathbf{w} \in V$ and all scalars $\alpha, \beta \in F$.

If (K1) and (K2), but not necessarily (K3) are satisfied, then $\langle \cdot, \cdot \rangle$ is called a bi-linear form (in the real case $F = \mathbb{R}$) or a sesqui-linear form (in the complex case $F = \mathbb{C}$).

In the real case, (K1) and (K2) imply that

$$\langle \alpha \mathbf{v} + \beta \mathbf{u}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle + \beta \langle \mathbf{u}, \mathbf{w} \rangle, \quad (5.1)$$

while in the complex case, they imply that

$$\langle \alpha \mathbf{v} + \beta \mathbf{u}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle + \beta \langle \mathbf{u}, \mathbf{w} \rangle. \quad (5.2)$$

Property (K3) ensure that we can sensibly define the norm (or length) of a vector as

$$|\mathbf{v}| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \quad (5.3)$$

since $\langle \mathbf{v}, \mathbf{v} \rangle$ must be greater than zero. Also, property (K1) demands that $\langle \mathbf{v}, \mathbf{v} \rangle$ is real since $\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle^*$.

The Cauchy-Schwarz inequality and the triangle inequality are then

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq |\mathbf{v}| |\mathbf{w}| \text{ and } |\mathbf{v} + \mathbf{w}| \leq |\mathbf{v}| + |\mathbf{w}|. \quad (5.4)$$

For a real scalar product, in analogy with the dot product, the Cauchy-Schwarz inequality allows the definition of the angle $\angle(\mathbf{v}, \mathbf{w}) \in [0, \pi]$ between two non-zero vectors \mathbf{v}, \mathbf{w} by

$$\cos(\angle(\mathbf{v}, \mathbf{w})) \equiv \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{|\mathbf{v}| |\mathbf{w}|}. \quad (5.5)$$

For any scalar product, the two vectors \mathbf{v} and \mathbf{w} are called orthogonal if and only if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

For example, the standard scalar product in \mathbb{C}^n is defined as $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^\dagger \mathbf{w} = \sum_{i=1}^n v_i^* w_i$ and the Minkowski product in \mathbb{R}^4 is defined as $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \eta \mathbf{w} = -v_0 w_0 + v_1 w_1 + v_2 w_2 + v_3 w_3$ where $\eta = \text{diag}(-1, 1, 1, 1)$. Note that the Minkowski product thus not satisfies property (K3), for example, for $\mathbf{v} = (-1 \ 1 \ 1 \ 1)^T$, $\langle \mathbf{v}, \mathbf{v} \rangle = -1$. These kind of scalar products are called a bi-linear form (or a sesqui-linear form in the complex case where $F = \mathbb{C}$).

Since definition 5.1.1 can also be used to define scalar products in arbitrary vector spaces. Consider the vector space of (real or complex-valued) functions $f : [a, b] \rightarrow \mathbb{R}$ or \mathbb{C} on an interval $[a, b] \subset \mathbb{R}$. A scalar product for such functions can be defined by

$$\langle f, g \rangle \equiv \int_a^b dx f(x)^* g(x). \quad (5.6)$$

A scalar product can also be defined for real matrices, since the real $n \times n$ matrices form a vector space V , where

$$\langle A, B \rangle \equiv \text{Tr}(A^T B) = \sum_{i,j} A_{ij} B_{ij}. \quad (5.7)$$

For complex matrices a scalar product can be defined analogously simply by replacing the transpose with a hermitian conjugate.

5.2 Orthonormal Basis, Gram-Schmidt Procedure

An important observation is the following lemma:

Lemma 5.2.1. *Pairwise orthogonal and non-zero vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent.*

Proof. Start with

$$\sum_{i=1}^k \alpha_i \mathbf{v}_i = 0 \quad (5.8)$$

and take the scalar product of this equation with one of the vectors, say \mathbf{v}_j . Since $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \quad \forall i \neq j$ it follows that $\alpha_j |\mathbf{v}_j|^2 = 0$. Since $\mathbf{v}_j \neq 0$ its norm is positive, $|\mathbf{v}_j| > 0$, so $\alpha_j = 0$.

□

This motivates the following definition:

Definition 5.2.1. A basis $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$ of a vector space V with a scalar product is called *ortho-normal* if and only if

$$\langle \boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j \rangle = 0, \quad (5.9)$$

i.e. if the basis vectors are pairwise orthogonal and have length one.

For example, the basis of standard unit vectors, $\mathbf{e}_1, \dots, \mathbf{e}_n$ of $\mathbb{R}^n(\mathbb{C}^n)$ is an ortho-normal basis with respect to the standard scalar product on $\mathbb{R}^n(\mathbb{C}^n)$.

The vectors

$$\boldsymbol{\epsilon}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \boldsymbol{\epsilon}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad (5.10)$$

also form an orthonormal basis on \mathbb{R}^2 with respect to the standard scalar product, *i.e.* $\boldsymbol{\epsilon}_i^T \boldsymbol{\epsilon}_j = \delta_{ij}$.

The vectors

$$\boldsymbol{\epsilon}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ i \end{pmatrix} \text{ and } \boldsymbol{\epsilon}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2i \end{pmatrix} \quad (5.11)$$

on the other hand, form an orthonormal basis on \mathbb{C}^2 with respect to the standard scalar product, *i.e.* $\boldsymbol{\epsilon}_i^\dagger \boldsymbol{\epsilon}_j = \delta_{ij}$.

For the vector space of real $n \times n$ matrices, the matrices $E_{(ij)}$ defined in eq. (1.26) form an ortho-normal basis with respect to the scalar product defined in eq. (5.7)

An ortho-normal basis has many advantages compared to an arbitrary basis of a vector space. For example, consider the coordinates of a vector $\mathbf{v} \in V$ relative to an ortho-normal basis $\{\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n\}$. Of course, we can write \mathbf{v} as a linear combination $\mathbf{v} = \sum_{i=1}^n \alpha_i \boldsymbol{\epsilon}_i$ with some coordinates α_i , but, in the general case, these coefficients need to be determined by solving a system of linear equations. For an ortho-normal basis, we can just take the scalar product of this equation with $\boldsymbol{\epsilon}_j$, leading to

$$\langle \boldsymbol{\epsilon}_j, \mathbf{v} \rangle = \left\langle \boldsymbol{\epsilon}_j, \sum_{i=1}^n \alpha_i \boldsymbol{\epsilon}_i \right\rangle = \sum_{i=1}^n \alpha_i \langle \boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_i \rangle = \sum_{i=1}^n \alpha_i \delta_{ij} = \alpha_j \quad (5.12)$$

So in summary, the coordinates of a vector \mathbf{v} relative to an ortho-normal basis $\{\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n\}$ can be computed as

$$\mathbf{v} = \sum_{i=1}^n \alpha_i \boldsymbol{\epsilon}_i \iff \alpha_i = \langle \boldsymbol{\epsilon}_i, \mathbf{v} \rangle. \quad (5.13)$$

We now introduce the Gram-Schmidt procedure, which states and tells us which transform any basis of a vector space into an ortho-normal basis:

Theorem 5.2.2 (Gram-Schmidt Procedure). *If $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis of vector space V , then there exists an ortho-normal basis $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$ of V such that $\text{Span}(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_k) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k) \quad \forall k = 1, \dots, n$.*

Proof. The proof is constructive. The first vector of our prospective ortho-normal basis is obtained by simply normalizing \mathbf{v}_1 , *i.e.*

$$\boldsymbol{\epsilon}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}. \quad (5.14)$$

Clearly, $|\boldsymbol{\epsilon}_1| = 1$ and $\text{Span}(\boldsymbol{\epsilon}_1) = \text{Span}(\mathbf{v}_1)$. Suppose we have already constructed the first $k - 1$ vectors $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{k-1}$, mutually orthogonal, normalized and such that $\text{Span}(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_j) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_j)$ for all $j = 1, \dots, k - 1$. The next vector, $\boldsymbol{\epsilon}_k$, is then constructed by first subtracting from \mathbf{v}_k its projections onto $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{k-1}$ and then normalizing, so

$$\mathbf{v}'_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{\epsilon}_i, \mathbf{v}_k \rangle \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_k = \frac{\mathbf{v}'_k}{|\mathbf{v}'_k|}. \quad (5.15)$$

Obviously, $|\boldsymbol{\epsilon}_k| = 1$ and for any vector $\boldsymbol{\epsilon}_j$ with $j < k$ we have

$$\langle \boldsymbol{\epsilon}_j, \mathbf{v}'_k \rangle = \langle \boldsymbol{\epsilon}_j, \mathbf{v}_k \rangle - \sum_{i=1}^{k-1} \langle \boldsymbol{\epsilon}_i, \mathbf{v}_k \rangle \langle \boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_i \rangle = \langle \boldsymbol{\epsilon}_j, \mathbf{v}_k \rangle - \langle \boldsymbol{\epsilon}_j, \mathbf{v}_k \rangle = 0. \quad (5.16)$$

Hence, $\boldsymbol{\epsilon}_k$ is orthogonal to all vectors $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{k-1}$. Moreover, since $\text{Span}(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{k-1}) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_{k-1})$ and \mathbf{v}_k and $\boldsymbol{\epsilon}_k$ only differ by a re-scaling and terms proportional to $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{k-1}$, it follows that $\text{Span}(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_k) = \text{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$. □

A useful application of the Gram-Schmidt procedure is to consider the vector space of quadratic polynomials in one variable $x \in [1, -1]$ with real coefficients and a scalar product defined by $\langle f, g \rangle = \int_{-1}^1 f(x)g(x)dx$

We start with the standard monomial basis $v_1 = 1, v_2 = x, v_3 = x^2$ (which is not ortho-normal since, say, $\int_{-1}^1 (1)(x)dx \neq 0$). To find $\boldsymbol{\epsilon}_1$. To find an ortho-normal basis, we apply the Gram-Schmidt procedure

1. To find $\boldsymbol{\epsilon}_1$:

$$\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \int_{-1}^1 dx = 2, \quad \boldsymbol{\epsilon}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \frac{1}{\sqrt{2}}. \quad (5.17)$$

