Linear Algebra

Haydn Cheng

Contents

1 Vectors and Vector Spaces

1.1 Vectors in $\mathbb{F}^n$

Vectors are mathemtical objects that satisfy certain properties, and they can be geometrical arrows in space, polynomials, matrices, *etc.*

Field (denoted by $\mathbb{F}$) contains elements (called scalars) that can be combined (multiplied) with the basis chosen for the vector space to construct any arbitrary vectors in the vector space. Examples include the real numbers $\mathbb{R}$ and the complex numbers $\mathbb{C}$.[1]

The simplest n-dimensional vector space containing n-dimensional vectors is $\mathbb{F}^n$, which is defined as the set of all ordered n-tuples of elements (called the components of a vector) in $\mathbb{F}$[2]

$$\mathbb{F}^n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in F, i = 1, 2, \ldots, n\} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}. \tag{1.1}$$

The standard unit vectors are the n vectors in $\mathbb{F}^n$ where $n-1$ elements chosen to form the n-tuples are 0, and the remaining element is 1:

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

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

[2] Note that we have not associate the vector to an arrow in space (yet), up till now we are still in an abstract mathematical discussion.

When we write a vector in $\mathbb{F}^n$ we assume that the underlying basis is the standard unit vectors, unless otherwise specified,[3] otherwise the components of the vector would be entirely different.[4]

As we will see in section 3.5, we can define an associated vector (known as the coordinate vector) in the vector space $\mathbb{F}^n$ for every n-dimensional vector that we want to describe. This is because a vector in $\mathbb{F}^n$ simply stores n elements from $\mathbb{F}$ into an array, which acts as the components of a more complicated vector.

More often than not it is much easier if we visualize a vector in $\mathbb{F}^n$ as a geometrical arrow in space with basis $\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}$, *etc.*, so $\mathbf{v} = a_1\hat{\mathbf{x}} + a_2\hat{\mathbf{y}} + a_3\hat{\mathbf{z}} + \cdots$, as an intermediate between $\mathbb{F}^n$ and an more abstract vector space, since this is the (second) simpliest kind of vector one can think of but it is also much more common and related to physics and daily life.

Example: Google's search alogorithm.

Question: Consider an internet with n websites labeled by $k = 1, \ldots, n$. The site labeled as k has been referenced to by other sites $L_k \subset 1, \ldots, n$. 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

$$x_k = n_k. \tag{1.3}$$

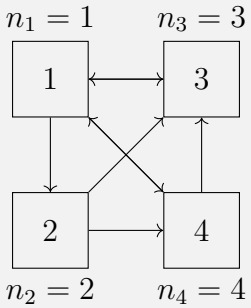
An improved version would be to take into account the page rank of the pages that reference this page, as it is more desirable to be referenced by a page with higher page rank. So

$$x_k = \sum_{j \in L_k} x_j \tag{1.4}$$

would be a improvement to the formula. Futhermore, as a reference is more valuable if a page only reference a samll number of page, we may define the page rank as

$$x_k = \sum_{j \in L_k} \frac{x_j}{n_j}. \tag{1.5}$$

A simple example is illustrated below:



[3]We will ocassionally add a prime besides the bracket to denote the changed basis

[4]In fact, changing the basis is a topic we will devote an entire section into.

As shown in the figure, $n = 4, n_1 = 3, n_2 = 2, n_3 = 1, n_4 = 2, L_1 = 3, 4, L_2 = 1, L_3 = 1, 2, 4$ and $L_4 = 1, 2$. So the formulas for assigning the page rank becomes

$$\begin{aligned} x_1 &= \frac{x_3}{1} + \frac{x_4}{2} \\ x_2 &= \frac{x_1}{3} \\ x_3 &= \frac{x_1}{3} + \frac{x_2}{2} + \frac{x_4}{2} \\ x_4 &= \frac{x_1}{3} + \frac{x_2}{2}. \end{aligned} \tag{1.6}$$

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

$$A\mathbf{x} = \mathbf{x}, \tag{1.7}$$

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \text{ and } A = \begin{pmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{pmatrix}. \tag{1.8}$$

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

1.2 Vector Spaces

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 $0 + \mathbf{v} = \mathbf{v}$.

(A3) "Inverse element": There exists an inverse, $-\mathbf{v}$ with $\mathbf{v} + (-\mathbf{v}) = 0$.

(A4) "Commutativity": $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$.

(A5) $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$.

(A6) $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{w}$.

(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.,*[5]

(B1) $\mathbf{w}_1 + \mathbf{w}_2 \in W$ *for all* $\mathbf{w}_1, \mathbf{w}_2 \in W$*.*

(B2) $\alpha\mathbf{w} \in W$ *for all* $\alpha \in \mathbf{F}$ *and* $\mathbf{w} \in W$*.*

The sum of $U + W$ of two sub vector spaces is called direct if and only if $U \cap W = \mathbf{0}$ and a direct sum is written as $U \oplus W$.

Example: Constructing Scalar Multiplication.

Question: A vector addition operation $\oplus$ is defined for the set of n-dimensional vectors, $\mathbb{R}^n$ by

$$\mathbf{x} \oplus \mathbf{y} \equiv \mathbf{x} + \mathbf{y} - \mathbf{a}, \tag{1.9}$$

where $\mathbf{a}$ is a fixed vector and the addtion and subtraction operations on the right hand side are defined conventionally.
Construct a corresponding scalar multiplication operation $\odot$ such that $\mathcal{R}^n$, together with operations $\oplus$ and $\odot$ forms a vector space over the field $\mathcal{R}^n$ and show that, with these definitions, scalar multiplication $\odot$ is distributive over vector addition $\oplus$.

Solution: The "null vector" and the inverse of a general vector $\mathbf{x}$,*i.e.,* $\mathbf{0}$ and $\mathbf{x}'$, respectively satisfy

$$\mathbf{x} \oplus \mathbf{0} = \mathbf{x} \quad \text{and} \quad \mathbf{x} \oplus \mathbf{x}' = \mathbf{0}. \tag{1.10}$$

Using the definition of $\oplus$ defined in the question we find

$$\mathbf{0} = \mathbf{a} \quad \text{and} \quad \mathbf{x}' = 2\mathbf{a} - \mathbf{x}. \tag{1.11}$$

Now to construct the scalar multiplication $\odot$ we test the property $0 \odot \mathbf{v} = \mathbf{0}$ (which is not an one of the 8 axioms but can be derived from them) to get

$$0 \odot \mathbf{v} = \mathbf{0} = \mathbf{a} \implies \alpha \odot \mathbf{v} = \alpha(?) + \mathbf{a}. \tag{1.12}$$

Then we test the axiom $(A8)$ to get n

$$1 \odot \mathbf{v} = \mathbf{v} \implies \alpha \odot \mathbf{v} = \alpha(? - \mathbf{a}) + \mathbf{a} = \alpha(\mathbf{v} - \mathbf{a}) + \mathbf{a}. \tag{1.13}$$

The distributive property can be shown easily, as

$$\alpha \odot (\mathbf{x} \oplus \mathbf{y}) = \alpha \odot (\mathbf{x} + \mathbf{y} - \mathbf{a}) = \alpha(\mathbf{x} + \mathbf{y} - 2\mathbf{a}) + \mathbf{a}, \tag{1.14}$$

and

$$\alpha \odot (\mathbf{x} \oplus \mathbf{y}) = \alpha \odot \mathbf{x} + \alpha \odot \mathbf{y} = \alpha(\mathbf{x} - \mathbf{a}) + \mathbf{a} + \alpha(\mathbf{y} - \mathbf{a}) + \mathbf{a} = \alpha(\mathbf{x} + \mathbf{y} - 2\mathbf{a}) + \mathbf{a}. \tag{1.15}$$

[5]The rigorous proof of a sub vector space is a vector space itself is provided in section A.2.

In hindsight the scalar multiplication can be constructed easily we notice that the expression $\mathbf{x} - \mathbf{a}$ can be thought of as the coordinates of $\mathbf{x}$ relative to a new origin $\mathbf{a}$. We first sclar these relative coordinates by α in the standard way: $\alpha(\mathbf{x} - \mathbf{a})$ and then we shift the result back by adding $\mathbf{a}$ to get the vector in the original system.

1.3 Span and Linear Independence

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

$$\text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k) \equiv \left\{ \sum_{i=1}^{k} \alpha_i \mathbf{v}_i \mid \alpha_i \in F \right\}. \tag{1.16}$$

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

$$\sum_{i=1} \alpha_i \mathbf{v}_i = \mathbf{0} \implies \alpha_i = 0 \text{ for all } i. \tag{1.17}$$

Otherwise, the vectors are called linearly dependent.

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

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

$$\mathbf{v}_i = -\frac{1}{\alpha_i} \sum_{j \neq i} \alpha_j \mathbf{v}_j, \tag{1.18}$$

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

In general, to determine whether a column vector is linearly dependent of others, and to find the linear combination coefficients, we have to solve, for example,

$$x \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} + y \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}. \tag{1.19}$$

There is no simple way to find x, y by eye, and the recommended method is to row reduce A to its reduced row echelon form R

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} = B, \tag{1.20}$$

so that we immediately have $(3,-3,0) = -(1,0,0) + 2(2,-1,-6)$. This method will be explored in greater details in section 3.7.2.

□

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

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

$$\mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \tag{1.21}$$

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

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 = \begin{pmatrix} \alpha_3 \\ \alpha_1 + \alpha_2 + \alpha_3 \\ \alpha_1 + 2\alpha_2 - \alpha_3 \end{pmatrix} = \mathbf{0} \tag{1.22}$$

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 Trigonometric Functions.

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

$$\frac{d^2y}{dx^2} = -y \tag{1.23}$$

are linearly indpendent.

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

$$\alpha \sin x + \beta \cos x = 0, \tag{1.24}$$

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.

Example: Linear Dependence of Three Four-Dimensional Vectors.

Question: Determine whether the vectors $\begin{pmatrix} 1 \\ 2 \\ 0 \\ -3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \\ 4 \end{pmatrix}$ and $\begin{pmatrix} -3 \\ 6 \\ -4 \\ 1 \end{pmatrix}$ are linearly indpendent.

Solution: One way to check their linearly independecy is to check whether setting the linaer combination of them implies that the coefficients of each vector is zero. Alternatively, we can find the rank of the matrix A formed by the vectors and find its rank. The matrix can be row reduce to

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & 6 \\ 0 & 1 & -4 \\ -3 & -4 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & -4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.25}$$

Therefore the rank of A is two and the vectors are linearly depedent.

To test whether three three-dimensional vectors are linear indepedence, the fastest way is to compute the determinant and check whether or not it is equals to zero.

1.4 Basis and Dimension

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

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

(C2) $V = Span(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

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

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i. \tag{1.26}$$

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

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

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \sum_{i=1}^{n} \beta_i \mathbf{v}_i. \tag{1.27}$$

Taking the difference of these two equations implies

$$\sum_{i=1}^{n} (\alpha_i - \beta_i) \mathbf{v}_i = 0 \tag{1.28}$$

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

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

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

Proof. Consider the first arbitrary vector $\mathbf{w}_1$. If $\mathbf{w}_1 = 0$ then the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_m$ are linearly dependent (since the coefficient of $\mathbf{w}_1$ can be any arbirary constant when we write down the linaer combination of the vectors $\mathbf{w}_1, \cdots, \mathbf{w}_n$ 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

$$\mathbf{w}_1 = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \tag{1.29}$$

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

$$\mathbf{v}_1 = \frac{1}{\alpha_1}\left(\mathbf{w}_1 - \sum_{i=2}^{n} \alpha_i \mathbf{v}_i\right). \tag{1.30}$$

This shows that we use $\mathbf{w}_1$ to replace $\mathbf{v}_1$ in basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ such that $V = \mathrm{Span}(\mathbf{w}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$, 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 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ can now be written as a linear combination of the basis $\mathbf{w}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ simply by utilizing eq. (1.30).

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

$$\mathbf{w}_{n+1} = \sum_{i=1}^{n} \beta_i \mathbf{w}_i \tag{1.31}$$

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

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

Example: Standard Unit Vectors form a Basis.

Question: Prove that the standard unit vectors $\mathbf{e}_1, \cdots, \mathbf{e}_n$ 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,

$$\sum_{i=1}^{n} \alpha_i \mathbf{e}_i = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = 0 \tag{1.32}$$

which is only true if $\alpha_i = 0$ for all i.
Thus the standard unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 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

$$\dim_{\mathbb{R}}(\mathbb{R}^n) = \dim_{\mathbb{C}}(\mathbb{C}^n) = n. \tag{1.33}$$

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

Example: Basis of a Polynomial.