2. To find ϵ_2 first compute \mathbf{v}'_2

$$\langle \epsilon_1, \mathbf{v}_2 \rangle = \int_{-1}^1 dx \frac{x}{\sqrt{2}} = 0, \quad \mathbf{v}'_2 = \mathbf{v}_2 - \langle \epsilon_1, \mathbf{v}_2 \rangle \epsilon_1 = x, \quad (5.18)$$

and then normalize

$$\langle \mathbf{v}'_2, \mathbf{v}'_2 \rangle = \int_{-1}^1 dx x^2 = \frac{2}{3}, \quad \epsilon_2 = \frac{\mathbf{v}'_2}{|\mathbf{v}'_2|} = \sqrt{\frac{3}{2}} x. \quad (5.19)$$

3. To find ϵ_3 first compute \mathbf{v}'_3

$$\begin{aligned} \langle \epsilon_1, \mathbf{v}_3 \rangle &= \frac{1}{\sqrt{2}} \int_{-1}^1 dx x^2 = \frac{\sqrt{2}}{3}, \quad \langle \epsilon_2, \mathbf{v}_3 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^1 dx x^3 = 0, \\ \mathbf{v}'_3 &= \mathbf{v}_3 - \langle \epsilon_1, \mathbf{v}_3 \rangle \epsilon_1 - \langle \epsilon_2, \mathbf{v}_3 \rangle \epsilon_2 = x^2 - \frac{1}{3} \end{aligned} \quad (5.20)$$

and normalize

$$\langle \mathbf{v}'_3, \mathbf{v}'_3 \rangle = \int_{-1}^1 dx \left(x^2 - \frac{1}{3} \right)^2 = \frac{8}{45}, \quad \epsilon_3 = \frac{\mathbf{v}'_3}{|\mathbf{v}'_3|} = \sqrt{\frac{5}{8}} (3x^2 - 1). \quad (5.21)$$

So, in summary, the ortho-normal polynomial basis is

$$\epsilon_1 = \frac{1}{\sqrt{2}}, \quad \epsilon_2 = \sqrt{\frac{3}{2}} x, \quad \epsilon_3 = \sqrt{\frac{5}{8}} (3x^2 - 1). \quad (5.22)$$

These are the first three of an infinite family of ortho-normal polynomials, referred to as Legendre polynomials, which play an important role in mathematical physics.

To see how an ortho-normal basis helps simplify things, we start with two vectors

$$\mathbf{v} = \sum_i \alpha_i \epsilon_i, \quad \alpha_i = \langle \epsilon_i, \mathbf{v} \rangle \quad (5.23)$$

$$\mathbf{w} = \sum_i \beta_i \epsilon_i, \quad \beta_i = \langle \epsilon_i, \mathbf{w} \rangle \quad (5.24)$$

and compute their scalar product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i,j} \alpha_i^* \beta_j \langle \epsilon_i, \epsilon_j \rangle = \sum_i \alpha_i^* \beta_i = \sum_i \langle \mathbf{v}, \epsilon_i \rangle \langle \epsilon_i, \mathbf{w} \rangle. \quad (5.25)$$

Suppose we would like to compute the representing matrix A of a linear map $f : V \rightarrow V$ relative to an ortho-normal basis $\{\epsilon_1, \dots, \epsilon_n\}$ of V . In general, following lemma 3.4.1, the entries A_{ij} of the matrix A can be obtained from

$$f(\epsilon_j) = \sum_i A_{ij} \epsilon_i. \quad (5.26)$$

Taking the scalar product of this equation with ϵ_k results in the simple formula

$$A_{ij} = \langle \epsilon_i, f(\epsilon_j) \rangle. \quad (5.27)$$

In physics, the RHS of this expression is often referred to as a matrix element of the map f . It is worth noting that a linear map is uniquely determined by its matrix elements.

Lemma 5.2.3. *If two linear maps $f : V \rightarrow V$ and $g : V \rightarrow V$ satisfy $\langle \mathbf{v}, f(\mathbf{w}) \rangle = \langle \mathbf{v}, g(\mathbf{w}) \rangle$ (or $\langle f(\mathbf{v}), \mathbf{w} \rangle = \langle g(\mathbf{v}), \mathbf{w} \rangle$) for all $\mathbf{v}, \mathbf{w} \in V$ then $f = g$.*

Proof. By linearity of the scalar product in the second argument the assumption implies that $\langle \mathbf{v}, f(\mathbf{w}) - g(\mathbf{w}) \rangle = 0$ for all $\mathbf{v}, \mathbf{w} \in V$. In particular, if we choose $\mathbf{v} = f(\mathbf{w}) - g(\mathbf{w})$, it follows that $f(\mathbf{w}) - g(\mathbf{w}) = 0$. Since this holds for all \mathbf{w} it follows that $f = g$. The alternative statement follows simply by applying (K1) in definition 5.1.1 \square

Example: AL Example 6.5

Question: For a fixed vector $\mathbf{n} \in \mathbb{R}^3$, we consider the linear map $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined by

$$f(\mathbf{v}) = (\mathbf{n} \cdot \mathbf{v})\mathbf{n}. \quad (5.28)$$

Evidently, this map projects vectors into the direction of \mathbf{n} . Compute the matrix A representing this linear map relative to the ortho-normal basis given by the three standard unit vectors \mathbf{e}_i .

Solution: Using eq. (5.27), we have

$$A_{ij} = \mathbf{e}_i \cdot f(\mathbf{e}_j) = (\mathbf{n} \cdot \mathbf{e}_i)(\mathbf{n} \cdot \mathbf{e}_j) = n_i n_j. \quad (5.29)$$

and, hence, $A_{ij} = n_i n_j$ or, in matrix notation

$$A = \begin{pmatrix} n_1^2 & n_1 n_2 & n_1 n_3 \\ n_1 n_2 & n_2^2 & n_2 n_3 \\ n_1 n_3 & n_2 n_3 & n_3^2 \end{pmatrix} \quad (5.30)$$

We end this section by introducing perpendicular spaces, which is defined as

$$W^\perp = \{ \mathbf{v} \in V \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0 \quad \forall \mathbf{w} \in W \}. \quad (5.31)$$

In other words, W^\perp consists of all vectors which are orthogonal to all vector in W . For example, if $W \subset \mathbb{R}^3$, is a plane through the origin then W^\perp is the line through the origin perpendicular to this plane. The following statements are intuitive and will be helpful for our treatment of eigenvectors and eigenvalues in the next section.

Lemma 5.2.4. *For a sub vector space $W \subset V$ of a finite dimensional vector space V with a scalar product, the following holds:*

(L1) W^\perp is a sub vector space of V .

$$(L2) \quad W \cap W^\perp = \{0\}$$

$$(L3) \quad \dim(W) + \dim(W^\perp) = \dim(V)$$

Proof. (L1) If $\mathbf{v}_1, \mathbf{v}_2 \in W^\perp$ then clearly $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 \in W^\perp$ so from Def. 1.2, W^\perp is a sub vector space.

(L2) If $\mathbf{v} \in W \cap W^\perp$ then $\langle \mathbf{v}, \mathbf{v} \rangle = 0$, but from (K3) in definition 5.1.1 this implies that $\mathbf{v} = 0$.

(L3) Choose an ortho-normal basis $\{\epsilon_1, \dots, \epsilon_k\}$ of W and define the linear map $f : V \rightarrow V$ by $f(\mathbf{v}) = \sum_{i=1}^k \langle \epsilon_i, \mathbf{v} \rangle \epsilon_i$ (a projection onto W). Clearly $\text{Im}(f) \subset W$. For $\mathbf{w} \in W$ it follows that $f(\mathbf{w}) = \mathbf{w}$ so that $\text{Im}(f) = W$. Moreover, $\text{Ker}(f) = W^\perp$ and the claim follows from eq. (3.12) applied to the map f .

□

5.3 Adjoint Linear Maps

The definition of adjoint linear map is:

Definition 5.3.1. For a linear map $f : V \rightarrow V$ on a vector space V with scalar product, an adjoint linear map, $f^\dagger : V \rightarrow V$ is a map satisfying

$$\langle \mathbf{v}, f\mathbf{w} \rangle = \langle f^\dagger \mathbf{v}, \mathbf{w} \rangle \quad (5.32)$$

for all $\mathbf{v}, \mathbf{w} \in V$.

In other words, a linear map can be “moved” into the other argument of the scalar product by taking its adjoint. The following properties of the adjoint map are relatively easy to show.

Lemma 5.3.1 (Properties of the adjoint). For the adjoint f^\dagger of the linear map f ,

(M1) f^\dagger is uniquely determined.

(M2) $(f^\dagger)^\dagger = f$

(M3) $(f + g)^\dagger = f^\dagger + g^\dagger$

(M4) $(\alpha f)^\dagger = \alpha^* f^\dagger$

(M5) $(f \circ g)^\dagger = g^\dagger \circ f^\dagger$

(M6) $(f^{-1})^\dagger = (f^\dagger)^{-1}$, if f is invertible.

Proof. (M1) For two adjoints f_1, f_2 for f we have $\langle f_1(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, f(\mathbf{w}) \rangle = \langle f_2(\mathbf{v}), \mathbf{w} \rangle$ for all $\mathbf{v}, \mathbf{w} \in V$. Then lemma 5.2.3 implies that $f_1 = f_2$.

(M2) $\langle (f^\dagger)^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, f^\dagger(\mathbf{w}) \rangle = \langle f(\mathbf{v}), \mathbf{w} \rangle$. Comparing the LHS and RHS together with lemma 5.2.3 shows that $(f^\dagger)^\dagger = f$.

(M3) $\langle (f+g)^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, (f+g)(\mathbf{w}) \rangle = \langle \mathbf{v}, f(\mathbf{w}) \rangle + \langle \mathbf{v}, g(\mathbf{w}) \rangle = \langle f^\dagger(\mathbf{v}), \mathbf{w} \rangle + \langle g^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle (f^\dagger + g^\dagger)(\mathbf{v}), \mathbf{w} \rangle$ and the claim follows from lemma 5.2.3.

(M4) $\langle (\alpha f)^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, (\alpha f)(\mathbf{w}) \rangle = \alpha \langle \mathbf{v}, f(\mathbf{w}) \rangle = \alpha \langle f^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle (\alpha^* f^\dagger)(\mathbf{v}), \mathbf{w} \rangle$ and lemma 5.2.3 leads to the stated result.

(M5) $\langle (f \circ g)^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, (f \circ g)(\mathbf{w}) \rangle = \langle f^\dagger(\mathbf{v}), g(\mathbf{w}) \rangle = \langle g^\dagger \circ f^\dagger(\mathbf{v}), \mathbf{w} \rangle$.