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

Solution: We start with setting

$$\sum_{i=0}^{d} \alpha_i \mathbf{x}^i = 0 \tag{1.34}$$

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 $\alpha_k = 0$ for all 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

$$E_{(i,j)} = \begin{pmatrix} 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}, ^a \tag{1.35}$$

where $i = 1, \ldots, n$ and $j = 1, \ldots, m$ 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.

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

Example: Semi-Magic Squares.

Question: A semi-magic square is a 3×3 matrix of (rational) numbers such that all rows and columns sum up to the same total. Show that the matrices

$$M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, M_2 = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, M_3 = \begin{pmatrix} 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, M_4 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \end{pmatrix}, M_5 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{1.36}$$

form a basis of the semi-magic squares.

Solution:

Example: Page Rank.

Question: Consider an "internet" with n sites, labeled by $i = 1, \ldots, n$. Each site i contains n_i links to some of the other sites and it is linked to by the pages $L_i \subset \{1, \ldots, n\}$. The page rank x_i on each site i is defined by

$$x_i = \sum_{j \in L_i} \frac{x_i}{n_j}. \tag{1.37}$$

Show that the above equation can be written in matrix form as $A\mathbf{x} = 0$. Then show that a non-trivial solution always exists.
Analyze a simple example with $n = 4$ and the following struture of links shown in fig. 1.1.
Write down the page rank equations for this case and solve them. Which page is ranked highest?

Solution:

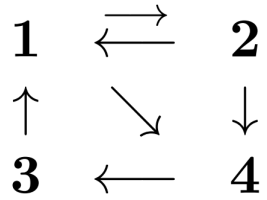


Figure 1.1

Bl

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

(D1) *V has a basis.*

(D2) *Every linearly independent set $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$ 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 linaerly independent set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ forms a basis.*

(D4) *If $\dim_F(V) = \dim_F(W)$ and $V \subset W$ for two vector spaces V and W then $V = W$.*

Proof. $(D1)$ Since $V = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, 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 spoan, so $V = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1})$. This process can be continued until the remainng set of vectors is linaerly independent.

$(D2)$ If $V = \text{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ we are finished. If not there exists a vector $\mathbf{v}_{k+1} \notin \mathit{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. Hence the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}$ must be linaerly 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 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 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 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ must span the space and they form a basis.

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

□

2 Geometrical Applications of Vectors in $\mathbb{R}^n$

2.1 Scalar and Vector Products

2.1.1 Kronecker Delta and Levi-Civita Symbol

The kronecker delta is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.1}$$

The Levi-Civita symbol is defined by

$$\epsilon_{ijk} = \begin{cases} +1 & \text{if } (i,j,k) = (1,2,3), (2,3,1), (3,1,2), \\ -1 & \text{if } (i,j,k) = (2,1,3), (3,2,1), (1,3,2), \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

In Einstein summation convention, there is are pairs of dummy indices and some free indices. The free indices indicate which component of the vector we are describing, and only appear once in a term, but there can be more than one free indices. The repeated (twice) indices are the dummy indices, which implies summation over them. No indices can appear more than twice.

Note, however, for example, that $x_i^2 \neq x_i x_i$ when we use the Einstein summation convention. The former is the square of the i^{th} component of the vector $\mathbf{x}$, and the latter is summing over the square of all components of the vector $\mathbf{x}$.

One important equality regarding the Levi-Civita Symbol is

$$\epsilon_{ijk}\epsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}, \tag{2.3}$$

which is true, since $\epsilon_{ijk}\epsilon_{ilm} = \sum_{i=1}^{3} \epsilon_{ijk}\epsilon_{ilm} = +1$ if $(j,l) = (k,m)$ and $\epsilon_{ijk} = -1$ if $(j,l) = (m,k)$ (and in any other combinations of (j,k,l,m) we get zero).

2.1.2 Scalar Product in $\mathbb{R}^n$

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

$$\mathbf{a} \cdot \mathbf{b} \equiv \sum_{i=1}^{n} a_i b_i = a_i b_i = \delta_{ij} a_i b_j, \tag{2.4}$$

given that the basis is orthogonal. (The dot product for under a general basis will be covered in chapter 5).

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

$$|\mathbf{a}| \equiv \sqrt{\mathbf{a} \cdot \mathbf{a}} = \left(\sum_{i=1}^{n} a_i^2 \right)^{\frac{1}{2}} \tag{2.5}$$

given that the basis is orthogonal.

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

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|. \tag{2.6}$$

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

$$0 \leq |\mathbf{a} \pm \mathbf{b}|^2 = (\mathbf{a} \pm \mathbf{b}) \cdot (\mathbf{a} \pm \mathbf{b}) = |\mathbf{a}|^2 + 2(\mathbf{a} \cdot \mathbf{b}) + |\mathbf{b}|^2 = 2(1 \pm \mathbf{a} \cdot \mathbf{b}) \implies |\mathbf{a} \cdot \mathbf{b}| \leq 1. \tag{2.7}$$

□

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*

$$|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}| \tag{2.8}$$

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

$$-1 \leq \cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|a||b|} \leq 1 \implies \mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta. \tag{2.9}$$

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

2.1.3 Vector Product in $\mathbb{R}^3$

Scalar Triple Product $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$

The scalar triple product $\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C})$ represents the volume of the parallelepiped generated by $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ (as shown in fig. 2.1) since $|\mathbf{B}\times\mathbf{C}|$ is the base area, and $|\mathbf{A}\cos\theta|$ is the altitude. Therefore, by considering different pairs of bases and altitudes, the following identity can be shown

$$\mathbf{A}\cdot(\mathbf{B}\times\mathbf{C}) = \mathbf{B}\cdot(\mathbf{C}\times\mathbf{A}) = \mathbf{C}\cdot(\mathbf{A}\times\mathbf{B}). \tag{2.10}$$

An easy way to remember this identity is that it adopts the same "clockwise convention" as $\hat{\mathbf{x}}\times\hat{\mathbf{y}} = \hat{\mathbf{z}}$ & $\hat{\mathbf{y}}\times\hat{\mathbf{z}} = \hat{\mathbf{x}}$ & $\hat{\mathbf{z}}\times\hat{\mathbf{x}} = \hat{\mathbf{y}}$.

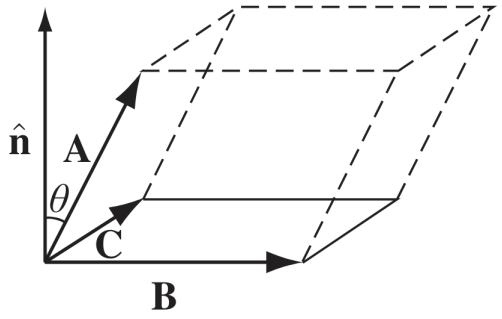


Figure 2.1

Vector Triple Product $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$

The vector triple product $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})$ can be simplified with the so-called $\mathbf{BAC}-\mathbf{CAB}$ identity

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B}). \tag{2.11}$$

This can be shown by

$$\begin{aligned}
(\mathbf{a}\times(\mathbf{b}\times\mathbf{c}))_i &= \epsilon_{ijk}a_j\,(\mathbf{b}\times\mathbf{c})_k \\
&= \epsilon_{ijk}a_j\epsilon_{kmn}b_m c_n \\
&= (\delta_{im}\delta_{jn} - \delta_{in}\delta_{jm})a_j b_m c_n \\
&= a_j b_i c_j - a_j b_j c_i \\
&= ((\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{c})_i
\end{aligned} \tag{2.12}$$

With eq. (2.10) and eq. (2.11) in hand, it is never necessary for an expression to contain more than one cross product in any term.

Example: Griffiths (5th ed.) P.8

Question: Simplify $(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D})$ and $\mathbf{A}\times(\mathbf{B}\times(\mathbf{C}\times\mathbf{D}))$.

Solution: For the first equality, using eqs. (2.10) and (2.11), we have

$$\begin{aligned}(\mathbf{A}\times\mathbf{B})\cdot(\mathbf{C}\times\mathbf{D}) &= \mathbf{C}\cdot(\mathbf{D}\times(\mathbf{A}\times\mathbf{B}))\\ &= \mathbf{C}\cdot(\mathbf{A}(\mathbf{B}\cdot\mathbf{D})-\mathbf{B}(\mathbf{D}\cdot\mathbf{A}))\\ &= (\mathbf{A}\cdot\mathbf{C})(\mathbf{B}\cdot\mathbf{D})-(\mathbf{A}\cdot\mathbf{D})(\mathbf{B}\cdot\mathbf{C}).\end{aligned} \tag{2.13}$$

Alternatively, we have

$$\begin{aligned}((\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}))_i &= (\mathbf{a}\times\mathbf{b})_i(\mathbf{c}\times\mathbf{d})_i\\ &= \epsilon_{ijk}a_jb_k\epsilon_{imn}c_md_n\\ &= \epsilon_{jki}\epsilon_{imn}a_jb_kc_md_n\\ &= (\delta_{jm}\delta_{kn}-\delta_{jn}\delta_{mk})a_jb_kc_md_n\\ &= a_jc_jb_kd_k-a_jd_jb_kc_k\\ &= (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})-(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c}).\end{aligned} \tag{2.14}$$

For the second equality,

$$\begin{aligned}\mathbf{A}\times(\mathbf{B}\times(\mathbf{C}\times\mathbf{D})) &= \mathbf{A}\times(\mathbf{C}(\mathbf{B}\cdot\mathbf{D})-\mathbf{D}(\mathbf{B}\cdot\mathbf{C}))\\ &= (\mathbf{B}\cdot\mathbf{D})(\mathbf{A}\times\mathbf{C})-(\mathbf{B}\cdot\mathbf{C})(\mathbf{A}\times\mathbf{D})\\ (\text{ or }\ \mathbf{A}\times(\mathbf{B}\times(\mathbf{C}\times\mathbf{D})) &= \mathbf{B}(\mathbf{A}\cdot(\mathbf{C}\times\mathbf{D}))-(\mathbf{C}\times\mathbf{D})(\mathbf{A}\cdot\mathbf{B}))\end{aligned} \tag{2.15}$$

Example: Griffiths (5th ed.) Problem 1.6

Question: When does $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=(\mathbf{A}\times\mathbf{B})\times\mathbf{C}$?

Solution: Since

$$\begin{aligned}&\mathbf{A}\times(\mathbf{B}\times\mathbf{C})+\mathbf{B}\times(\mathbf{C}\times\mathbf{A})+\mathbf{C}\times(\mathbf{A}\times\mathbf{B})\\ &=\mathbf{B}(\mathbf{A}\cdot\mathbf{C})-\mathbf{C}(\mathbf{A}\cdot\mathbf{B})+(\mathbf{C}(\mathbf{B}\cdot\mathbf{A})-\mathbf{A}(\mathbf{B}\cdot\mathbf{C}))+(\mathbf{A}(\mathbf{C}\cdot\mathbf{B})-\mathbf{B}(\mathbf{C}\cdot\mathbf{A}))\\ &=0,\end{aligned} \tag{2.16}$$

Therefore, $\mathbf{A}\times(\mathbf{B}\times\mathbf{C})=(\mathbf{A}\times\mathbf{B})\times\mathbf{C}$ when $\mathbf{B}\times(\mathbf{C}\times\mathbf{A})=0$.
Thus, either $\mathbf{A}/\!/\mathbf{C}$ *or* $\mathbf{A}\perp\mathbf{B}\perp\mathbf{C}$

Example: Quadraple Cross Product.

Question: Simplify $(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{b})$.

Solution:

$$\begin{aligned}((\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{b}))_i &= \epsilon_{ijk}(\epsilon_{jpq}a_pb_q)(\epsilon_{krs}c_rb_s)\\ &= \epsilon_{ijk}\epsilon_{krs}\epsilon_{jpq}a_pb_qc_rb_s\\ &= (\delta_{ir}\delta_{js}-\delta_{is}\delta_{rj})\epsilon_{jpq}a_pb_qc_rb_s\\ &= \epsilon_{jpq}a_pb_q(c_ib_j-b_ic_j)\\ &= ((\mathbf{a}\times\mathbf{b})\cdot\mathbf{b})c_i-((\mathbf{a}\times\mathbf{b})\cdot\mathbf{c})b_i\\ &= (\mathbf{b}\cdot(\mathbf{a}\times\mathbf{c}))b_i.\end{aligned} \tag{2.17}$$

Example: Lagrange's Identity.

Question: Prove the Lagrange's Identity

$$|\mathbf{a}\times\mathbf{b}|^2 = |\mathbf{a}|^2|\mathbf{b}|^2-(\mathbf{a}\cdot\mathbf{b})^2. \tag{2.18}$$

Find the matrix M such that the above relation may be written in the form

$$|\mathbf{a}\times\mathbf{b}|^2 = \mathbf{a}^TM\mathbf{a}. \tag{2.19}$$

Solution: We have

$$|\mathbf{a}\times\mathbf{b}|^2 = |\mathbf{a}|^2\left|\mathbf{b}^2\right|\sin^2\theta = |\mathbf{a}|^2|\mathbf{b}|^2(1-\cos^2\theta) = |\mathbf{a}|^2|\mathbf{b}|^2\left(1-\frac{(\mathbf{a}\cdot\mathbf{b})^2}{|\mathbf{a}|^2|\mathbf{b}|^2}\right) = |\mathbf{a}|^2|\mathbf{b}|^2-(\mathbf{a}\cdot\mathbf{b})^2. \tag{2.20}$$

We note that

$$\mathbf{a}^2\mathbf{b}^2 = \mathbf{a}^T(|\mathbf{b}|^2\mathbb{I})\mathbf{a} \quad\text{and}\quad (\mathbf{a}\cdot\mathbf{b})^2 = \mathbf{a}^T(\mathbf{b}\mathbf{b}^T)\mathbf{a}, \tag{2.21}$$

so

$$M = |\mathbf{b}|^2\mathbb{I}-\mathbf{b}\mathbf{b}^T = \begin{pmatrix}(b_2^2+b_3^2) & -b_1b_2 & -b_1b_3\\ -b_2b_1 & (b_1^2+sd) & -b_2b_3\\ -b_3b_1 & -b_3b_2 & (b_1^2+b_2^2)\end{pmatrix}. \tag{2.22}$$

2.2 Equations of Lines, Planes and Spheres

2.2.1 Equation of Lines

The general equation of a line is

$$\mathbf{r} = \mathbf{a}+\lambda\mathbf{b} \Longleftrightarrow (\mathbf{r}-\mathbf{a})\times\mathbf{b} = \mathbf{0} \Longleftrightarrow \frac{x-a_x}{b_x} = \frac{y-a_y}{b_y} = \frac{z-a_z}{b_z} = \text{constant}, \tag{2.23}$$

where $\mathbf{r}$ is the position vector of a general point on the line, $\mathbf{a}$ is the position vector of

some fixed point on the line and $\mathbf{b}$ is some fixed vector along the direction of the line.

2.2.2 Equation of Planes

The general equation of a plane is

$$\mathbf{r} = \mathbf{a} + \lambda\mathbf{b} + \mu\mathbf{c} \iff (\mathbf{r} - \mathbf{a}) \cdot \hat{\mathbf{n}} = \mathbf{0} \iff \mathbf{r} \cdot \hat{\mathbf{n}} = \hat{n}_x x + \hat{n}_y y + \hat{n}_z z = \text{constant}, \quad (2.24)$$

where $\mathbf{a}$ is the position vector of some fixed point on the plane, $\mathbf{b}$ and $\mathbf{c}$ are some fixed (linearly independent) vectors along the direction of the plane and $\hat{\mathbf{n}}$ is the normal vector of the surface, which is related to $\mathbf{b}$ and $\mathbf{c}$ by $\hat{\mathbf{n}} = (\mathbf{b} \times \mathbf{c})/(|\mathbf{b}||\mathbf{c}|)$.

The constant in the above equation is the shortest distance from the origin to the plane, given that $\hat{\mathbf{n}}$ is normalized.

Example: Line of Intersection of Two Planes.

Question: Find the direction of the line of intersection of the two planes $x + 3y - z = 5$ and $2x - 2y + 4z = 3$.

Solution: The two planes have normal vectors

$$\hat{\mathbf{n}}_1 = \hat{\mathbf{x}} + 3\hat{\mathbf{y}} - \hat{\mathbf{z}} \text{ and } \hat{\mathbf{n}}_2 = 2\hat{\mathbf{x}} - 2\hat{\mathbf{y}} + 4\hat{\mathbf{z}}. \quad (2.25)$$

The direction of the intersecting line mst be parallel to both planes and hence perpendicular to both normals. Therfore,

$$\mathbf{p} = \hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2 = 10\hat{\mathbf{x}} - 6\hat{\mathbf{y}} - 8\hat{\mathbf{z}}. \quad (2.26)$$

Example: Planes and Spheres.

Question: Find the radius ρ of the circle that is the intersection of the plane $\hat{\mathbf{n}} \cdot \mathbf{r} = p$ and the sphere of radius a centered on the point with position vector $\mathbf{c}$.

Solution: The equation of the sphere is

$$|\mathbf{r} - \mathbf{c}|^2 = a^2 \implies r^2 - 2\mathbf{r} \cdot \mathbf{c} + c^2 = a^2, \quad (2.27)$$

while the equation of the circle of intersection is

$$|\mathbf{r} - \mathbf{b}|^2 = \rho^2, \quad (2.28)$$

where $\mathbf{r}$ is restricted to lie in the plane and $\mathbf{b}$ is the position vector of the circle's center.
From simple geometry, we have

$$\mathbf{b}-\mathbf{c}=\lambda\hat{\mathbf{n}} \quad \text{and} \quad \rho^2+|\mathbf{b}-\mathbf{c}|^2=a^2, \tag{2.29}$$

which gives $\mathbf{b}=\mathbf{c}+\sqrt{a^2-\rho^2}\hat{\mathbf{n}}$. Substituting into the the equation of the circle, we have

$$r^2-2\mathbf{r}\cdot\left(\mathbf{c}+\sqrt{a^2-\rho^2}\hat{\mathbf{n}}\right)+c^2+2(\mathbf{c}\cdot\hat{\mathbf{n}})\sqrt{a^2-\rho^2}+a^2-\rho^2=\rho^2. \tag{2.30}$$

We finally yields $\rho=\sqrt{a^2-(p-c\cdot\hat{\mathbf{n}})^2}$.

Example: Circle of Intersection of Two Spheres.

Question: The equations of two spheres are

$$\begin{cases}(\mathbf{r}-\mathbf{a})\cdot(\mathbf{r}-\mathbf{a}) &= \alpha^2,\\ (\mathbf{r}-\mathbf{b})\cdot(\mathbf{r}-\mathbf{b}) &= \beta^2.\end{cases} \tag{2.31}$$

Find the condition for the spheres to intersect. For the case where the spheres do intersect, find the position vector of the center of the circle of intersectio $\mathbf{R}$.

Solution: The spheres intersect if the distnace d between their ceneres is less than or equal to the sum of their radii and greater than or equal to the difference of their radii, so

$$||\beta|-|\alpha||\leq|\mathbf{b}-\mathbf{a}|\leq|\alpha|+|\beta|. \tag{2.32}$$

Expanding the two equations of the two spheres and subtracting them gives the equation for the circle of intersection

$$\begin{aligned}2\mathbf{r}\cdot(\mathbf{b}-\mathbf{a})&=\alpha^2-\beta^2-|\mathbf{a}|^2+|\mathbf{b}|^2\\ \mathbf{r}\cdot\frac{\mathbf{b}-\mathbf{a}}{|\mathbf{b}-\mathbf{a}|}&=\frac{(\alpha^2-\beta^2-|\mathbf{a}|^2+|\mathbf{b}|^2)}{2|\mathbf{b}-\mathbf{a}|}.\end{aligned} \tag{2.33}$$

We then let $\mathbf{R}=\mathbf{a}+t(\mathbf{b}-\mathbf{a})$ and substitute into the equation of the circle above to get

$$2(\mathbf{a}+t(\mathbf{b}-\mathbf{a}))\cdot(\mathbf{b}-\mathbf{a})=\alpha^2-\beta^2-|\mathbf{a}|^2+|\mathbf{b}|^2 \implies t=\frac{\alpha^2-\beta^2+|\mathbf{b}-\mathbf{a}|^2}{2|\mathbf{b}-\mathbf{a}|^2}, \tag{2.34}$$

so

$$\mathbf{R}=\frac{1}{2}\left(\mathbf{a}+\mathbf{b}+\frac{(\alpha^2-\beta^2)(\mathbf{b}-\mathbf{a})}{|\mathbf{b}-\mathbf{a}|^2}\right). \tag{2.35}$$

Alternatively, we can project $\mathbf{a}$ (or $\mathbf{b}$) onto the plane of the circle of intersection, and get the same answer

$$\mathbf{R}=\mathbf{a}-\mathbf{a}_\perp=\mathbf{a}-\left(\mathbf{a}\cdot\frac{\mathbf{b}-\mathbf{a}}{|\mathbf{b}-\mathbf{a}|}-\frac{(\alpha^2-\beta^2-|\mathbf{a}|^2+|\mathbf{b}|^2)}{2|\mathbf{b}-\mathbf{a}|}\right)\frac{\mathbf{b}-\mathbf{a}}{|\mathbf{b}-\mathbf{a}|}, \tag{2.36}$$

where we have used the distance from a point to a plane formula (eq. (2.43)).

Example: Intersection between a Plane and a sphere.

Question: Let $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ be constant vectors and r be a real constant. Find the locus of points for $\mathbf{x}$ resulting from the simultaneous solution of the equations

$$\mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot \mathbf{b} \quad \text{and} \quad |\mathbf{x} - \mathbf{c}| = r. \tag{2.37}$$

Solution: For negative r we have no intersections because the norm of any vector must be non-negative. For $r = 0$ we must have $\mathbf{x} = \mathbf{c}$, given taht $\mathbf{c} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b}$ is satisfied. For $r > 0$, the intersection is a cirlce centered at $\mathbf{c} + \mathbf{d}$ with raidus $R = \sqrt{r^2 - |\mathbf{d}|^2}$. To find the vector $\mathbf{d}$, we use the charateristic facts that $\mathbf{c} + \mathbf{d}$ lies on the plane and $\mathbf{d}$ is parallel to $\mathbf{a}$ to get

$$\mathbf{d} \cdot \mathbf{a} = \mathbf{a} \cdot \mathbf{b} - \mathbf{a} \cdot \mathbf{c} \quad \text{and} \quad \mathbf{d} \times \mathbf{a} = 0. \tag{2.38}$$

By crossing the second identity with $\mathbf{a}$ and using the vector identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$, we get

$$\mathbf{d} = \frac{(\mathbf{a} \cdot (\mathbf{b} - \mathbf{c}))\mathbf{a}}{|\mathbf{a}|^2}. \tag{2.39}$$

The locus of $\mathbf{x}$ is then given by $\mathbf{a} \cdot \mathbf{x} = \mathbf{a} \cdot \mathbf{b}$ and

$$|\mathbf{x} - \mathbf{c} - \mathbf{d}| = \left| \mathbf{x} - \mathbf{c} - \frac{(\mathbf{a} \cdot (\mathbf{b} - \mathbf{c}))\mathbf{a}}{|\mathbf{a}|^2} \right| = \sqrt{r^2 - \frac{(\mathbf{a} \cdot \mathbf{b} - \mathbf{a} \cdot \mathbf{c})^2}{|\mathbf{a}|^2}}. \tag{2.40}$$

To combine the two constriants we simply set the sum of square to be zero

$$(\mathbf{a} \cdot \mathbf{x} - \mathbf{a} \cdot \mathbf{b})^2 + \left(\left| \mathbf{x} - \mathbf{c} - \frac{(\mathbf{a} \cdot (\mathbf{b} - \mathbf{c}))\mathbf{a}}{|\mathbf{a}|^2} \right| - \sqrt{r^2 - \frac{(\mathbf{a} \cdot \mathbf{b} - \mathbf{a} \cdot \mathbf{c})^2}{|\mathbf{a}|^2}} \right)^2 = 0. \tag{2.41}$$

2.3 Distances Between Points, Lines and Planes

2.3.1 Distances From a Point to a Line

The minimum distance d from a point P with position vector $\mathbf{p}$ to the line $\mathbf{r} = \mathbf{a} + \lambda \mathbf{b}$ is given by

$$d = |\mathbf{p} - \mathbf{a}| \sin\theta = \left| (\mathbf{p} - \mathbf{a}) \times \hat{\mathbf{b}} \right|. \tag{2.42}$$

2.3.2 Distances From a Point to a Plane

The minimum distance d from a point P with position vector $\mathbf{p}$ to the plane $(\mathbf{r}-\mathbf{a})\cdot\hat{\mathbf{n}}=0$ is given by

$$d = |\mathbf{p}-\mathbf{a}|\sin\theta = |(\mathbf{p}-\mathbf{a})\cdot\hat{\mathbf{n}}|. \tag{2.43}$$

Note that the sign of d indicate whether the point P is on the same or opposite side of the surface as the origin.

2.3.3 Distances From a Line to a Line

Consider two lines $\mathbf{r}=\mathbf{a}+\lambda\mathbf{a}'$ and $\mathbf{r}'=\mathbf{b}+\mu'\mathbf{b}'$. To find the minimum distance d between the two lines, it is easier to imagine two parallel planes, each consists of one of the lines.