(M6) From (v) we have $\text{id}_V = (f \circ f^{-1})^\dagger = f^\dagger \circ (f^{-1})^\dagger$. This means $(f^{-1})^\dagger$ is the inverse of f^\dagger and, hence, $(f^\dagger)^{-1} = (f^{-1})^\dagger$.

□

Let us now consider two matrices A and B describing f and f^\dagger relative to an ortho-normal basis $\epsilon_1, \dots, \epsilon_n$ of V , so A and B are given by

$$A_{ij} = \langle \epsilon_i, f(\epsilon_j) \rangle \text{ and } B_{ij} = \langle \epsilon_i, f^\dagger(\epsilon_j) \rangle. \quad (5.33)$$

These matrices are related by

$$B_{ij} = \langle \epsilon_i, f^\dagger(\epsilon_j) \rangle = \langle f^\dagger(\epsilon_j), \epsilon_i \rangle^* = \langle \epsilon_j, f(\epsilon_i) \rangle^* = A_{ji}^* = A_{ij}^\dagger. \quad (5.34)$$

Therefore, if A represents f then the hermitian conjugate A^\dagger represents f^\dagger . Also this show that by reversing the above argument and defining f^\dagger as the linear map associated to A^\dagger , the adjoint of a linear map always exist, which was non-trivial from the definition.

We can verify this in the case of the standard scalar product in \mathbb{R}^n or \mathbb{C}^n by

$$\langle \mathbf{v}, A\mathbf{w} \rangle = \mathbf{v}^\dagger A\mathbf{w} = (A^\dagger \mathbf{v})^\dagger \mathbf{w} = \langle A^\dagger \mathbf{v}, \mathbf{w} \rangle. \quad (5.35)$$

A particularly important class of linear maps are those which coincide with their adjoint.

Definition 5.3.2. *A linear map $f : V \rightarrow V$ on a vector space V with scalar product is called self-adjoint(or hermitian) if and only if $f = f^\dagger$.*

For a more abstract example of a self-adjoint linear map, consider the vector space of (infinitely many times) differentiable functions $\varphi : [a, b] \rightarrow \mathbb{C}$, satisfying $\varphi(a) = \varphi(b)$, with scalar product

$$\langle \varphi, \psi \rangle = \int_a^b dx \varphi(x)^* \psi(x). \quad (5.36)$$

The derivative operator

$$D = -i \frac{d}{dx} \quad (5.37)$$

defines a linear map on this space, and we would like to check that it is self-adjoint. Performing an integration by parts, we find

$$\begin{aligned}
\langle \varphi, D\psi \rangle &= -i \int_a^b dx \varphi(x)^* \frac{d\psi}{dx}(x) = -i [\varphi(x)^* \psi(x)]_a^b + i \int_a^b dx \frac{d\varphi}{dx}(x)^* \psi(x) \\
&= \int_a^b dx, (D\varphi)(x)^* \psi(x) = \langle D\varphi, \psi \rangle.
\end{aligned} \tag{5.38}$$

Hence, D is indeed Hermitian. Note that the boundary term vanishes due to the boundary condition on our functions and that including the factor of i in the definition of D is crucial for the sign to work out correctly. In quantum mechanics, physical quantities are represented by Hermitian operators. In this context, the present operator D plays an important role as it corresponds to linear momentum.

5.4 Unitary Maps

We define a unitary (also known as orthogonal) map as follows:

Definition 5.4.1. *Let V be a vector space with a real (hermitian) scalar product. A linear map $f : V \rightarrow V$ is called orthogonal (unitary) if and only if*

$$\langle f(\mathbf{v}), f(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle \quad \forall \mathbf{v}, \mathbf{w} \in V. \tag{5.39}$$

Lemma 5.4.1 (Properties of Unitary Maps). *For any unitary map f ,*

(N1) *f can also be characterized by $f^\dagger \circ f = \text{id}_V$*

(N2) *f are invertible and $f^{-1} = f^\dagger$.*

(N3) *The composition of unitary maps is a unitary map.*

(N4) *The inverse, f^\dagger , of f is unitary.*

Proof. (N1) Using the adjoint map, the condition eq. (5.39) can be re-written as $\langle \mathbf{v}, f^\dagger \circ f(\mathbf{w}) \rangle = \langle \mathbf{v}, \text{id}_V(\mathbf{w}) \rangle$. From lemma 5.3.1 a function is uniquely determined by its matrix elements so that orthogonal and unitary operators can also be defined by the condition $f^\dagger \circ f = \text{id}_V$.

(N2) A direct consequence of (N1).

(N3) For two unitary maps f, g , satisfying $\langle f(\mathbf{v}), f(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle g(\mathbf{v}), g(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$, it follows that $\langle f \circ g(\mathbf{v}), f \circ g(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$ and hence, that $f \circ g$ is unitary.

(N4) From $\langle f(\mathbf{v}), f(\mathbf{w}) \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$, writing $\mathbf{v}' = f(\mathbf{v}), \mathbf{w}' = f(\mathbf{w})$ it follows that $\langle \mathbf{v}', \mathbf{w}' \rangle = \langle f^{-1}(\mathbf{v}'), f^{-1}(\mathbf{w}') \rangle$ so that $f^{-1} = f^\dagger$ is unitary.

□

We begin with the real case where the condition (N1) turns into

$$A^T A = \mathbb{I} \quad \Longleftrightarrow \quad A^{-1} = A^T \quad \Longleftrightarrow \quad \mathbf{A}^i \cdot \mathbf{A}^j = \delta_{ij}, \tag{5.40}$$

where the last step can be shown by $(A^T A)_{ij} = A_{ki} A_{kj} = \mathbf{A}^i \cdot \mathbf{A}^j = \mathbb{I} = \delta_{ij}$. Matrices A satisfying this condition are called orthogonal matrices and they can be characterized,

equivalently, by either one of the three conditions above. The simplest way to check if a given matrix is orthogonal is usually to verify the condition on the LHS. The condition in the middle tells us it is easy to compute the inverse of an orthogonal matrix—it is simply the transpose. And, finally, the condition on the RHS says that the column vectors of an orthogonal matrix form an ortho-normal basis with respect to the standard scalar product (the dot product). In fact, since a real scalar product, written in terms of an ortho-normal basis, corresponds to the dot product, we expect that orthogonal matrices are precisely those matrices which leave the dot product invariant. Indeed, we have

$$A^T A = \mathbb{I}_n \iff (A\mathbf{v})^T (A\mathbf{w}) = \mathbf{v}^T \mathbf{w} \quad \text{for all } \mathbf{v}, \mathbf{w} \in \mathbb{R}^n. \quad (5.41)$$

The set of all $n \times n$ orthogonal matrices is also denoted by $O(n)$. Taking the determinant of the LHS condition in eq. (5.40) gives $1 = \det(\mathbb{I}) = \det(AA^T) = \det(A) \det(A^T) = \det(A)^2$ so that

$$\det(A) = \pm 1 \quad (5.42)$$

for any orthogonal matrix. The subset of orthogonal matrices A with determinant $\det(A) = +1$ is called “special orthogonal matrices or rotations” and denoted by $SO(n)$. Since the matrices leave the dot product invariant they do not change the lengths of vectors and angles between vectors and the positive sign condition excludes orthogonal matrices which contain reflections, they describe rotations of a body. The relation between orthogonal matrices with positive and negative determinants is easy to understand. Consider an orthogonal matrix A with $\det(A) = -1$ and the orthogonal matrix $F = \text{diag}(-1, 1, \dots, 1)$ with $\det(F) = -1$ which corresponds to a reflection in the first coordinate direction. Then the matrix $R = FA$ is a rotation since $\det(R) = \det(F) \det(A) = 1$. This means every orthogonal matrix A can be written as a product

$$A = FR \quad (5.43)$$

of a rotation R and a reflection F .

The most prominent example of the applications of orthogonal matrices is to describe rotations in two and three dimensions.

To find the explicit form of two-dimensional rotation matrices, start with a general 2×2 matrix:

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad (5.44)$$

where a, b, c, d are real numbers. We impose the conditions $R^T R = \mathbb{I}_2$ and $\det(R) = 1$. This gives:

$$R^T R = \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \det(R) = ad - bc = 1. \quad (5.45)$$

A solution to the equations can be written as $a = d = \cos(\theta)$, $c = -b = \sin(\theta)$ for some angle θ . Thus, the two-dimensional rotation matrices can be written in the form:

$$R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}. \quad (5.46)$$

For the rotation of an arbitrary vector $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, we get:

$$\mathbf{x}' = R\mathbf{x} = \begin{pmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{pmatrix}. \quad (5.47)$$

It is easy to verify explicitly that $|\mathbf{x}'| = |\mathbf{x}|$, as must be the case, and that the cosine of the angle between \mathbf{x} and \mathbf{x}' is given by

$$\cos(\langle \mathbf{x}', \mathbf{x} \rangle) = \frac{\mathbf{x}' \cdot \mathbf{x}}{|\mathbf{x}'||\mathbf{x}|} = \frac{(x \cos \theta - y \sin \theta)x + (x \sin \theta + y \cos \theta)y}{|\mathbf{x}|^2} = \cos \theta. \quad (5.48)$$

This result means we should interpret $R(\theta)$ as a rotation by an angle θ . From the addition theorems of sin and cos it also follows easily that

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2), \quad (5.49)$$

i.e. the rotation angle adds up for subsequent rotations, as one would expect. Note, the above equation also implies that two-dimensional rotations commute, since $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2) = R(\theta_2 + \theta_1) = R(\theta_2)R(\theta_1)$, again a property intuitively expected.

To find the explicit form for three-dimensional rotations we could, in principle, use the same approach as in two dimensions and impose all relevant constraints on an arbitrary 3×3 matrix. However, this leads to a set of equations in 9 variables and is much more complicated. However, it is easy to obtain special three-dimensional rotations from two-dimensional ones. For example, the matrices

$$R_1(\theta_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ 0 & \sin \theta_1 & \cos \theta_1 \end{pmatrix} \quad (5.50)$$

clearly satisfy $R_1(\theta_1)^T R_1(\theta_1) = \mathbb{I}_3$ and $\det(R_1(\theta_1)) = 1$ and are, hence, rotation matrices. They describe a rotation by an angle θ_1 around the first coordinate axis. Analogously, rotation matrices around the other two coordinate axes can be written as

$$R_2(\theta_2) = \begin{pmatrix} \cos \theta_2 & 0 & \sin \theta_2 \\ 0 & 1 & 0 \\ -\sin \theta_2 & 0 & \cos \theta_2 \end{pmatrix}^1, \quad R_3(\theta_3) = \begin{pmatrix} \cos \theta_3 & -\sin \theta_3 & 0 \\ \sin \theta_3 & \cos \theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.51)$$

It turns out that general three-dimensional rotation matrices can be obtained as products of the above three special types. For example, we can write a three-dimensional rotation matrix as $R(\theta_1, \theta_2, \theta_3) = R_1(\theta_1)R_2(\theta_2)R_3(\theta_3)$, *i.e.* as subsequent rotations around the three coordinate axes. Of course, there are different ways of doing this, another choice

¹The odd signs can be easily justified since the columns vectors are the images of $\mathbf{e}_1, \mathbf{e}_2$ and \mathbf{e}_3 .

frequently used in physics being $R(\psi, \theta, \phi) = R_3(\psi)R_1(\theta)R_3(\phi)$. The angles ψ, θ, ϕ in this parametrization are also called the *Euler angles* and in this case, the rotation is combined from a rotation by ϕ around the z -axis, then a rotation by θ around the x -axis and finally another rotation by ψ around the (new) z -axis. The Euler angles are particularly useful to describe the motion of tops in classical mechanics.

Finally, we note that, unlike their two-dimensional counterparts, three-dimensional rotations do not, in general, commute. For example, apart from special choices for the angles $R_1(\theta_1)R_2(\theta_2) \neq R_2(\theta_2)R_1(\theta_1)$.

Suppose we have a stationary coordinate system with coordinates $\mathbf{x} \in \mathbb{R}^3$ and another coordinate system with coordinates $\mathbf{y} \in \mathbb{R}^3$, which is rotating relative to the first one. Such a set-up can be used to describe the mechanics of objects in rotating systems and has many applications, for example to the physics of tops or the laws of motion in rotating systems such as the earth (see below). Mathematically, the relation between these two coordinate systems can be described by the equation

$$\mathbf{x} = R(t)\mathbf{y}, \quad (5.52)$$

where $R(t)$ are time-dependent rotation matrices. This means the matrices $R(t)$ satisfy

$$R(t)^T R(t) = \mathbb{I}_3, \quad (5.53)$$

(as well as $\det(R(t)) = 1$) for all times t . In practice, we can write rotation matrices in terms of rotation angles. The time-dependence of $R(t)$ then means that the rotation angles are functions of time. For example, a rotation around the z -axis with constant angular speed ω can be written as

$$R(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) & 0 \\ \sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (5.54)$$

In physics, a rotation is often described by the angular velocity $\boldsymbol{\omega}$, a vector whose direction indicates the axis of rotation and whose length gives the angular speed. It is very useful to understand the relation between $R(t)$ and $\boldsymbol{\omega}$. To do this, define the matrix

$$W = R^T \dot{R}, \quad (5.55)$$

where the dot denotes the time derivative and observe, by differentiating the above equation with respect to time, that

$$R^T \dot{R} + \dot{R}^T R = 0. \quad (5.56)$$

Hence, W is an anti-symmetric matrix and can be written in the form

$$W = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \quad \text{or} \quad W_{ij} = \epsilon_{ikj} \omega_k. \quad (5.57)$$

The three independent entries ω_i of this matrix define the angular velocity $\boldsymbol{\omega} = (\omega_1 \ \omega_2 \ \omega_3)^T$. To see that this makes sense let us work out the matrix W for $R(t)$.

$$W = \omega \begin{pmatrix} \cos(\omega t) & \sin(\omega t) & 0 \\ -\sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\sin(\omega t) & -\cos(\omega t) & 0 \\ \cos(\omega t) & -\sin(\omega t) & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (5.58)$$

Comparison with the general form of W then shows that the angular velocity for this case is given by $\boldsymbol{\omega} = (0, 0, \omega)$, indicating a rotation with angular speed ω around the z -axis, as expected.

In the previous sections we have shown that the multiplication of an anti-symmetric 3×3 matrix with a vector can be written as a cross-product, so that

$$W\mathbf{b} = \boldsymbol{\omega} \times \mathbf{b} \quad (5.59)$$

for any vector $\mathbf{b} = (b_1 \ b_2 \ b_3)^T$. This can also be directly verified using the matrix form of W together with the definition of the cross product or, more elegantly, by the index calculation $W_{ij}b_j = \epsilon_{ikj}\omega_k b_j = (\boldsymbol{\omega} \times \mathbf{b})_i$. This relation can be used to re-write expressions involving W in terms of the angular velocity $\boldsymbol{\omega}$.

For a simple application of this formalism, consider an object moving with velocity $\dot{\mathbf{y}}$ relative to the rotating system. What is its velocity relative to the stationary coordinate system? Differentiating $\mathbf{x} = R(t)\mathbf{y}$ gives

$$\dot{\mathbf{x}} = R\dot{\mathbf{y}} + \dot{R}\mathbf{y} = R(\dot{\mathbf{y}} + W\mathbf{y}) = R(\dot{\mathbf{y}} + \boldsymbol{\omega} \times \mathbf{y}). \quad (5.60)$$

The velocity $\dot{\mathbf{x}}$ in the stationary system has, therefore, two contributions, namely the velocity $\dot{\mathbf{y}}$ relative to the rotating system and the velocity $\boldsymbol{\omega} \times \mathbf{y}$ due to the rotation itself.

We now investigate the Newton's second law in rotating frame.

We start with multiplying $\mathbf{F} = m\ddot{\mathbf{x}}$ on both sides by the rotation matrix $R(t)$ to get

$$\mathbf{F}_R = mR^T\ddot{\mathbf{x}}, \quad (5.61)$$

where \mathbf{F}_R is the force in the rotating coordinate system.

Taking the time derivatives of \mathbf{x} , we get

$$\dot{\mathbf{x}} = R\dot{\mathbf{y}} + \dot{R}\mathbf{y} \text{ and } \ddot{\mathbf{x}} = R\ddot{\mathbf{y}} + 2\dot{R}\dot{\mathbf{y}} + \ddot{R}\mathbf{y}. \quad (5.62)$$

Substituting $\ddot{\mathbf{x}}$, we have

$$m\ddot{\mathbf{y}} = \mathbf{F}_R - 2mR^T\dot{R}\dot{\mathbf{y}} - mR^T\ddot{R}\mathbf{y}. \quad (5.63)$$

The two extra terms can be simplified further by recalling the definition $W = R^T\dot{R}$ and thus $\dot{W} = R^T\ddot{R} + \dot{R}^T\dot{R} = R^T\ddot{R} + (\dot{R}^T R)(R^T \dot{R}) = R^T\ddot{R} - W^2$. With these results we have

$$m\ddot{\mathbf{y}} = \mathbf{F}_R - 2mW\dot{\mathbf{y}} - mW^2\mathbf{y} - m\dot{W}\mathbf{y} = \mathbf{F}_R - 2m\boldsymbol{\omega} \times \dot{\mathbf{y}} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{y}) - 2m\dot{\boldsymbol{\omega}} \times \mathbf{y}. \quad (5.64)$$

The three extra terms on the right are Coriolis force, centrifugal force and Euler force, respectively.

We now turn to the complex case, where the characterizing conditions becomes

$$A^\dagger A = 1 \iff A^{-1} = A^\dagger \iff (\mathbf{A}^i)^\dagger \mathbf{A}^j = \delta_{ij} \quad (5.65)$$

and can be verified in \mathbb{C}^n since

$$A^\dagger A = \mathbb{I}_n \iff (A\mathbf{v})^\dagger (A\mathbf{w}) = \mathbf{v}^\dagger \mathbf{w} \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{C}^n. \quad (5.66)$$

The set of all $n \times n$ unitary matrices is denoted by $U(n)$, where naturally $O(n) \subset U(n)$. From the determinant of unitary matrices we conclude that $1 = \det(\mathbb{I}) = \det(A^\dagger A) = \det(A)^* \det(A) = |\det(A)|^2$. Hence the determinant of unitary matrices has complex modulus of 1, so

$$|\det(A)| = 1. \quad (5.67)$$

The unitary matrices U with $\det(U) = 1$ are called “special unitary matrices” and denoted by $SU(n)$. Naturally, $SO(n) \subset SU(n)$.

For an arbitrary unitary $n \times n$ matrix A we can always find a complex phase ζ such that $\zeta^n = \det(A)$. Then, the matrix $U = \zeta^{-1}A$ is a special unitary since $\det(U) = \det(\zeta^{-1}A) = \zeta^{-n} \det(A) = 1$. This means every unitary matrix A can be written as a product

$$A = \zeta U \quad (5.68)$$

of a special unitary matrix U and a complex phase ζ .

We can find all two-dimensional special unitary matrices by using the same method as for two-dimensional rotations. We start with an arbitrary complex 2×2 matrix

$$U = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \quad (5.69)$$

where $\alpha, \beta, \gamma, \delta$ are complex numbers and impose the conditions $U^\dagger U = \mathbb{I}_2$ and $\det(U) = 1$. After a short calculation we find that every two-dimensional special unitary matrix can be written in terms of two complex numbers α, β as

$$U = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix} \quad \text{where} \quad |\alpha|^2 + |\beta|^2 = 1. \quad (5.70)$$

This shows that two-dimensional special unitary matrices depend on two complex parameters α, β subject to the (real) constraint $|\alpha|^2 + |\beta|^2 = 1$ and, hence, on three real

parameters. Inserting the special choice $\alpha = \cos \theta, \beta = -\sin \theta$ the above equation we recover the two-dimensional rotation matrices we found earlier in the real case, so that $SO(2) \subset SU(2)$, as expected from our general discussion.

Eigenvectors and Eigenvalues

Recall from lemma 3.4.1 that the matrix representing a linear map is computed by writing the images of the basis vectors as linear combinations of the basis, with the coefficients from each image forming a column of the matrix. Suppose that the image of a basis vector \mathbf{v} is simply a multiple of itself, so $f(\mathbf{v} = \lambda\mathbf{v})$ for some number λ . In this case, the corresponding column of the representing matrix only has one non-zero entry, λ , in the diagonal.