The common normal unit vector of the two planes is

$$\hat{\mathbf{n}} = \frac{\mathbf{b}\times\mathbf{a}}{|\mathbf{b}\times\mathbf{a}|}, \tag{2.44}$$

and the minimum distance d is

$$d = |\mathbf{b}-\mathbf{a}|\sin\theta = \left|(\mathbf{b}-\mathbf{a})\cdot\frac{\mathbf{b}'\times\mathbf{a}'}{|\mathbf{b}'\times\mathbf{a}'|}\right|. \tag{2.45}$$

2.3.4 Distances From a Line to a Plane

The minimum distance d from a line $\mathbf{r}=\mathbf{a}+\lambda\mathbf{b}$ to a plane $(\mathbf{r}-\mathbf{a}')\cdot\hat{\mathbf{n}}=0$ is always zero unless the line is parallel to the plane (*i.e.,* $\mathbf{b}\cdot\hat{\mathbf{n}}=0$). In those cases, the minimum distance d is

$$d = |(\mathbf{a}'-\mathbf{a})\cdot\hat{\mathbf{n}}|. \tag{2.46}$$

2.3.5 Ratio Theorem

Let $\mathbf{a}$ and $\mathbf{b}$ be the position vectors of point A and B respectively. If point P with position vector $\mathbf{p}$ divides the line segement AB in the ratio $\lambda:\mu$, then $\mathbf{p}$ is given by

$$\mathbf{p} = \frac{\mu\mathbf{a}+\lambda\mathbf{b}}{\mu+\lambda}. \tag{2.47}$$

Example: Position Vector of the Centroid of a Triangle.

Question: The vertices of triangle ABC have position vectors $\mathbf{a},\mathbf{b}$ and $\mathbf{c}$ (refer to fig. 2.2). Find the position vector of the centroid G of the triangle.

Solution: The position vectors of D and E are

$$\mathbf{d} = \frac{\mathbf{a}}{2} + \frac{\mathbf{b}}{2} \quad \text{and} \quad \mathbf{e} = \frac{\mathbf{a}}{2} + \frac{\mathbf{c}}{2}. \tag{2.48}$$

A general point on CD that divides the line in the ratio $\lambda : 1 - \lambda$ has the position vector

$$\mathbf{r} = (1 - \lambda)\mathbf{c} + \lambda\mathbf{d} = (1 - \lambda)\mathbf{c} + \frac{\lambda}{2}(\mathbf{a} + \mathbf{b}). \tag{2.49}$$

Similarly, a general point on BE that divides the line in the ratio $\mu : 1 - \mu$ can be expressed as

$$\mathbf{r} = (1 - \mu)\mathbf{b} + \mu\mathbf{e} = (1 - \mu)\mathbf{b} + \frac{\mu}{2}(\mathbf{a} + \mathbf{c}). \tag{2.50}$$

Thus at G, which is the intersection of CD and BE, we require

$$(1 - \lambda)\mathbf{c} + \frac{\lambda}{2}(\mathbf{a} + \mathbf{b}) = (1 - \mu)\mathbf{b} + \frac{\mu}{2}(\mathbf{a} + \mathbf{c}). \tag{2.51}$$

Equating the coefficients, we have

$$\lambda = \mu, \frac{\lambda}{2} = 1 - \mu \quad \text{and} \quad 1 - \lambda = \frac{\mu}{2} \implies \mathbf{g} = \frac{1}{3}(\mathbf{a} + \mathbf{b} + \mathbf{c}). \tag{2.52}$$

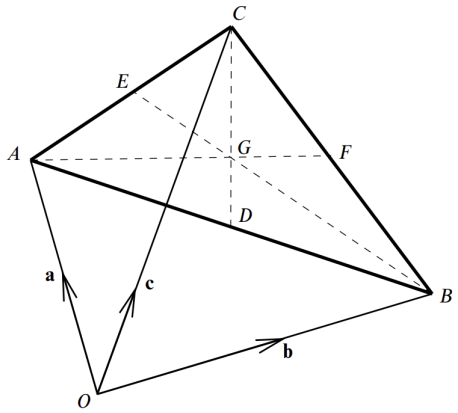


Figure 2.2

Example: Shortest Distance between Lines.

Question: Find λ at the ponit at which the two lines

$$\begin{cases} \mathbf{r}_1 &= \mathbf{a} + \lambda\mathbf{b}, \\ \mathbf{r}_2 &= \mathbf{c} + \mu\mathbf{d} \end{cases} \tag{2.53}$$

are at the shortest distance.

Solution: At the shortest distance, we have

$$\Delta\mathbf{r} \cdot \mathbf{b} = \Delta\mathbf{r} \cdot \mathbf{d} = 0. \tag{2.54}$$

Substitute $\Delta\mathbf{r} = \mathbf{c} - \mathbf{a} + \mu\mathbf{d} - \lambda\mathbf{b}$, we have a linear system of equations for λ and μ

$$\begin{cases} (\mathbf{c}-\mathbf{a})\cdot\mathbf{b} + \mu(\mathbf{d}\cdot\mathbf{b}) - \lambda(\mathbf{b}\cdot\mathbf{b}) &= 0, \\ (\mathbf{c}-\mathbf{a})\cdot\mathbf{d} + \mu(\mathbf{d}\cdot\mathbf{d}) - \lambda(\mathbf{b}\cdot\mathbf{d}) &= 0. \end{cases} \tag{2.55}$$

Solving for λ gives

$$\lambda = \frac{(\mathbf{c}-\mathbf{a})\cdot((\mathbf{d}\cdot\mathbf{d})\mathbf{b} - (\mathbf{b}\cdot\mathbf{d})\mathbf{d})}{|\mathbf{b}|^2|\mathbf{d}|^2 - (\mathbf{b}\cdot\mathbf{d})^2}. \tag{2.56}$$

Example: Collisions between Particles.

Question: Two balls of diameter d are fired along trajectories $\mathbf{r}_1 = \mathbf{a} + \mathbf{p}t$ and $\mathbf{r}_2 = \mathbf{b} + \mathbf{q}t$. Derive an expression of $\mathbf{a}, \mathbf{b}, \mathbf{p}$ and $\mathbf{q}$ for which the balls will collide.

Solution: The condition for collision is

$$|\Delta\mathbf{r}|_{\min} = |\mathbf{a}-\mathbf{b}+(\mathbf{p}-\mathbf{q})t| = |\mathbf{p}-\mathbf{q}|^2t^2 + 2(\mathbf{a}-\mathbf{b})(\mathbf{p}-\mathbf{q})t + |\mathbf{a}-\mathbf{b}|^2 - d^2 = 0. \tag{2.57}$$

Requiring the determinant to be non-negative we get

$$((\mathbf{a}-\mathbf{b})\cdot(\mathbf{p}-\mathbf{q}))^2 \geq |\mathbf{p}-\mathbf{q}|^2(|\mathbf{a}-\mathbf{b}|^2 - d^2). \tag{2.58}$$

2.4 Reciprocal Vectors

The two sets of vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{a}', \mathbf{b}', \mathbf{c}'$ are called reciprocal sets if $\mathbf{a}\cdot\mathbf{a}' = \mathbf{b}\cdot\mathbf{b}' = \mathbf{c}\cdot\mathbf{c}' = 1$ while other possible combinations of dot product is zero.

They are related by

$$\mathbf{a}' = \frac{\mathbf{b}\times\mathbf{c}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})}, \quad \mathbf{b}' = \frac{\mathbf{c}\times\mathbf{a}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})} \text{ and } \mathbf{c}' = \frac{\mathbf{a}\times\mathbf{b}}{\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})}. \tag{2.59}$$

The reciprocal vectors are useful, since the components of an arbitrary vector can be represented by a dot product

$$\mathbf{v} = (\mathbf{v}\cdot\mathbf{a}')\mathbf{a} + (\mathbf{v}\cdot\mathbf{b}')\mathbf{b} + (\mathbf{c}\cdot\mathbf{c}')\mathbf{c}, \tag{2.60}$$

which can be easily verified by taking the scalar product of $\mathbf{v}$ and, say $\mathbf{a}'$ to get

$$\mathbf{v}\cdot\mathbf{a}' = (\mathbf{v}\cdot\mathbf{a}')(\mathbf{a}\cdot\mathbf{a}') + (\mathbf{v}\cdot\mathbf{b}')(\mathbf{b}\cdot\mathbf{a}') + (\mathbf{v}\cdot\mathbf{c}')(\mathbf{c}\cdot\mathbf{a}') = (\mathbf{v}\cdot\mathbf{a}'). \tag{2.61}$$

If $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are mutually orthogonal then $\mathbf{a}' = \mathbf{a}, \mathbf{b}' = \mathbf{b}$ and $\mathbf{c}' = \mathbf{c}$, in those cases we have

$$\mathbf{v} = (\mathbf{v} \cdot \mathbf{a})\mathbf{a} + (\mathbf{v} \cdot \mathbf{b})\mathbf{b} + (\mathbf{v} \cdot \mathbf{c})\mathbf{c}, \tag{2.62}$$

as expected.

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 (or range) of x under f. In symbols,*

$$f : X \to Y, \quad x \mapsto f(x). \tag{3.1}$$

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

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

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 $f : X \to Y$ be a map between two sets X and Y. The map f is called*

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

$$f(x) = f(\tilde{x}) \implies x = \tilde{x} \text{ for all } x, \tilde{x} \in X. \tag{3.3}$$

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

$$\mathrm{Im}(f) = Y \tag{3.4}$$

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 $f : X \to Y$ and $g : Y \to Z$ the composite map $g \circ f : X \to Z$ is defined by*

$$(g \circ f)(x) \equiv g(f(x)) \tag{3.5}$$

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

$$h \circ (g \circ f) = (h \circ g) \circ f \tag{3.6}$$

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* $\mathrm{id}_X : X \to X$ *maps every element in* X *onto itself, i.e.,*

$$\mathrm{id}_X(x) = x \text{ for all } x \in X. \tag{3.7}$$

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

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

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

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 domain.[1] □

Theorem 3.1.2 (Inverse of a Composite Map)**.** *If the maps* $f : X \to Y$ *and* $g : Y \to Z$ *are both bijective, then the composite map* $f \circ g : Y \to Z$ *is also bijective and hence by theorem 3.1.1 has an inverse given by*[2]

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

3.2 Linear Maps

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

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

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

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

[1] The rigorous proof of this theorem can be found in section A.3.

[2] The theorem is intuitive and trivial but the rigorous proof is given in section A.4.

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

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

Lemma 3.2.1 (Properties of a Linear Map (1))**.** *A linaer map $f : V \to W$ between two vector spaces V and W over$\mathbb{F}$ has the following properties:*[3]

(F 1) The zero vectors are mapped onto each other, so $f(\mathbf{0}) = 0$. Hence $\mathbf{0} \in \ker(f)$.

(F 2) The kernel of f is a sub vector space of V.

(F 3) The image of f is a sub vector space of W .

(F 4) f is surjective $\iff \operatorname{Im}(f) = W \iff \dim \operatorname{Im}(f) = \dim(W)$

(F 5) f is injective $\iff \ker(f) = \{\mathbf{0}\} \iff \dim \ker(f) = 0$

(F 6) The scalar multiple αf, where $\alpha \in F$, is linear.

(F 7) For another linear map $g : V \to W$, the sum $f + g$ is linear.

(F 8) For another linear map $g : W \to U$, the composition $g \circ f$ is linear.[4]

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,*

$$\operatorname{rank}(f) \equiv \dim \operatorname{Im}(f). \tag{3.11}$$

To be concrete, consider a linear map $f : \mathbb{R}^3 \to \mathbb{R}^2$ and assume that $\dim \ker(f) = 2$, *i.e.,* the kernel of f is a plane in $\mathbb{R}^3$ (which passes through the origin since any sub vector space contains the zero vector). 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 $f(\mathbf{v}_1 - \mathbf{v}_2) = 0$ 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 $f : V \to W$ we have*

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

Lemma 3.2.3 (Properties of a Linear Map (2))**.** *For a linaer map $f : V \to W$ we have:*[6]

[3]Since the properties are rather intuitive and trivial, the rigorous proof for these properties is provided in section A.5.

[4]From (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 iteself, denoted $\operatorname{Hom}(V, W)$, called the homomorphisms from V to W.

[5]Since this formula is relatively intuitive and imaginable, the rigorous proof is given in section A.6.

[6]The rigorous proof is given in section A.7 due to its simplicity and imaganability.

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

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

Example: Dimension of the Sum of Two Sub Vector Spaces.