Definition 6.0.1 (Definition of Eigenvectors and Eigenvalues). *For a linear map $f : V \rightarrow V$ on a vector space V over F the number $\lambda \in F$ is called an eigen value of f if there is a non-zero vector \mathbf{v} such that*

$$f(\mathbf{v}) = \lambda\mathbf{v}. \quad (6.1)$$

In this case, \mathbf{v} is called an eigenvector of f with eigenvalue λ .

In short, an eigenvector is a vector which is just “scaled by the action of a linear map”.

To discuss this we first introduce the idea of eigenspaces. The eigenspace for $\lambda \in F$ is defined by

$$\text{Eig}_f(\lambda) = \ker(f - \lambda \text{id}_V), \quad (6.2)$$

which collects all the eigenvectors for λ .

Evidently λ is an eigenvalue of f precisely when $\dim \text{Eig}_f(\lambda) > 0$. If $\dim \text{Eig}_f(\lambda) = 1$ the eigenvalue λ is called non-degenerate, which means that up to re-scaling there is only one eigenvector for λ and if $\dim \text{Eig}_f(\lambda) > 1$ the eigenvalue λ is called degenerate, which means there are at least two linearly independent eigenvectors for λ .

We see that λ is an eigenvalue of f precisely when $\ker(f - \lambda \text{id}_V)$ is non-trivial. From ?? and eq. (3.12), they implies that $f - \text{id}_V$ is not bijective thus not invertible which is equivalent to

$$\chi_f(\lambda) \equiv \det(f - \lambda \text{id}_V) = 0, \quad (6.3)$$

where $\chi_f(\lambda)$ is called the characteristic polynomial of the linear map for the eigenvalue λ , which is a polynomial of order n in λ whose coefficients depend on f . Some general properties of the characteristic polynomial are given below:

Lemma 6.0.1 (Properties of a Characteristic Polynomial). *The characteristic polynomial $\chi_A(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \dots + c_1\lambda + c_0$ of an $n \times n$ matrix A has the following properties:*

(O1) $\chi_{PAP^{-1}} = \chi_A$, so the characteristic polynomial is basis-independent.

(O2) The coefficients c_i of the characteristic polynomial are basis-independent.

(O3) $c_n = (-1)^n$, $c_{n-1} = (-1)^{n-1} \sum_{i=1}^n A_{ii}$, $c_0 = \det(A)$.

Proof. (O1) $\chi_{PAP^{-1}}(\lambda) = \det(PAP^{-1} - \lambda\mathbb{I}) = \det(P(A - \lambda\mathbb{I})P^{-1}) = \det(P) \det(A - \lambda\mathbb{I}) \det(P^{-1}) = \det(A - \lambda\mathbb{I}) = \chi_A(\lambda)$.

(O2) This is a direct consequence of item (O1), since two polynomials are identical only if the coefficients of every power of x are equal, this is due to the fact that the vectors $1, x, x^2, \dots$ are linearly independent.

(O3) It is trivial for the c_0 case. For c_1 and c_2 , notice that the terms of order λ^n and λ^{n-1} only receive contributions from the product of the diagonal elements, so that

$$\chi_A(\lambda) = \prod_{i=1}^n (A_{ii} - \lambda) + \mathcal{O}(\lambda^{n-2}) = (-1)^n \lambda^n + (-1)^{n-1} \left(\sum_{i=1}^n A_{ii} \right) \lambda^{n-1} + \mathcal{O}(\lambda^{n-2}). \quad (6.4)$$

□

The above lemma shows that the determinant of a matrix is basis-independent, which we have already shown in section 4.2. Further, it shows that the trace of a matrix A

$$\text{Tr}(A) \equiv \sum_{i=1}^n A_{ii} \quad (6.5)$$

is also basis-independent. However this can be seen more directly by noting that $\text{Tr}(AB) = \sum_{i,j} A_{ij} B_{ji} = \sum_{i,j} B_{ji} A_{ij} = \text{Tr}(BA)$, so $\text{Tr}(PAP^{-1}) = \text{Tr}(P^{-1}PA) = \text{Tr}(A)$.

6.1 Diagonalization of matrices

In order to diagonalize a matrix, we start with

Definition 6.1.1. *We say a linear map $f : V \rightarrow V$ can be diagonalized if there exist a basis of V such that the matrix which describes f relative to this basis is diagonal. Further, we say an $n \times n$ matrix A with entries in F can be diagonalized if there is an invertible $n \times n$ matrix P with entries in F such that $\hat{A} \equiv P^{-1}AP$ is diagonal.*

The key statement relating eigenvectors and eigenvalues to diagonalization of a linear map is

Lemma 6.1.1. *A linear map $f : V \rightarrow V$ can be diagonalised iff there exists a basis of V consisting of eigenvectors of f . Relative to such a basis of eigenvectors, f is described by a diagonal matrix with the eigenvalues along the diagonal.*

Proof. The entries of the matrix A which describes f relative to a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of V are obtained from

$$f(\mathbf{v}_j) = \sum_i A_{ij} \mathbf{v}_i, \quad (6.6)$$

as discussed in lemma 3.4.1. From this equation, if $A = \text{diag}(\lambda_1, \dots, \lambda_n)$ is diagonal, then $f(\mathbf{v}_j) = \lambda_j \mathbf{v}_j$ and the basis vectors \mathbf{v}_j are eigenvectors with eigenvalues λ_i . \square

The analogous statement for matrices is

Lemma 6.1.2. *The $n \times n$ matrix A with entries in F can be diagonalized if and only if A has n eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ which form a basis of F^n . In this case, if we define the matrix*

$$P = (\mathbf{v}_1, \dots, \mathbf{v}_n) \quad (6.7)$$

whose columns are the eigenvectors of A , it follows that

$$P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (6.8)$$

where λ_i are the eigenvalues for \mathbf{v}_i .

Proof. We assume that we have a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of eigenvectors with eigenvalues λ_i so that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Define the matrix $P = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ whose columns are the eigenvectors of A . Since the eigenvectors form a basis of F^n , the matrix P is invertible. Then

$$\begin{aligned} P^{-1}AP &= P^{-1}A(\mathbf{v}_1, \dots, \mathbf{v}_n) = P^{-1}(A\mathbf{v}_1, \dots, A\mathbf{v}_n) = P^{-1}(\lambda_1 \mathbf{v}_1, \dots, \lambda_n \mathbf{v}_n) \\ &= P^{-1}(\mathbf{v}_1, \dots, \mathbf{v}_n) = \underbrace{P^{-1}(\mathbf{v}_1, \dots, \mathbf{v}_n)}_{=P} \text{diag}(\lambda_1, \dots, \lambda_n) = \text{diag}(\lambda_1, \dots, \lambda_n). \end{aligned} \quad (6.9)$$

\square

If a matrix A can be diagonalized, with eigenvalues $\lambda_1, \dots, \lambda_n$, so that $P^{-1}AP = \text{diag}(\lambda_1, \dots, \lambda_n)$, then the basis-independence of the determinant and the trace implies that

$$\det(A) = \prod_{i=1}^n \lambda_i \text{ and } \text{Tr}(A) = \sum_{i=1}^n \lambda_i. \quad (6.10)$$

While ?? provides a general criterion for a matrix to be diagonalizable it requires calculation of all the eigenvectors and checking whether they form a basis. It would be helpful to have a simpler condition, at least for some classes of matrices, which can simply be “read off” from the matrix. To this end we prove

Theorem 6.1.3. *Let V be a vector space over $\mathbb{R}(\mathbb{C})$ with real (hermitian) scalar product $\langle \cdot, \cdot \rangle$. If $f : V \rightarrow V$ is self-adjoint then*

(P1) All eigenvalues of f are real.

(P2) Eigenvectors for different eigenvalues are orthogonal.

Proof. (P1) For the real case, the first part of the statement is of course trivial. For the complex case, we start with an eigenvector $\mathbf{v} \neq 0$ of f with corresponding eigenvalue λ , so that

$$f(\mathbf{v}) = \lambda \mathbf{v}. \quad (6.11)$$

Then

$$\lambda \langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, \lambda \mathbf{v} \rangle = \langle \mathbf{v}, f(\mathbf{v}) \rangle = \langle f(\mathbf{v}), \mathbf{v} \rangle = \langle \lambda \mathbf{v}, \mathbf{v} \rangle = \lambda^* \langle \mathbf{v}, \mathbf{v} \rangle. \quad (6.12)$$

In the third step we have used the fact that f is self-adjoint and can, hence, be moved from one argument of the scalar product into the other. Since $\mathbf{v} \neq 0$ and, hence, $\langle \mathbf{v}, \mathbf{v} \rangle \neq 0$, it follows that $\lambda = \lambda^*$, so the eigenvalue is real.

(P2) Consider two eigenvectors $\mathbf{v}_1, \mathbf{v}_2$, so that

$$f(\mathbf{v}_1) = \lambda_1 \mathbf{v}_1 \quad \text{and} \quad f(\mathbf{v}_2) = \lambda_2 \mathbf{v}_2, \quad (6.13)$$

with different eigenvalues, $\lambda_1 \neq \lambda_2$. Then

$$(\lambda_1 - \lambda_2) \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \lambda_1 \mathbf{v}_1, \mathbf{v}_2 \rangle - \langle \mathbf{v}_1, \lambda_2 \mathbf{v}_2 \rangle = \langle f(\mathbf{v}_1), \mathbf{v}_2 \rangle - \langle \mathbf{v}_1, f(\mathbf{v}_2) \rangle = 0. \quad (6.14)$$

Since $\lambda_1 - \lambda_2 \neq 0$, this means $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$, and the two eigenvectors are orthogonal. \square

Theorem 6.1.4. *Let V be a vector space over \mathbb{C} with hermitian scalar product $\langle \cdot, \cdot \rangle$. If $f : V \rightarrow V$ is self-adjoint, it has an ortho-normal basis, $\epsilon_1, \dots, \epsilon_n$ of eigenvectors.*

Proof. The proof is by induction in n , the dimension of the vector space V . For $n = 1$, the assertion is trivial. Assume that it is true for all dimensions $k < n$. We would like to show that it is true for dimension n . The characteristic polynomial χ_f of f has at least one zero, λ , over the complex numbers. Since f is self-adjoint, λ is real from the previous theorem.