Question: Let U and W be two sub vector spaces of V. Show that the dimensions of the above vector spaces are related by

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W). \tag{3.13}$$

Solution: Let the dimensions of $U \cap W, U$ and W be $p, p + m$ and $p + n$ respectively. Then $(\mathbf{v}_1, \cdots, \mathbf{v}_p), (\mathbf{v}_1, \cdots, \mathbf{v}_p, \mathbf{u}_1, \cdots, \mathbf{u}_m), (\mathbf{v}_1, \cdots, \mathbf{v}_p, \mathbf{w}_1, \cdots, \mathbf{w}_n)$ are the basis $U \cap W, U$ and W respectively.
To prove that $\dim(U + W) = m + n + p$, we need to prove that $(\mathbf{v}_1, \cdots, \mathbf{v}_p, \mathbf{u}_1, \cdots, \mathbf{u}_m, \mathbf{n}_1, \cdots, \mathbf{n}_n)$ forms a basis for $U + W$. Firstly, they clearly span the space $U + W$, so we just need to prove the linear independency. Set the linear combination of them to be zero

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_p\mathbf{v}_p + \beta_1\mathbf{u}_1 + \cdots + \beta_m\mathbf{u}_m + \gamma_1\mathbf{w}_1 + \cdots + \gamma_n\mathbf{w}_n = 0, \tag{3.14}$$

then the vector

$$\mathbf{x} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_p\mathbf{v}_p + \beta_1\mathbf{u}_1 + \cdots + \beta_m\mathbf{u}_m = -(\gamma_1\mathbf{w}_1 + \cdots + \gamma_n\mathbf{w}_n) = 0 \tag{3.15}$$

belongs to both U and W, so $\mathbf{x} \in U \cap W$, thus $\mathbf{x}$ can also be written as

$$\mathbf{x} = \delta_1\mathbf{v}_1 + \cdots + \delta_p\mathbf{v}_p. \tag{3.16}$$

Thus, we get

$$\mathbf{x} - \mathbf{x} = \delta_1\mathbf{v}_1 + \cdots + \delta_p\mathbf{v}_p - (\gamma_1\mathbf{w}_1 + \cdots + \gamma_n\mathbf{w}_n) = 0 \implies (\delta_1, \ldots, \delta_p, \gamma_1, \ldots, \gamma_n) = 0, \tag{3.17}$$

since the set of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{w}_1, \ldots, \mathbf{w}_n)$ are linearly independent. From this we get also that

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_p\mathbf{v}_p + \beta_1\mathbf{u}_1 + \cdots + \beta_m\mathbf{u}_m = 0 \implies (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_m) = 0. \tag{3.18}$$

Therefore the set of vectors $(\mathbf{v}_1, \cdots, \mathbf{v}_p, \mathbf{u}_1, \cdots, \mathbf{u}_m, \mathbf{w}_1, \cdots, \mathbf{w}_n)$ forms a basis for the vector space $U + W$, and the dimension of $U + W$ is the number of vectors in its basis, which is $\dim(U + W) = m + n + p = \dim(U) + \dim(W) - \dim(U \cap W)$.

3.3 Coordinates Maps $f : \mathbb{F}^n \to V$

To map a n-dimensional column vector $\boldsymbol{\alpha} \in \mathbb{F}^n$ (where the basis is not necessarily the standard unit vectors) with componenets α_i to a n-dimensional vector $\mathbf{v} \in V$ over$\mathbb{F}$ with basis $\mathbf{v}_1, \cdots, \mathbf{v}_n$, we can define the coordinate map $\varphi : \mathbb{F}^n \to V$ as

$$\varphi(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i. \tag{3.19}$$

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

$$\begin{aligned} \varphi(\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2) &= \sum_{i=1}^{n} (\alpha_{1_i} + \alpha_{2_i})\mathbf{v}_i = \sum_{i=1}^{n} \alpha_{1_i}\mathbf{v}_i + \alpha_{2_i}\mathbf{v}_i = \varphi(\boldsymbol{\alpha}) + \varphi(\boldsymbol{\alpha}) \\ \varphi(a\boldsymbol{\alpha}) &= \sum_{i=1}^{n} (a\alpha_i)\mathbf{v}_i = a\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = a\varphi(\boldsymbol{\alpha}). \end{aligned} \tag{3.20}$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ forms a basis it is clear that $\mathrm{Im}(\varphi) = V$ 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 $\varphi^{-1} : V \to \mathbb{F}^n$. The inverse map assigns to a vector $\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \in V$ its coordinate vector $\boldsymbol{\alpha} \in \mathbb{F}^n$

$$\varphi^{-1}(\mathbf{v}) = \varphi^{-1}\left(\sum_{i=1}^{n} \alpha_i \mathbf{v}_i\right) = \boldsymbol{\alpha}. \tag{3.21}$$

In short, we can associate any vector with a vector in $\mathcal{F}^n$ by a means of coorinates map.

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 $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$ by means of a coordinate map φ.

3.4 Matrices

3.4.1 Linear Maps $f : \mathbb{F}^n \to \mathbb{F}^m$

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

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \tag{3.22}$$

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.

It is important to note that the entries of A depends on both the basis chosen for the doamin and the co-domain. The relations between different matrices for different basis choice representing the same linear map is discussed in section 3.6.

The matrix multiplication of a $m \times n$ matrix on a n-dimensional vector is defined as

$$A\mathbf{v} \equiv \begin{pmatrix} a_{11}v_1 + \cdots + a_{1n}v_n \\ \vdots \\ a_{m1}v_1 + \cdots + a_{mn}v_n \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 \cdot \mathbf{v} \\ \vdots \\ \mathbf{A}_m \cdot \mathbf{v} \end{pmatrix} = \sum_{i=1}^{m} v_i \mathbf{A}^i \tag{3.23}$$

.

As we can see, there are useful interpretation for the matrix multiplication $A\mathbf{v}$, the row picture illustrates that the rows of $A\mathbf{v}$ come from dot products of $\mathbf{v}$ with the rows of $\mathbf{A}$, and the column picuture shows that the columns of $A\mathbf{v}$ come from linear combination of A with entries of $\mathbf{v}$, so we can imagine A as an operator that transform $\mathbf{v}$ to another vector $A\mathbf{v}$.

The two ways of visualizing a matrix multiplied by a vector is illustrated in fig. 3.1.

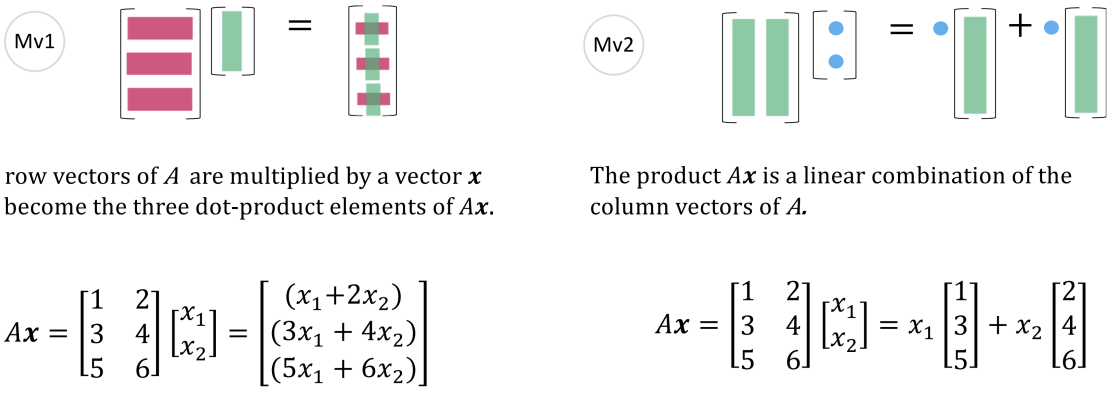


Figure 3.1

The row picture is easier for computation but the column picture is easier for understanding.

Using index notation,

$$(A\mathbf{v})_i = \sum_{j=1}^{n} a_{ij}v_j = a_{ij}v_j, \tag{3.24}$$

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 begin described.

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

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2)_i &= a_{ij}(v_{1_j} + v_{2_j}) = a_{ij}v_{1_j} + a_{ij}w_{2_j} = f(\mathbf{v}_1)_i + f(\mathbf{v}_2)_i, \\ f(\alpha\mathbf{v})_i &= a_{ij}(\alpha v_j) = \alpha(a_{ij}v_j) = \alpha f(\mathbf{v})_i. \end{aligned} \tag{3.25}$$

In above, we have proved that every $m \times n$ matrix multiplication represent a linear map action from $\mathbb{F}^n$ to $\mathbb{F}^m$. However, we have not shown the converse is true, *i.e.,* whether every linear map action from $\mathbb{F}^n$ to $\mathbb{F}^m$ can be represented by a $m \times n$ matrix.

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

$$f(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\tilde{\mathbf{e}}_i \tag{3.26}$$

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

$$f(\mathbf{v}) = f\left(\sum_{j=1}^{n} v_j \mathbf{e}_j\right) = \sum_{j=1}^{n} v_j f(\mathbf{e}_j) = \sum_{j=1}^{n} v_j \sum_{i=1}^{m} a_{ij}\tilde{\mathbf{e}}_i = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}v_j\right)\tilde{\mathbf{e}}_i.^{7} \tag{3.27}$$

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

$$f(\mathbf{v})_i = \sum_{j=1}^{n} a_{ij}v_j = (A\mathbf{v})_i, \tag{3.28}$$

where we used eq. (3.24) to establish the second equality.

Lemma 3.4.1. *Every linear map $f : \mathbb{F}^n \to \mathbb{F}^m$ can be written in terms of a $m \times n$ matrix A, such that $f(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$. If $f(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\tilde{\mathbf{e}}_i$ for the standard unit vectors $\mathbf{e}_i$ of $\mathbb{F}^n$ and $\tilde{\mathbf{e}}_i$ of $\mathbb{F}^m$, 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 $f : \mathbb{R}^3 \to \mathbb{R}^3$ 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)

$$\begin{aligned} f(\mathbf{v}_1 + \mathbf{v}_2) &= \mathbf{n} \times (\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{n} \times \mathbf{v}_1 + \mathbf{n} \times \mathbf{v}_2 = f(\mathbf{v}_1) + f(\mathbf{v}_2) \\ f(\alpha\mathbf{v}) &= \mathbf{n} \times (\alpha\mathbf{v}) = \alpha\mathbf{n} \times \mathbf{v} = \alpha f(\mathbf{v}). \end{aligned} \tag{3.29}$$

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 $\mathbf{n} = \begin{pmatrix} n_1 & n_2 & n_3 \end{pmatrix}^T$ we find by explicit computation that

[7] This changing of order of summation is similar to what has been done in section A.8.

$$
\begin{aligned}
f(\mathbf{e}_1) &= \mathbf{n} \times \mathbf{e}_1 = n_3\mathbf{e}_2 - n_2\mathbf{e}_3 \\
f(\mathbf{e}_2) &= \mathbf{n} \times \mathbf{e}_2 = -n_3\mathbf{e}_1 + n_1\mathbf{e}_3 \\
f(\mathbf{e}_3) &= \mathbf{n} \times \mathbf{e}_3 = n_2\mathbf{e}_1 - n_1\mathbf{e}_2.
\end{aligned} \tag{3.30}
$$

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 desire matrix is

$$
A = \begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_2 \\ -n_2 & n_1 & 0 \end{pmatrix}, \tag{3.31}
$$

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

$$
A_{ij} = f(\mathbf{e}_j)_i = [\mathbf{n} \times \mathbf{e}_j]_i = \epsilon_{ikl} n_k [\mathbf{e}_j]_l = \epsilon_{ikl} n_k \delta_{jl} = \epsilon_{ikj} n_k \tag{3.32}
$$

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

3.4.2 Basic Properties of Matrices

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

$$
\mathbb{I}_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \text{ or } (\mathbb{I}_n)_{ij} = \delta_{ij} \implies (\mathbb{I}\mathbf{v})_i = \delta_{ij} v_j = v_i. \tag{3.33}
$$

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 $D_{ij} = 0$ for all $i \neq j$. It can be written as

$$
D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \equiv \text{diag}(d_1, \ldots, d_n). \tag{3.34}
$$

A useful property of diagonal matrices is that the product is commutative, *i.e.,* $AB = BA$.

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 $A^* : \mathbb{F}^m \to \mathbb{F}^n$ of a matrix $A : \mathbb{F}^m \to \mathbb{F}^n$ 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 (hermitian conjugate) of an $m \times n$ matrix $A : \mathbb{F}^n \to \mathbb{F}^m$ is a $n \times m$ matrix $A^T(A^\dagger) : \mathbb{F}^m \to \mathbb{F}^n$ obtained by exchanging the rows and columns of A (and taking the complex conjugate of every entries). In component form, this means $(A^T)_{ij} = A_{ji}((A^\dagger)_{ij} = A^*_{ji})$.

A square matrix A is said to be symmetric (hermitian) if $A = A^T(A = A^\dagger)$, or $A_{ij} = A_{ji}(A_{ij} = A^*_{ji})$ for all entries, and is called anti-symmetric (or skew-symmetric) if $A =$

$-A^T (A = -A^\dagger)$, or $A_{ij} = -A_{ji} (A_{ij} = -A^*_{ji})$ for all entries (note that all diagonal entries A_{ii} of an anti-symmetric (hermitian) matrix vanish). Note that any matrix $A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T) (A = \frac{1}{2}(A + A^\dagger) + \frac{1}{2}(A - A^\dagger))$ can be written as a sum of a symmetric (hermitian) and an anti-symmetric (hermitian) matrix.

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

Another important class of matrices is normal matrices, which satsifies $AA^T = A^T A (AA^\dagger = A^\dagger A)$. Examples of normal matrices include symmetric (hermitian) and orthogonal (unitary) matrices (the latter of which will be introduced in section 5.7). Note that if A is normal then so too is its inverse A^{-1}, since $A^{-1}(A^{-1})^\dagger = A^{-1}(A^\dagger)^{-1} = (A^\dagger A)^{-1} = (AA^\dagger)^{-1} = (A^\dagger)^{-1}A^{-1} = A^{-1^\dagger}A^{-1}$.

Example: Deterimants of Anti-Symmetric Matrices.

Question: If A is a $n \times n$ anti-symmetric matrix, show that $|A| = 0$ if n is odd.

Solution: For an anti-symmetric matrix, we have

$$A^T = -A \implies |A| = \left|A^T\right| = |-A| = (-1)^n |A|. \tag{3.35}$$

Thus if n is odd then $|A| = -|A| \implies |A| = 0$.

Example: Commutativity of Transpose and Powers.

Question: Prove that $(A^T)^n = (A^n)^T$.

Solution: $(A^n)^T = (A \cdots A)^T = (A^T \cdots A^T) = (A^T)^n$.

3.4.3 Rank of Matrices

From eq. (3.23) we see that the image of $\mathbf{v}$ is given by a linear combination of the column vectors $\mathbf{A}^i$ with coefficients equatl 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 $\mathbf{v}$. Thus observation tells us that

$$\mathrm{Im}(A) = \mathrm{Span}(\mathbf{A}^1, \ldots, \mathbf{A}^m), \tag{3.36}$$

and is called the column space of the matrix, 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 $\mathrm{rank}(f) = \dim \mathrm{Im}(f)$. Treating matrix multiplication as the action of a linear map, we have

$$\begin{aligned} \mathrm{rank}(A) &= \dim \mathrm{Im}(A) = \dim \mathrm{Span}(\mathbf{A}^1, \ldots, \mathbf{A}^m) \\ &= \text{maximal number of linear independent column vectors of } A. \end{aligned} \tag{3.37}$$

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]

A nice interpretation of matrix multiplication $A = CR$ is that matrix C contains all the

Example: Kernel and Image of A Matrix.

Question: Consider the 3×3 matrix

$$A = \begin{pmatrix} -1 & 4 & 3 \\ 2 & -3 & -1 \\ 3 & 2 & 5 \end{pmatrix}, \tag{3.38}$$

which defines a map $A : \mathbb{R}^3 \to \mathbb{R}^3$.
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 ahve to solve $A\mathbf{v} = \mathbf{0}$. With $\mathbf{v} = \begin{pmatrix} x & y & z \end{pmatrix}^T$ 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

$$\ker(A) = \text{Span} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \text{ and } \text{Im}(A) = \text{Span} \left(\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ -3 \\ 2 \end{pmatrix} \right). \tag{3.39}$$

3.4.4 Multiplication of Matrices

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 $\mathbb{F}^n \xrightarrow{A} \mathbb{F}^m \xrightarrow{B} \mathbb{F}^r$. We would like to determine the matrix C which describes the composite map $B \circ A : \mathbb{F}^n \to \mathbb{F}^r$. By straightforward computation we find

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

$$(B(A\mathbf{v}))_i = \sum_{j=1}^{n} B_{ij}(A\mathbf{v})_j = \sum_{j=1}^{n} B_{ij}\left(\sum_{k=1}^{n} A_{jk}\mathbf{v}_k\right) = \sum_{k=1}^{n}\left(\left(\sum_{j=1}^{n} B_{ij}A_{jk}\right)\mathbf{v}_k\right) = \sum_{k=1}^{n} C_{ik}v_k = (C\mathbf{v})_i$$
$$\implies C_{ik} = \sum_{j=1}^{n} B_{ij}A_{jk} = \mathbf{B}_i \cdot \mathbf{A}^k \tag{3.40}$$

This is an extension of the row picture in eq. (3.23), where each entry in $C = BA$ is a dot product between a different row in B and a different column in A. We can also easily extend the column picture in eq. (3.23) just by imagining the matrix B multiplies on the column vectors in A individually to produce the column vecotrs in C. So we can imagine B as an operator that transform the column vectors in A to another set of vectors, before AB transform a vector $\mathbf{v}$ to another vector.

Alternatively, we can view matrix multiplications as linear combinations of rows or entries of outer product, but these two interpretations are seldom used.

The four interpretations for matrix multiplications are illustrated in fig. 3.2.

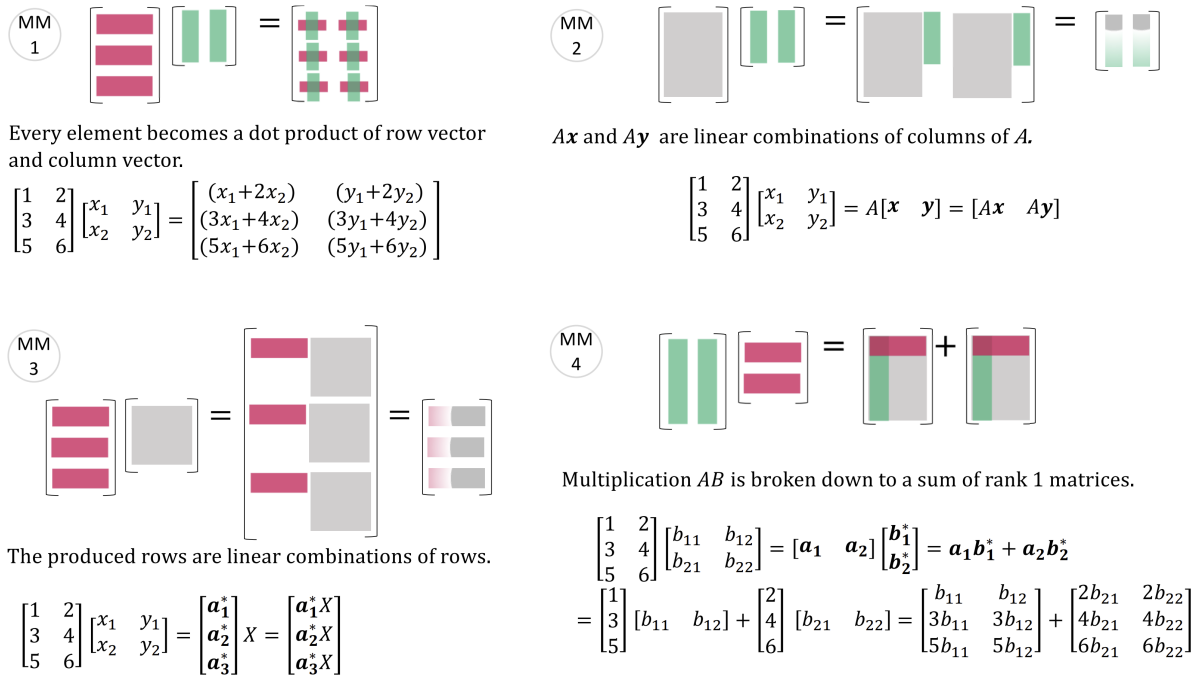


Figure 3.2

So again, the row picture is easier for calculation and the column picture is easier for understanding.

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

$$\begin{aligned}(A(BC))_{ij} = A_{ik}(BC)_{kj} = A_{ik}B_{kl}C_{lj} = (AB)_{il}C_{lj} = ((AB)C)_{ij} \\ \implies A(BC) = (AB)C.\end{aligned} \tag{3.41}$$

A useful relationship involving multiplication and transposition of matrices is

$$\begin{aligned}((AB)^T)_{ij} = (AB)_{ji} = A_{jk}B_{ki} = B_{ki}A_{jk} = (B^T)_{ik}(A^T)_{kj} = (B^TA^T)_{ij} \\ \implies (AB)^T = B^TA^T.\end{aligned} \tag{3.42}$$

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

Using matrix terminology, the dot product of two column vectors with the same dimension $\mathbf{v}$ and $\mathbf{w}$ can be written as

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T\mathbf{w}. \tag{3.43}$$

3.4.5 Inverse of Matrices

For a matrix A representing a linear map $f : \mathbb{F}^n \to \mathbb{F}^m$ to have an inverse, it must be bijective as proved in theorem 3.1.1. Thus from $(F5)$ and $(G1)$ we have $\dim\ker(f) = \mathbf{0}$ 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 $A : \mathbb{F}^n \to \mathbb{F}^n$ is invertible if and only if its rank is maximal, i.e.,* $\text{rank}(A) = n$*. If A and B are two invertible $n \times n$ matrices we have:*[9]

(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 $(A^T)^{-1} = (A^{-1})^T$.

The properties of matrix inverse is best understood by visualizing the operation of A as transfomrming all vectors from one place to another. If the rank of a $m \times n$ matrix A is not maximal, then $\text{Im}(A) \neq \mathcal{R}^m$, which means some line is mapped to zero. And there is no inverse since no matrix can undo the damage and map the zero vectors back to some finite vector.

Property (H2) almost becomes trivial: if you put on socks and then shoes, the first to be taken off are the shoes.

At this stage, only inverse of square matrix interests us, because it has the extra nice property that the left inverse and the right inverse are the same .

3.5 Linear Maps $f : V \to W$

We start with a linear map $f : V \to W$ over $\mathbb{F}$ with dimensions n and m, respectively.

[9]The proof is given in section A.9 as always.

We introduce a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V and a basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ 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 $\boldsymbol{\alpha} = \begin{pmatrix}\alpha_1 & \ldots & \alpha_n\end{pmatrix}^T$ and $\mathbf{w} = \sum_{j=1}^m \beta_j \mathbf{w}_j$ with coordinate vectors $\boldsymbol{\beta} = \begin{pmatrix}\beta_1 & \ldots & \beta_m\end{pmatrix}^T$, respectively.

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

The situation so for can be summarized by the diagram

$$\begin{array}{ccc} V & \xrightarrow{\ f\ } & W \\ {\scriptstyle\varphi}\uparrow & & \uparrow{\scriptstyle\psi} \\ F^n & \xrightarrow{A=?} & F^m \end{array} \tag{3.44}$$

Abstractly, the matrix A is given by

$$A = \psi^{-1} \circ f \circ \varphi. \tag{3.45}$$

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 $\mathbb{F}^n$, so

$$\begin{aligned} A\mathbf{e}_j = \psi^{-1} \circ f \circ \varphi(\mathbf{e}_j) = \psi^{-1} \circ f(\mathbf{v}_j) = \psi^{-1}\left(\sum_{i=1}^m a_{ij}\mathbf{w}_i\right) \\ = \sum_{i=1}^m a_{ij}\psi^{-1}(\mathbf{w}_i) = \sum_{i=1}^m a_{ij}\tilde{\mathbf{e}}_i. \end{aligned} \tag{3.46}$$

where $\tilde{\mathbf{e}}_i$ are the standard unit vectors in $\mathbb{F}^m$. 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 $\mathbf{e}_i \in \mathbb{F}^n$).

Lemma 3.5.1 (Essence of Linear Algebra)**.** *Let $f : V \to W$ be a linear map, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ a basis of V and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 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*

$$f(\mathbf{v}_j) = \sum_{i=1}^m a_{ij}\mathbf{w}_i. \tag{3.47}$$

We have $\operatorname{rank}(A)$ *= rank(f) and, in particular, A is invertible if and only if f is.*[10]

[10]The proof of this claim is provided in section A.11.

3.6 Change of Basis

We consider a linear map $f : V \to V$ from a vector space to itself and we would like to change the basis in both vector spaces. The two sets of basis vectors, coordinate maps, coordinate vectors and the representing matrices are then denoted by

$$\begin{array}{llll} \text{basis of } V & \text{coordinate map} & \text{coordinate vector} & \text{representing matrix} \\ \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n & \varphi(\boldsymbol{\alpha}) = \sum_{i=1}^n \alpha_i \boldsymbol{v}_i & \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T & A = \varphi^{-1} \circ f \circ \varphi \\ \boldsymbol{v}'_1, \ldots, \boldsymbol{v}'_n & \varphi'(\boldsymbol{\alpha}') = \sum_{i=1}^n \alpha'_i \boldsymbol{v}'_i & \boldsymbol{\alpha}' = (\alpha'_1, \ldots, \alpha'_n)^T & A' = \varphi'^{-1} \circ f \circ \varphi' \end{array} \tag{3.48}$$

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

$$\begin{aligned} A' &= \varphi'^{-1} \circ f \circ \varphi' = \varphi'^{-1} \circ \varphi \circ \varphi^{-1} \circ f \circ \varphi \circ \varphi'^{-1} \circ \varphi' \\ &= \underbrace{\varphi'^{-1} \circ \varphi}_{P} \circ \underbrace{\varphi^{-1} \circ f \circ \varphi}_{A} \circ \underbrace{\varphi^{-1} \circ \varphi'}_{P^{-1}} = PAP^{-1}. \end{aligned} \tag{3.49}$$

When acting the above equation on a primed coordinate vector $\boldsymbol{\alpha'}$, the first thing we obtain on the RHS is $P^{-1}\boldsymbol{\alpha'}$, which is the corresponding unprimed coordinate vector on which the matrix A on the RHS can sesibly act, converting it into another unprimed coordinate vector, representing the action of the linear map. The final action of P converts this back into a primed coordinate vector. Altogether, this is the action of the matrix A' on $\boldsymbol{\alpha'}$.