Consider the eigenspace $W = \text{Eig}_f(\lambda)$. Since λ is an eigenvalue, $\dim(W) > 0$. Vectors $\mathbf{v} \in W^\perp$ and $\mathbf{w} \in W$ are perpendicular, $\langle \mathbf{w}, \mathbf{v} \rangle = 0$, so

$$\langle \mathbf{w}, f(\mathbf{v}) \rangle = \langle f(\mathbf{w}), \mathbf{v} \rangle = \langle \lambda \mathbf{w}, \mathbf{v} \rangle = \lambda \langle \mathbf{w}, \mathbf{v} \rangle = 0. \quad (6.15)$$

This means that $f(\mathbf{v})$ is perpendicular to \mathbf{w} so that, whenever $\mathbf{v} \in W^\perp$, then also $f(\mathbf{v}) \in W^\perp$. As a result, W^\perp is invariant under f , and we can restrict f to W^\perp , i.e. consider $g = f|_{W^\perp}$. Since $\dim(W^\perp) < n$, there is an ortho-normal basis $\epsilon_1, \dots, \epsilon_k$ of W^\perp consisting of eigenvectors of g (which are also eigenvectors of f) by the induction assumption.

Add to this ortho-normal basis of W^\perp an ortho-normal basis of W (which, by definition of W , consists of eigenvectors of f with eigenvalue λ). Since $\dim(W) + \dim(W^\perp) = n$ (see Lemma 6.3), and pairwise orthogonal vectors are linearly independent, this list of vectors forms an ortho-normal basis of V , consisting of eigenvectors of f .

In summary, these results mean that every real symmetric (hermitian) matrix can be diagonalized, has an ortho-normal basis $\epsilon_1, \dots, \epsilon_n$ of eigenvectors with corresponding real eigenvalues $\lambda_1, \dots, \lambda_n$, and the diagonalizing matrix $P = (\epsilon_1, \dots, \epsilon_n)$ is orthogonal (unitary), such that

$$P^\dagger A P = \text{diag}(\lambda_1, \dots, \lambda_n). \quad (6.16)$$

□

Theorem 6.1.5. *Let V be a vector space over \mathbb{C} with hermitian scalar product $\langle \cdot, \cdot \rangle$. If $f : V \rightarrow V$ is self-adjoint, it has an ortho-normal basis, $\epsilon_1, \dots, \epsilon_n$ of eigenvectors.*

Proof. The proof is by induction in n , the dimension of the vector space V . For $n = 1$, the assertion is trivial. Assume that it is true for all dimensions $k < n$. We would like to show that it is true for dimension n . The characteristic polynomial χ_f of f has at least one zero, λ , over the complex numbers. Since f is self-adjoint, λ is real from the previous theorem.

Consider the eigenspace $W = \text{Eig}_f(\lambda)$. Since λ is an eigenvalue, $\dim(W) > 0$. Vectors $\mathbf{v} \in W^\perp$ and $\mathbf{w} \in W$ are perpendicular, $\langle \mathbf{w}, \mathbf{v} \rangle = 0$, so

$$\langle \mathbf{w}, f(\mathbf{v}) \rangle = \langle f(\mathbf{w}), \mathbf{v} \rangle = \langle \lambda \mathbf{w}, \mathbf{v} \rangle = \lambda \langle \mathbf{w}, \mathbf{v} \rangle = 0. \quad (6.17)$$

This means that $f(\mathbf{v})$ is perpendicular to \mathbf{w} so that, whenever $\mathbf{v} \in W^\perp$, then also $f(\mathbf{v}) \in W^\perp$. As a result, W^\perp is invariant under f , and we can restrict f to W^\perp , i.e. consider $g = f|_{W^\perp}$. Since $\dim(W^\perp) < n$, there is an ortho-normal basis $\epsilon_1, \dots, \epsilon_k$ of W^\perp consisting of eigenvectors of g (which are also eigenvectors of f) by the induction assumption.

Add to this ortho-normal basis of W^\perp an ortho-normal basis of W (which, by definition of W , consists of eigenvectors of f with eigenvalue λ). Since $\dim(W) + \dim(W^\perp) = n$ (see lemma 5.2.4), and pairwise orthogonal vectors are linearly independent, this list of vectors forms an ortho-normal basis of V , consisting of eigenvectors of f .

In summary, these results mean that every real symmetric (hermitian) matrix can be diagonalized, has an ortho-normal basis $\epsilon_1, \dots, \epsilon_n$ of eigenvectors with corresponding real eigenvalues $\lambda_1, \dots, \lambda_n$, and the diagonalizing matrix $P = (\epsilon_1, \dots, \epsilon_n)$ is orthogonal (unitary), such that

$$P^\dagger A P = \text{diag}(\lambda_1, \dots, \lambda_n). \quad (6.18)$$

□

In summary, these results mean that every real symmetric (hermitian) matrix can be diagonalized, has an ortho-normal basis $\epsilon_1, \dots, \epsilon_n$ of eigenvectors with corresponding real eigenvalues $\lambda_1, \dots, \lambda_n$ and the diagonalizing matrix $P = \text{lr}(\epsilon_1, \dots, \epsilon_n)$ is orthogonal (unitary), such that

$$P^\dagger A P = \text{diag}(\lambda_1, \dots, \lambda_n). \quad (6.19)$$

Rotational Transformations

Assume a point P has coordinates (x, y, z) and (x', y', z') with respect to the unprimed coordinate system and the primed coordinate system respectively (the origins of two coordinate system coincide) so we have

$$x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} = x'\hat{\mathbf{x}}' + y'\hat{\mathbf{y}}' + z'\hat{\mathbf{z}}'. \quad (7.1)$$

Dotting this equation with $\hat{\mathbf{x}}', \hat{\mathbf{y}}'$ and $\hat{\mathbf{z}}'$, we have

$$\begin{aligned} x' &= (\hat{\mathbf{x}}' \cdot \hat{\mathbf{x}})x + (\hat{\mathbf{x}}' \cdot \hat{\mathbf{y}})y + (\hat{\mathbf{x}}' \cdot \hat{\mathbf{z}})z \\ y' &= (\hat{\mathbf{y}}' \cdot \hat{\mathbf{x}})x + (\hat{\mathbf{y}}' \cdot \hat{\mathbf{y}})y + (\hat{\mathbf{y}}' \cdot \hat{\mathbf{z}})z \\ z' &= (\hat{\mathbf{z}}' \cdot \hat{\mathbf{x}})x + (\hat{\mathbf{z}}' \cdot \hat{\mathbf{y}})y + (\hat{\mathbf{z}}' \cdot \hat{\mathbf{z}})z. \end{aligned} \quad (7.2)$$

Writing in matrix form, we have

$$\mathbf{OP}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{x}}' \cdot \hat{\mathbf{x}} & \hat{\mathbf{x}}' \cdot \hat{\mathbf{y}} & \hat{\mathbf{x}}' \cdot \hat{\mathbf{z}} \\ \hat{\mathbf{y}}' \cdot \hat{\mathbf{x}} & \hat{\mathbf{y}}' \cdot \hat{\mathbf{y}} & \hat{\mathbf{y}}' \cdot \hat{\mathbf{z}} \\ \hat{\mathbf{z}}' \cdot \hat{\mathbf{x}} & \hat{\mathbf{z}}' \cdot \hat{\mathbf{y}} & \hat{\mathbf{z}}' \cdot \hat{\mathbf{z}} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \boldsymbol{\lambda}(\mathbf{OP}) \quad (7.3)$$

where $\boldsymbol{\lambda}$ is called the rotation matrix.

By the same fashion,

$$\mathbf{OP} = \boldsymbol{\lambda}^T(\mathbf{OP}'). \quad (7.4)$$

From these two equations, we have

$$\mathbf{OP} = \boldsymbol{\lambda}^T \boldsymbol{\lambda}(\mathbf{OP}'), \quad (7.5)$$

so

$$\boldsymbol{\lambda}^T \boldsymbol{\lambda} = \mathbf{I} \quad (7.6)$$

which means that the rotation matrix $\boldsymbol{\lambda}$ is orthogonal with $\boldsymbol{\lambda}^T = \boldsymbol{\lambda}^{-1}$.

However, only 3 of the 9 entries of $\boldsymbol{\lambda}$ are actually independent which implies that all rotations can be expressed in terms of rotations about three orthogonal axes since

$$\hat{\mathbf{x}}' \cdot \hat{\mathbf{x}} + \hat{\mathbf{x}} \cdot \hat{\mathbf{y}} + \hat{\mathbf{x}} \cdot \hat{\mathbf{z}} = 1 \text{ etc.} \quad (7.7)$$

by the Pythagorean Theorem which limits 3 degrees of freedom.

Also, the orthogonality of the coordinate systems requires

$$\hat{\mathbf{x}}' \cdot \hat{\mathbf{y}} = (\hat{\mathbf{x}}' \cdot \hat{\mathbf{x}})(\hat{\mathbf{y}}' \cdot \hat{\mathbf{x}}) + (\hat{\mathbf{x}}' \cdot \hat{\mathbf{y}})(\hat{\mathbf{y}}' \cdot \hat{\mathbf{y}}) + (\hat{\mathbf{x}}' \cdot \hat{\mathbf{z}})(\hat{\mathbf{y}}' \cdot \hat{\mathbf{z}}) = 0 \text{ etc.} \quad (7.8)$$

These six relations can be written in a nice and neat form

$$\sum_j ((\hat{\mathbf{i}}' \cdot \hat{\mathbf{j}})(\hat{\mathbf{k}}' \cdot \hat{\mathbf{j}})) = \delta_{ik}. \quad (7.9)$$

With this formulation, it is clear that finite rotation do not commute as

$$\boldsymbol{\lambda}_A \boldsymbol{\lambda}_B \neq \boldsymbol{\lambda}_B \boldsymbol{\lambda}_A. \quad (7.10)$$

as one can easily see by trying out random numbers.

Appendices

A.1 A Sub Vector Space is a Vector Space

Here we provide a rigorous proof that the sub vector space $W \subset V$ satisfy all the requirements in definition 1.2.1.

Lemma A.1.1. $0 \cdot \mathbf{v} = \mathbf{0} \quad \forall \mathbf{v} \in V$.

Proof. Since $0\mathbf{v} = (0 + 0)\mathbf{v} \stackrel{(A6)}{=} 0\mathbf{v} + 0\mathbf{v}$ and $\mathbf{0} + 0\mathbf{v} \stackrel{(A2)}{=} 0\mathbf{v}$ it follows that $0\mathbf{v} = \mathbf{0}$. \square

Lemma A.1.2. $(-1)\mathbf{v} = -\mathbf{v} \quad \forall \mathbf{v} \in V$.