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

$$\boldsymbol{\alpha'} = \varphi'^{-1}(\mathbf{v}) = \varphi'^{-1} \circ \varphi(\boldsymbol{\alpha}) = P\boldsymbol{\alpha}. \tag{3.50}$$

Hence, P converts unprimed coordinate vectors $\boldsymbol{\alpha}$ into the corresponding primed coordinate vector $\boldsymbol{\alpha'}$. In other words, the columns of P^{-1} are the new basis expressed in terms of the old basis.

Another convention that we may use in later texts is to use $A' = P^{-1}AP$, but then $\boldsymbol{\alpha}' = P^{-1}\boldsymbol{\alpha}$.

Example: Change of Basis of Linear Map $f : \mathbb{R}^2 \to \mathbb{R}^2$.

Question: Consider the linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ represented by the matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$. By default, the matrix is described relative to the standard unit vectors $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Now determine the matrix A' which describes the same linear map relative to the basis $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ in both domain and co-domain.

Solution: The images of the basis vectors under A are

$$A\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -1\mathbf{v}_1 - 2\mathbf{v}_2 \text{ and } A\mathbf{v}_2 = \begin{pmatrix} -1 \\ -2 \end{pmatrix} = -1\mathbf{v}_1 + 0\mathbf{v}_2. \tag{3.51}$$

Using the lemma 3.5.1, we arrange the coefficients from $A\mathbf{v}_1$ into the first column amd the coefficients from $A\mathbf{v}_2$ into the second column we get the matrix

$$A' = \begin{pmatrix} -1 & -1 \\ -2 & 0 \end{pmatrix}, \tag{3.52}$$

which is the matrix representing the linear map f relative to the basis $\mathbf{v}_1$ and $\mathbf{v}_2$. Alternatively, we can figure out the change of basis matrix P from the relation between the coordinates vector

$$\alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \alpha_1'\mathbf{v}_1 + \alpha_2'\mathbf{v}_2 = \begin{pmatrix} \alpha_1' - \alpha_2' \\ 2\alpha_1' + \alpha_2' \end{pmatrix}. \tag{3.53}$$

So we have

$$P^{-1} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} \implies A' = PAP^{-1} = \begin{pmatrix} -1 & -1 \\ -2 & 0 \end{pmatrix}. \tag{3.54}$$

Example: Change of Basis of Linear Map $f : \mathbb{R}^3 \to \mathbb{R}^2$.

Question: Consider the linear map $f : \mathbb{R}^3 \to \mathbb{R}^2$ represented by the matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ which project a 3-dimensional vector into the xy-plane. By default, the matrix is described relative to the standard unit vectors $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Now determine the matrix A' which describes the same linear mape relative to the basis $\mathbf{e}_1$ and $\mathbf{e}_2$ in the domain and $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ in the co-domain.

Solution: The first method is to use lemma 3.5.1, and compute the images of the basis vectors in the domain $\mathbf{e}_1$ and $\mathbf{e}_2$ in terms of the basis vectors in the co-domain $\mathbf{v}_1$ and $\mathbf{v}_2$

$$A\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}(\mathbf{v}_1 - \mathbf{v}_2), A\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2) \text{ and } A\mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.55}$$

Arranging the coefficients into columns, we have the transformed matrix

$$A' = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix}. \tag{3.56}$$

Example: Active and Passive Transformation.

Question: The matrix

$$A = \begin{pmatrix} 1 & -1 & 0 \\ -3 & 0 & 1 \\ 2 & 01 & 1 \end{pmatrix} \tag{3.57}$$

is calculated with respect to the standard basis $\mathbf{e}_1 = (1,0,0), \mathbf{e}_2 = (0,1,0), \mathbf{e}_3 = (0,0,1)$. Find the matrix A' with respect to the basis

$$\mathbf{u}_1 = (1,1,0)_e, \quad \mathbf{u}_2 = (1,0,1)_e, \quad \mathbf{u}_3 = (0,1,1)_e, \tag{3.58}$$

where the subscript e indicate that it is with respect to the standard basis. For example, $\mathbf{u}_1 = (1,1,0)_s = \mathbf{e}_1 + \mathbf{e}_2$. With respect to the $\mathbf{u}$ basis $\mathbf{u}_1 = (1,0,0)_u$ is trivial.
An aritrary vector $\mathbf{x}$ can be represented in the standard basis as $\mathbf{x}_e$ and in the new basis as $\mathbf{x}_u$ and they are related by

$$\mathbf{x}_e = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)\, \mathbf{x}_u \equiv P\mathbf{x}_u. \tag{3.59}$$

To ensure this is not the other way round, we substitute $\mathbf{x}_u = (1,0,0)_u$ and find $\mathbf{x}_e = \mathbf{u}_1$ is indeed true.
A vector transformation in the new basis can then be represented by

$$A'\mathbf{x}_u = A'P^{-1}\mathbf{x}_e = P^{-1}A\mathbf{x}_e \implies A' = P^{-1}AP. \tag{3.60}$$

Solution:

3.7 Systems of Linear Equations

3.7.1 General Structures of Solutions

Consider a linear map $f : V \to W$. We would like to find all solutions $\mathbf{x} \in V$ of the inhomogeneous equation $f(\mathbf{x}) = \mathbf{b}$, where $\mathbf{b} \in W$ is a fixed vector. In other words, which vectors $\mathbf{x} \in V$ get mapped to $\mathbf{b} \in W$. We need to find first the solution to the associated homogenous equation $f(\mathbf{x}) = \mathbf{0}$, where the homogeneous solution is $\mathbf{x} = \ker(f)$.

Lemma 3.7.1. *If $\mathbf{x}_0 \in V$ solves the inhomogenous equation, i.e., $f(\mathbf{x}_0) = \mathbf{b}$, then $\mathbf{x}_0 + \ker(f)$ is the general solution of the inhomogenous equation.*

Proof. $f(\mathbf{x}_0 + \ker(f)) = f(\mathbf{x}_0) + f(\ker(f)) = f(\mathbf{x}_0) = \mathbf{b}$. □

Since every linear map f can be represented by a matrix A, this is equivalent to solving the matrix equation $A\mathbf{x} = \mathbf{b}$, where A is a $m \times n$ matrix and $\mathbf{x}$ and $\mathbf{b}$ are n-dimensional and m-dimensional column vectors respectively.

That is, we want to find the linear combinations (represented by entries in $\mathbf{x}$) of the

column vectors in A that produces the column vector $\mathbf{b}$. In other words, we want to find a vector $\mathbf{x}$ which gets mapped to $\mathbf{b}$ after the transformation. If we expand the matrix equation, we have m linear equations with n variables. The solutions can be no solution, an unique solution, or infinitely many solutions. The two latter case can be viewed as one category (there exists a solution with no or some free parameters).

To distinguish the cases of the general structure of the solution, it is extremely helpful to visualize that the action of A as transformating all vectors from one place to another, and the columns of A are the new basis (in terms of the old basis). The rank of the matrix is defined as the dimension of the image of A, which is the dimension of the vector space spanned by the column vectors of A.

For a general $m \times n$ matrix $A : \mathbb{F}^n \to \mathbb{F}^m$ where $m \neq n$, since the row rank is equals to the column rank, the maximum rank is $\max(\text{rank}(A)) = \min(m, n)$.

If $m > n$, then $\max(\text{rank}(A)) = n < m$; if $m \leq n$, then $\max(\text{rank}(A)) = m \leq n$. Depending on whether $\text{rank}(A) = m$ or $\text{rank}(A) \leq m$, the solution structure behave differently:

1. $\text{rank}(A) = m$: $\mathbf{b} \in \mathbb{F}^m = \text{Im}(A)$, thus we have a solution $\mathbf{x}_0$ for any choice of $\mathbf{b}$ and the general solution is given by $\mathbf{x}_0 + \ker(A)$[11] and the number of free parameters in this solution equals $\dim(\ker(A)) = n - \text{rank}(A) = n - m$.
   (a) $\text{rank}(A) = m = n$: There is no free parameter and we have an unique solution.
   (b) $\text{rank}(A) = m < n$: There is some free parameters and we have infinite number of solutions.
2. $\text{rank}(A) < m$: The columns of A does not span through the whole m-dimensional vector space that it is allowed to (not greater than m since the image of a $m \times n$ matrix is $\mathbb{F}^m$), so depending on $\mathbf{b}$ we have
   (a) $\mathbf{b} \in \text{Im}(A)$: The equations are consistent and we have a solution with $\dim(\ker(A)) = n - \text{rank}(A)$ free parameters. Depending on whether $\text{rank}(A) = n$ or $\text{rank}(A) < n$, we have:
      i. $\text{rank}(A) < m = n$: There is no free parameter and we have an unique solution.
      ii. $\text{rank}(A) < m < n$: There is some free parameters and we have infinite number of solutions.
   (b) $\mathbf{b} \notin \text{Im}(A)$: The equations are inconsistent and there is no solution.

The three relevant case of a square matrix are 1a), 2ai) and 2b.

For example, for a 4×6 matrix A with $\text{rank}(A) = 3$, there can be a solution only when $\mathbf{b} \in \text{Im}(A)$, *i.e.,* $\text{rank}(A' = (A \mid \mathbf{b})) = \text{rank}(A)$, and we expect a line $0 = 0$ at the last row of the augmented matrix, denoting a wasted dimension and there will be $6 - 3 = 3$ free paramters.

As another example, for a 5×3 matrix B with $\text{rank}(B) = 2$, there can be s olution only when $\mathbf{v} \in \mathfrak{Im}((B))$, *i.e.,* $\text{rank}(B' = (B \mid \mathbf{b})) = \text{rank}(B)$, and we expect a line $0 = 0$ at the end with $5 - 2 = 3$ free parameters.

[11] It is important to note that $\ker(A)$ is also a solution to the equation, but in most cases it is ignored.

It is very important to note that the above discussion is under the premise that $\mathbf{b} \neq 0$, *i.e.,* we are finding the particular solution to the inhomogeneous equation, and we have to add the homogeneous solution $\ker(A)$ to yield the complete solution, according to lemma 3.7.1.

As an alternative geometric interpretation, for example, for a system of 3 equations with 3 variables, each equation represents a plane in the 3-dimensional space. In general they intersect at one point in space. However, when the determinant is zero there is a non-trivial solution if the three planes intersect at a line in space, corresponding to a free parameter, or the three planes intersect at a plane in space, corresponding to 2 free parameters, and there is no solution if they do not intersect at all.

3.7.2 Gaussian Elimination

Definition 3.7.1. *The following manipulations of a matrix are called elementary row operations:*[12]

(I1) Exchange two rows.

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

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

Lemma 3.7.2. *The first r row vectors in row echelon form are linearly independent.*[13]

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

$$\operatorname{rank}(A) = r = \text{the number of steps in row echelon form.} \tag{3.61}$$

If every sub square matrices are invertible, then by operations $(I2)$ (and $(I3)$ but it is just a special case of $(I2)$ and is only for simplifying.), we can reduce any matrix A to its row echelon form, where all zero rows (if there is any) are at the bottom and the leading entry of every non-zero row (called the pivot) is on the right of the leading entry of every row above.

magic squares and coding theory

In fact, we can associate any operation of type $(I2)$ to a matrix $E_{(ij)}$, where l_{ij} times i^{th} row is being subtracted from the j^{th} row, with

$$E_{(ij)} = E_{R_j \to R_j - l_{ij} R_i} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & -l_{ij} & & j^{\text{th}} \text{ row} \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & i^{\text{th}} \text{ col.} & & 1 \end{pmatrix}. \tag{3.62}$$

For example, if we would like to subtract row 3 by 2 times row 2, we would need the matrix

[12] Analogous definitions hold for elementary column operations.

[13] The proof of this lemma is provided in section A.10.

$$E_{(23)} = E_{R_3 \to R_3 - 2R_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}. \tag{3.63}$$

The matrix associated with scalar multiplication of a certain row is just a special case where $i = j$.