Proof. Since $\mathbf{0} = 0\mathbf{v} = (1 + (-1))\mathbf{v} \stackrel{(A6, A8)}{=} \mathbf{v} + (-1)\mathbf{v}$ and $\mathbf{0} \stackrel{(A3)}{=} \mathbf{v} + (-\mathbf{v})$ it follows that $(-1)\mathbf{v} = -\mathbf{v}$. \square

Since $0\mathbf{w} = \mathbf{0}$ and $(-1)\mathbf{w} = -\mathbf{w}$, so from property (B2), the sub vector space W indeed contains the zero vector and an inverse for each vector $w \in W$, so W satisfies requirements (A2) and (A3). All other requirements in definition 1.2.1 are trivially satisfied for W simply by virtue of them being satisfied in V and W is defined as the subset of V . Hence W is indeed a vector space.

A.2 Relation between Inverse and Bijective Maps

Here we provide a rigorous proof for theorem 3.1.1.

Proof. We assume that $f : X \rightarrow Y$ has an inverse $g : Y \rightarrow X$ with $f \circ g = \text{id}_Y$ and $g \circ f = \text{id}_X$. To show that f is injective start we start with $f(x) = f(\tilde{x})$ and try to apply g from the left. It follows immediately that $x = \tilde{x}$. To show surjectivity we need to find for a given $y \in Y$, an $x \in X$ such that $f(x) = y$. We can choose $x = g(y)$ since then $f(x) = f \circ g(y) = \text{id}_Y(y) = y$. In conclusion f is bijective.

To show uniqueness we consider two maps $g : Y \rightarrow X$ and $\tilde{g} : Y \rightarrow X$ with $x = g \circ f(x) = \tilde{g} \circ f(x)$. Setting $y = f(x)$ it follows that $g(y) = \tilde{g}(y)$ and since f is surjective this holds for all $y \in Y$. Hence $g = \tilde{g}$. \square

A.3 Inverse of a Composite Map

Here we provide a rigorous proof for theorem 3.1.2.

Proof. This relation follows from $(f \circ g)^{-1} \circ (f \circ g) = \text{id}$ by definition and $(g^{-1} \circ f^{-1}) \circ (f \circ g) = g^{-1} \circ g = \text{id}$ where we used the associativity of composite maps. This implies that both $(f \circ g)^{-1}$ and $(g^{-1} \circ f^{-1})$ provide an inverse for $f \circ g$. Uniqueness of the inverse function then leads to the change of order on the RHS of eq. (3.9). \square

A.4 Properties of a Linear Map (1)

Here we provide a rigorous proof for the properties of linear map in lemma 3.2.1

Proof. (F1) $f(\mathbf{0}) = f(\mathbf{00}) \stackrel{(E2)}{=} \mathbf{0}$.

(F2) Checking the two conditions in definition 1.2.2, if $\mathbf{v}_1, \mathbf{v}_2 \in \ker(f)$, then by definition of the kernel, $f(\mathbf{v}_1) = f(\mathbf{v}_2) = \mathbf{0}$. It follows that $f(\mathbf{v}_1 + \mathbf{v}_2) \stackrel{(E1)}{=} f(\mathbf{v}_1) + f(\mathbf{v}_2) = \mathbf{0}$ so $\mathbf{v}_1 + \mathbf{v}_2 \in \ker(f)$ and condition (B1) is satisfied. For (B2), if $\mathbf{v} \in \ker(f)$ so that $f(\mathbf{v}) = \mathbf{0}$ it follows that $f(\alpha\mathbf{v}) \stackrel{(E2)}{=} \alpha f(\mathbf{v}) = \mathbf{0}$, hence $\alpha\mathbf{v} \in \ker(f)$.

(F3) If $\mathbf{w}_1, \mathbf{w}_2 \in \text{Im}(f)$, then by the definition of image we have $f(\mathbf{v}_1) = \mathbf{w}_1$ and $f(\mathbf{v}_2) = \mathbf{w}_2$ for some vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$. So $\mathbf{w}_1 + \mathbf{w}_2 = f(\mathbf{v}_1) + f(\mathbf{v}_2) \stackrel{(E1)}{=} f(\mathbf{v}_1 + \mathbf{v}_2) \in \text{Im}(f)$, hence condition (B1) is satisfied. For (B2), if $f(\mathbf{v}) = \mathbf{w}$ where $\mathbf{v} \in V$ and $\mathbf{w} \in \text{Im}(f)$, then $f(\alpha\mathbf{v}) \stackrel{(E1)}{=} \alpha f(\mathbf{v})$

(F4) The first part is simply the definition of surjectivity. The second part is trivial as the dimensions of same spaces are obviously the same.

(F5) Since $f(\mathbf{0}) = \mathbf{0}$ as shown in (F1), and f is injective, the only vector in V that get mapped to the zero vector $\mathbf{0}$ in W is the zero vector $\mathbf{0}$ in V . The second part is trivial once we know that $\ker(f) = \{\mathbf{0}\}$.

(F6) Simply check the linearity conditions for αf . Firstly, $(\alpha f)(\mathbf{v}_1 + \mathbf{v}_2) = \alpha f(\mathbf{v}_1 + \mathbf{v}_2) = \alpha(f(\mathbf{v}_1) + f(\mathbf{v}_2)) = \alpha f(\mathbf{v}_1) + \alpha f(\mathbf{v}_2) = (\alpha f)(\mathbf{v}_1) + (\alpha f)(\mathbf{v}_2)$. Secondly, $(\alpha f)(\beta\mathbf{v}) = \alpha(f(\beta\mathbf{v})) = \alpha\beta(f(\mathbf{v})) = \beta\alpha f(\mathbf{v}) = \beta(\alpha f)(\mathbf{v})$.

(F7) Checking the linearity conditions for $f + g$, we have firstly $(f + g)(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1 + \mathbf{v}_2) + g(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2) + g(\mathbf{v}_1) + g(\mathbf{v}_2) = f(\mathbf{v}_1) + g(\mathbf{v}_1) + f(\mathbf{v}_2) + g(\mathbf{v}_2) = (f + g)(\mathbf{v}_1) + (f + g)(\mathbf{v}_2)$. and $(\alpha(f + g))(\mathbf{v}) = ((f + g)(\alpha\mathbf{v})) = (f(\alpha\mathbf{v})) + g(\alpha\mathbf{v}) = \alpha f(\mathbf{v}) + \alpha g(\mathbf{v}) = \alpha(f + g)(\mathbf{v})$.

(F8) The two linearity conditions are satisfied since $(g \circ f)(\mathbf{v}_1 + \mathbf{v}_2) = g(f(\mathbf{v}_1 + \mathbf{v}_2)) = g(f(\mathbf{v}_1) + f(\mathbf{v}_2)) = g(f(\mathbf{v}_1)) + g(f(\mathbf{v}_2)) = (g \circ f)(\mathbf{v}_1) + (g \circ f)(\mathbf{v}_2)$ and also $(g \circ f)(\alpha\mathbf{v}) = g(f(\alpha\mathbf{v})) = g(\alpha f(\mathbf{v})) = \alpha g(f(\mathbf{v})) = \alpha(g \circ f)(\mathbf{v})$.

\square

A.5 Dimensions Relation of Domain, Kernel and Image

Here we provide a rigorous proof for theorem 3.2.2.

Proof. For simplicity of notation, set $k = \dim \ker(f)$ and $n = \dim(V)$. Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be the basis of $\ker(f)$ which we complete to a basis $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}, \dots, \mathbf{v}_n$ of V . We will show that $f(\mathbf{v}_{k+1}), \dots, f(\mathbf{v}_n)$ forms a basis of $\text{Im}(f)$. To satisfy condition (C2), we need to show that $\text{Im}(f)$ is spanned by $f(\mathbf{v}_{k+1}), \dots, f(\mathbf{v}_n)$. We begin with an arbitrary vector $\mathbf{w} \in \text{Im}(f)$. This vector must be the image of a $\mathbf{v} \in V$, where \mathbf{v} can be written as the linear combination of the basis in V

$$\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{v}_i \quad (\text{A.1})$$

Acting on this equation with f and using linearity we find

$$\mathbf{w} = f(\mathbf{v}) = f\left(\sum_{i=1}^n \alpha_i \mathbf{v}_i\right) = \sum_{i=1}^n \alpha_i f(\mathbf{v}_i) = \sum_{i=k+1}^n \alpha_i f(\mathbf{v}_i). \quad (\text{A.2})$$

In the last step, we have used the fact that for $f(\mathbf{v}_i) = \mathbf{0}$ for $i = 1, \dots, k$ by the definition of kernel.

Hence, we have written \mathbf{w} as a linear combination of the vectors $f(\mathbf{v}_{k+1}), \dots, f(\mathbf{v}_n)$, which, therefore, span the image of f .

To satisfy condition (C1), we have to show that the vectors $f(\mathbf{v}_{k+1}), \dots, f(\mathbf{v}_n)$ are linearly independent, so as usual we start with the equation

$$\sum_{i=k+1}^n \alpha_i f(\mathbf{v}_i) = \mathbf{0} \implies f\left(\sum_{i=k+1}^n \alpha_i \mathbf{v}_i\right) = \mathbf{0} \quad (\text{A.3})$$

which means that the vector $\sum_{i=k+1}^n \alpha_i \mathbf{v}_i$ is in the kernel of f and, since $\mathbf{v}_1, \dots, \mathbf{v}_k$ form a basis of the kernel, there are coefficients $\alpha_1, \dots, \alpha_k$ such that

$$\sum_{i=k+1}^n \alpha_i \mathbf{v}_i = -\sum_{i=1}^k \alpha_i \mathbf{v}_i \implies \sum_{i=1}^n \alpha_i \mathbf{v}_i = \mathbf{0}. \quad (\text{A.4})$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ form a basis of V it follows that all $\alpha_i = 0$ and, hence, $f(\mathbf{v}_{k+1}), \dots, f(\mathbf{v}_n)$ are linearly independent.

Altogether, it follows that $f(\mathbf{v}_{k+1}), \dots, f(\mathbf{v}_n)$ form a basis of $\text{Im}(f)$. Hence by counting the number of basis elements

$$\dim \text{Im}(f) = n - k = \dim(V) - \dim \ker(f). \quad (\text{A.5})$$

□

A.6 Properties of a Linear Map (2)

Here we provide a rigorous proof for the properties of linear map in lemma 3.2.3

Proof. (G1) If f is bijective, it is also injective and surjective. So from (F4) and (F5) in lemma 3.2.1, $\dim \ker(f) = 0$ and $\dim \text{Im}(f) = \dim(W)$. Then from eq. (3.12), $\dim(V) = \dim \ker(f) + \dim \text{Im}(f) = 0 + \dim(W) = \dim(W)$.

(G2) We set $\mathbf{w}_1 = f(\mathbf{v}_1)$, $\mathbf{w}_2 = f(\mathbf{v}_2)$ and $\mathbf{w} = f(\mathbf{v})$ and check the linearity conditions (E1): $f^{-1}(\mathbf{w}_1 + \mathbf{w}_2) = f^{-1}f((\mathbf{v}_1) + f(\mathbf{v}_2)) = f^{-1}(f(\mathbf{v}_1 + \mathbf{v}_2)) = \mathbf{v}_1 + \mathbf{v}_2 = f^{-1}(\mathbf{w}_1) + f^{-1}(\mathbf{w}_2)$ and (E2): $f^{-1}(\alpha \mathbf{w}) = f^{-1}(\alpha f(\mathbf{v})) = f^{-1}(f(\alpha \mathbf{v})) = \alpha \mathbf{v} = \alpha f^{-1}$

□

A.7 Row Rank is Equals to Column Rank

Here we provide a proof for theorem 3.4.2.

Proof. Suppose one row, say \mathbf{A}_1 , of a $m \times n$ matrix A can be written as a linear combination of the other rows

$$\mathbf{A}_1 = \sum_{j=2}^m \alpha_j \mathbf{A}_j \quad (\text{A.6})$$

then if we define

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} \text{ and } \mathbf{b}_i = \begin{pmatrix} A_{2i} \\ \vdots \\ A_{mi} \end{pmatrix}, \quad (\text{A.7})$$

then the column vectors of A can be written as

$$\mathbf{A}^i = \begin{pmatrix} A_{1i} \\ \mathbf{b}_i \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha} \cdot \mathbf{b}_i \\ \mathbf{b}_i \end{pmatrix}. \quad (\text{A.8})$$

Lemma A.7.1. *The entries of \mathbf{A}_1 are irrelevant for the linear independence of the column vectors \mathbf{A}_i , i.e. if the n column vectors \mathbf{b}_i are linearly independent, then the n column vectors \mathbf{A}^i are also linearly independent and vice versa.*

Proof. If the n column vectors \mathbf{b}_i are linearly independent, then we have

$$\sum_{i=1}^n \beta_i \mathbf{b}_i = 0 \iff \beta_i = 0 \quad \forall i \quad (\text{A.9})$$

To test whether the first row would make any difference, we try to set the linear combination of $\boldsymbol{\alpha} \cdot \mathbf{b}_i$ to be zero

$$\begin{aligned}
\sum_{i=1}^n \gamma_i (\boldsymbol{\alpha} \cdot \mathbf{b}_i) &= \left(\sum_{i=1}^n \gamma_i \sum_{j=2}^m \alpha_j b_{ji} \right) \\
&= \gamma_1 (\alpha_2 b_{21} + \alpha_3 b_{31} + \cdots + \alpha_m b_{m1}) + \\
&\quad \gamma_2 (\alpha_2 b_{22} + \alpha_3 b_{32} + \cdots + \alpha_m b_{m2}) + \cdots + \gamma_n (\alpha_2 b_{2n} + \alpha_3 b_{3n} + \cdots + \alpha_m b_{mn}) \\
&= \alpha_2 (\gamma_1 b_{21} + \gamma_2 b_{22} + \cdots + \gamma_n b_{2n}) + \\
&\quad \alpha_3 (\gamma_1 b_{31} + \gamma_2 b_{32} + \cdots + \gamma_n b_{3n}) + \cdots + \alpha_m (\gamma_1 b_{m1} + \gamma_2 b_{m2} + \cdots + \gamma_n b_{mn}) \\
&= \left(\sum_{j=2}^m \alpha_j \sum_{i=1}^n \gamma_i b_{ji} \right) = 0.
\end{aligned} \tag{A.10}$$

Now since $\alpha_2, \alpha_3, \dots, \alpha_m$ are arbitrary, thus the coefficients for α_i must be zero for all i . This implies that $\sum_{i=1}^n \gamma_i b_{ji} = 0 \quad \forall i$. But we have already proved that this condition implies $\gamma_i = 0 \quad \forall i$ above, since the n column vectors \mathbf{b}_i are linearly independent. Thus adding the entries from the first row would not affect the linear dependency of the column vectors. We can use the same argument if the column vectors are linearly dependent instead. \square

Then by dropping \mathbf{A}_1 from A we arrive at a matrix with one less row, but its row rank and column rank unchanged from that of A .

In this way, we can continue dropping linearly dependent row and column¹ vectors from A until we arrive at a matrix A' which has linearly independent row and column vectors and the same row and column ranks as A . But on a purely dimensional standpoint, a matrix with all row vectors and all column vectors linearly independent must be quadratic². Therefore the row rank must be equals to the column rank. \square

A.8 Properties of an Inverse of a Matrix

Here we provide proof for the properties listed in lemma 3.4.3.

Proof. (H1) This has already been shown in the main text.

(H2) These are direct consequences of eq. (3.9) for general maps.

(H3) From theorem 3.4.2 we know that $\text{rank}(A^T) = \text{rank}(A) = n$ and from (H1) we have shown that a matrix is invertible if and only if its rank is maximal, we conclude that A^T is indeed invertible which proves the first part of the claim. For the second part, we transpose the equation $A^{-1}A = AA^{-1} = \mathbb{I}$, to arrive at $A^T(A^{-1})^T = (A^{-1})^T A^T = \mathbb{I}$. On the other hand, we also have $A^T(A^T)^{-1} = (A^T)^{-1} A^T = \mathbb{I}$. Comparing the two equations shows that both $(A^{-1})^T$ and $(A^T)^{-1}$ provide an inverse for A^T and, hence from the uniqueness of the inverse, they must be equal. \square

¹Since rows and columns has not inherent difference and we have proven for the case of row vectors, so it must hold for column vectors as well.

²For example, consider a 3×2 matrix. Its three 2-dimensional row vectors cannot be linearly independent.

A.9 Rank of a Matrix in Upper Echelon Form

Here we provide a proof for lemma 3.4.4.

Proof. Consider the row vector at j_i^{th} row, if it can be written as the linear combination of the rest of the row vectors where

$$\mathbf{A}_{j_i} = \sum_{k=1, k \neq i}^r \alpha_k \mathbf{A}_k, \quad (\text{A.11})$$

then by inspection $\alpha_k = 0 \forall k < i$ since the entries in j_1^{th} column is non-zero only at the first row, so $\alpha_1 = 0$ ensures that the entry of the j_i^{th} row vector is zero at j_1^{th} . The same logic applies to the second row now that $\alpha_1 = 0$ which gives $\alpha_2 = 0$. This process continues for all $k < i$.

However, the j_i^{th} row vector cannot be expressed in terms of a linear combination of the row vectors for $k > i$, since all the entries in the j_i^{th} column are zero for $k > i$ but non-zero otherwise.

Therefore an arbitrary row vector in this (upper) echelon form cannot be expressed as a linear combination of the rest of the row vectors. Hence, the first r row vectors in this matrix are linearly independent.

□

A.10 Matrix Rank is Equals to Linear Map Rank

Here we provide a proof for the last claim in lemma 3.5.1

Proof. Since ?? implies that $\text{Im}(A) = \psi^{-1}(\text{Im}(f))$, if we denote by $\chi \equiv \psi^{-1}|_{\text{Im}(f)}$ we have $\dim \ker(\chi) = 0$ since ψ^{-1} is an isomorphism (as are all coordinate maps) and hence $\dim \ker(\psi^{-1}) = 0$. We can then use eq. (3.12) to get

$$\text{rank}(A) = \dim \text{Im}(A) = \text{rank}(\chi) = \dim \text{Im}(f) - \dim \ker(\chi) = \text{rank}(f). \quad (\text{A.12})$$

□

B.1 Fields

Fields with finite number of elements can also exist, the simplest example would be $\mathbb{F}_p = \{0, 1, \dots, p-1\}$, where p is a prime number. And the addition and multiplication in this field is defined to be

$$a + b = (a + b) \bmod p \text{ and } a \cdot b = (ab) \bmod p. \quad (\text{B.1})$$

where the modulus operation is introduced such that the result of addition and multiplication is brought back to the required range $\{0, 1, \dots, p-1\}$ whenever it exceeds $p-1$.

The smallest example of a field in this class is $\mathbb{F}_2 = \{0, 1\}$, where the addition and multiplication tables are

+	0	1
0	0	1
1	1	0

Table B.1: addition table for \mathbb{F}_2

·	0	1
0	0	0
1	0	1

Table B.2: multiplication table for \mathbb{F}_2

Note that taking into account the mod 2 operation, in this field we have $1 + 1 = 0$. Since the elements of \mathbb{F}_2 can be viewed as the two states of a bit, this field has important applications in computer science.

B.2 Matrices for Elementary Row Operations

The matrices corresponding to the 3 kinds of elementary row operations are given below as

$$\begin{aligned}
P_{R_i \leftrightarrow R_j} &= \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & 1 & i^{\text{th}} \text{ row} \\ & & & & \ddots & \\ & & & 1 & & 0 & j^{\text{th}} \text{ row} \\ & & & & & 1 & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix}, \\
P_{R_i \rightarrow \lambda R_i} &= \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \lambda & & i^{\text{th}} \text{ row} \\ & & & & 1 & \\ & & & & & \ddots & \\ & & i^{\text{th}} \text{ col.} & & & & 1 \end{pmatrix} \text{ and} \\
P_{R_i \rightarrow R_i + \lambda R_j} &= \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & \dots & \lambda & i^{\text{th}} \text{ row} \\ & & & \ddots & \vdots & \\ & & & & 1 & \\ & & & & & \ddots & \\ & & & j^{\text{th}} \text{ col.} & & & 1 \end{pmatrix}.
\end{aligned} \tag{B.2}$$