On the other hand, if some of the submatrices are non-invertible, we would need to permute the rows at the end to avoid zeros in pivots, since elimination of the whole matrix is also elimination of the submatrices at the same time.

We can associate any operation of type $(I1)$ to a permutation matrix P_{ij} to exchange the i^{th} and j^{th} row of the matrix, with

$$P_{(ij)} = P_{R_i \leftrightarrow R_j} = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & & 1 & & & i^{\text{th}} \text{ row} \\ & & & & \ddots & & & & \\ & & & 1 & & 0 & & & j^{\text{th}} \text{ row} \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix} \tag{3.64}$$

.

For example, if we would like to exchange the 2^{nd} and 3^{rd} row, we would need the matrix

$$P_{(23)} = P_{R_2 \leftrightarrow P_3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.65}$$

Usually we would use a single permutation matrix P to permute the rows at once.

To reduce a matrix to its row echelon form U, we continue multiply the row operations matrices on the left hand side. For a 3×3 matrix, that is

$$PE_{32}E_{31}E_{21}A = U \tag{3.66}$$

However, we can also have exchanged the rows at the start, for example, if the second and the thir row are exchanged, then we have

$$P_{23}E_{32}E_{31}E_{21}A = E_{23}E_{21}E_{31}P_{23}A \equiv EPA \equiv L^{-1}PA = U \implies PA = LU \equiv LDU' \tag{3.67}$$

In the last step, we decompose $U = DU'$, where D are the diagonal elements of U and the diagonal elements of U are replaced by 1 to form U', which has a greater symmetry.

The matrix L is a lower traingular matrix satisfying $EL = \mathbb{I}$ by definition, which undoes the row operations. For example, if we have

$$E = E_{32}E_{31}E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -l_{32} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -l_{31} & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -l_{21} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -l_{21} & 1 & 0 \\ l_{32}l_{21} - l_{31} & -l_{32} & 0 \end{pmatrix} \tag{3.68}$$

then the corresponding matrix that undoes the operations will be

$$L = L_{21}L_{31}L_{32} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_{31} & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & l_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 0 \end{pmatrix}, \tag{3.69}$$

which add back the rows in reverse order.

The row echelon form U can be further reduced into the reduced row echelon from R, which contains an identity matrix interspersed by linear combination coefficient of linear depndence columns in A.

In general, a matrix A can be factorized as $A = CR$, where C consists fo all linearly indepedent column vectors in A, so R is arguably more important than U, since it is completely determined from A once C is chosen and by convention we have C as the first set of independent column vectors from left to right. For example, we have

$$A = \begin{pmatrix} 1 & 7 & 3 & 35 \\ 2 & 14 & 6 & 70 \\ 2 & 14 & 9 & 97 \end{pmatrix} = CR = \begin{pmatrix} 1 & 3 \\ 2 & 6 \\ 2 & 9 \end{pmatrix} \begin{pmatrix} 1 & 7 & 0 & 8 \\ 0 & 0 & 1 & 9 \end{pmatrix}, \tag{3.70}$$

which means that there are only two columns in A that are independent, and we have $(7, 14, 14) = 7(1, 2, 2)$ and $(35, 70, 97) = 8(1, 2, 2) + 9(3, 6, 9)$, something that is not trivial at all.

3.7.3 Computing Inverse and Solving $A\mathbf{x} = \mathbf{b}$

To find the inverse of A, we continue to reduce the matrix until it reaches its reduced row echelon form R, for an invertible square $n \times n$ matrix, $R = \mathbb{I}_n$, since

Inverse

$$A \xrightarrow{\text{echelon form}} \begin{pmatrix} a'_{11} & & & * \\ & a'_{22} & & \\ & & \ddots & \\ 0 & & & a'_{nn} \end{pmatrix} \xrightarrow{(I1,I2)} \begin{pmatrix} a'_{11} & & & 0 \\ & a'_{22} & & \\ & & \ddots & \\ 0 & & & a'_{nn} \end{pmatrix} \xrightarrow{(I3)} \begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & & \\ 0 & & & 1 \end{pmatrix} = \mathbb{I}_n \tag{3.71}$$

In matrix terms, we have

$$\mathbb{I}_n = P_k \dots P_1 A \implies A^{-1} = P_k \dots P_1 \mathbb{I}_n. \tag{3.72}$$

This means that we can write A and $\mathbb{I}$ together as a single augmented matrix $A' = (A \mid \mathbb{I})$ and carry out the row operations in parallel to find A^{-1}. In principle we can multiply A^{-1} by $\mathbf{b}$ to find $\mathbf{x}$ but it is not done practically due to its inefficiency.

$A\mathbf{x} = \mathbf{0}$

To find the homogeneous solution to $A\mathbf{x} = \mathbf{0}$, *i.e.,* to find $\ker(A)$, we first factorize $A = CR$, then we choose the dependent coordinates to be the free parameters

The identical technique is used to solve $A\mathbf{x} = \mathbf{b}$ by elimination: reduce the augemnted matrix $A' = (A \mid \mathbf{b})$ so that the answers can be directly read off.

If the equations are consistent, we would either have an actual equation, or something like $0z = 0$, which implies a free parameter (which adds one to $\ker(A)$), which happens when $b \in \mathrm{Im}(A)$ by coincidence. If the equations are not consistent then we would have somthing like $0z = 3$.

4 Determinants

4.1 Definition of a Determinant

4.1.1 Permutations of Elements

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

$$S_n \equiv \{\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\} \mid \sigma \text{ is bijective}\}, \tag{4.1}$$

and this has $n!$ elements. A useful notation for a permutation mapping $1 \to \sigma(1), \ldots, n \to \sigma(n)$.

$$\sigma = \begin{pmatrix} 1 & \ldots & n \\ \sigma(1) & \ldots & \sigma(n) \end{pmatrix}.^{1} \tag{4.2}$$

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 that every permutation can be written as a composition of transpositions, so any $\sigma \in S_n$ can be written as $\sigma = \tau_1 \circ \ldots \circ \tau_k$, where $\tau_1, \ldots, \tau_k \in S_n$ 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

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

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

[1] Note, dispite the similar notation, this is not a matrix.

4.1.2 Leibniz Formula for Determinants

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

(J1) $\det(\ldots, \alpha\mathbf{a} + \beta\mathbf{b}, \ldots) = \alpha \det(\ldots, \mathbf{a}, \ldots) + \beta \det(\ldots, \mathbf{b}, \ldots)$
This means the determinant is linear in each argument.

(J2) $\det(\ldots, \mathbf{a}, \ldots, \mathbf{b}, \ldots) = -\det(\ldots, \mathbf{b}, \ldots, \mathbf{a}, \ldots)$
This means the determinant is completely anti-symmetric.

(J3) $\det(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$
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 $\det(A) \equiv \det(\mathbf{A}^1, \ldots, \mathbf{A}^n)$.

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

To derive an explicit formula for the determinant, we start with a $n \times n$ matrix A with entries a_{ij} whose column vectors we write as linear combinations of the standard unit vectors: $\mathbf{A}^i = \begin{pmatrix} a_{1i} & \cdots & a_{ni} \end{pmatrix}^T = \sum_{j=1}^n a_{ji}\mathbf{e}_j$. We find then

$$\begin{aligned}
\det(\mathbf{A}) = \det\big(\mathbf{A}^1, \ldots, \mathbf{A}^n\big) &= \det\left(\sum_{j_1=1}^n a_{j_1 1}\mathbf{e}_{j_1}, \ldots, \sum_{j_n=1}^n a_{j_n n}\mathbf{e}_{j_n}\right) \\
&\overset{(J1)}{=} \sum_{j_1,\ldots,j_n} a_{j_1 1} \ldots a_{j_n n} \det(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_n}) \overset{j_a=\sigma(a)}{=} \sum_{\sigma \in S_n} a_{\sigma(1)1} \ldots a_{\sigma(n)n} \det\big(\mathbf{e}_{\sigma(1)}, \ldots, \mathbf{e}_{\sigma(n)}\big) \\
&\overset{(J2)}{=} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \ldots a_{\sigma(n)n} \det(\mathbf{e}_1, \ldots, \mathbf{e}_n) \overset{(J3)}{=} \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \ldots a_{\sigma(n)n}
\end{aligned} \tag{4.4}$$

where in the first equality of the second line can be seen by first decomposing the determinant into n terms using $(J1)$ and appplying it to the first column vector $\mathbf{A}^1$. Repeating this step for other column vectors would yield the desired result. In the second step of the second line we replace $\sum_{j_1,\ldots,j_n}$ as $\sum_{\sigma \in S_n}$ since when we sum over all j_i for $i = 1, \ldots, n$, consider the term when $(j_1, ..., j_k, ..., j_n) = (2, \ldots, 6, \ldots, 1)$, since it is a possible permutation of $(1, \ldots, n)$, it is in S_n. For the cases such as $(j_1, ..., j_k, ..., j_n) = (5, \ldots, 5, \ldots, 2)$, where since it is not a possible permutation of $(1, \ldots, n)$, 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

$$\epsilon_{i_1 \ldots i_n} = \begin{cases} +1 & \text{if } i_1, \ldots, i_n \text{ is an even permutation of } 1, \ldots, n, \\ -1 & \text{if } i_1, \ldots, i_n \text{ is an odd permutation of } 1, \ldots, n, \\ 0 & \text{otherwise,} \end{cases} \tag{4.5}$$

the determinant of a matrix A can be written by

$$\det(A) = \epsilon_{i_1 \ldots i_n} a_{i_1 1} \ldots a_{i_n n}, \tag{4.6}$$

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

In three dimensions we find

$$\begin{aligned} \det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} &= \epsilon_{ijk} a_i b_j c_k \\ &= a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_2 b_1 c_3 - a_3 b_2 c_1 - a_1 b_3 c_2 \\ &= \langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \end{aligned} \tag{4.7}$$

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

$$\det \begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix} = a_1 \ldots a_n \tag{4.8}$$

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

For a large matrix, it is usually easier to first convert the matrix in to upper triangular form, by elementary row operations, which then the determinant is just the product of the diagonal entries.

The effect of the operations to the determinant is minimal, 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. This is true because it is equivalent of adding a multiple of one column to another because of $\det\left(A^T\right) = \det(A)$. Thus, $\det\left(\mathbf{A}^1, \ldots, \mathbf{A}^j, \ldots, \mathbf{A}^k + \lambda \mathbf{A}^j, \ldots, \mathbf{A}^n\right) = \det(\mathbf{A}^1, \ldots, \mathbf{A}^n) + \lambda \det(\mathbf{A}^1, \ldots, \mathbf{A}^j, \ldots, \mathbf{A}^j, \ldots, \mathbf{A}^n) = \det(\mathbf{A}^1, \ldots, \mathbf{A}^n)$.

Example: Tranformation Law for the Cross Product.

Question: Prove the identity

$$R(\mathbf{u} \times \mathbf{v}) = \det(R)(R\mathbf{u} \times R\mathbf{v}). \tag{4.9}$$

Solution: Using the Levi-Civita symbol we have

$$\begin{aligned} (R\mathbf{u} \times R\mathbf{v})_i &= \epsilon_{ijk} (R\mathbf{u})_j (Rv)_k = \epsilon_{ijk} R_{jm} u_m R_{kn} v_n \\ &= R_{jm} R_{kn} \epsilon_{ijk} u_m v_n = \det(R) R_{il} \epsilon_{lmn} u_m v_n = \det(R) \left(R(\mathbf{u} \times \mathbf{v})\right)_i . \end{aligned} \tag{4.10}$$

Example: Volumes and Determinants.

Question: Given that $\det(A) = 23$, find λ such that the volume corresponding to λA is the same as the volume corresponding to A^5.

Solution: The volume of the parallelepiped for λA is

$$V = |\det(\lambda A)| = |\lambda|^3 \det(A) = 23|\lambda|^3, \tag{4.11}$$

which is equal to the volume of the parallepiped for A^5, which is

$$V = \left|\det\left(A^5\right)\right| = |\det(A)|^5 = 23^5, \tag{4.12}$$

so we get $\lambda = 23^{4/3}$.

4.2 Properties of Determinants

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

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. (4.4) as

$$A_{\sigma(1)1} \dots A_{\sigma(n)n} = A_{j_1\sigma^{-1}(j_1)} \dots A_{j_n\sigma^{-1}(j_n)} = A_{1\sigma^{-1}(1)} \dots A_{n\sigma^{-1}(n)}. \tag{4.13}$$

From this observation, the determinant can be written as

$$\begin{aligned} \det(A) &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) A_{1\sigma^{-1}(1)} \dots A_{n\sigma^{-1}(n)} = \sum_{\sigma^{-1}\in S_n} \operatorname{sgn}(\sigma^{-1}) A_{1\sigma^{-1}(1)} \dots A_{n\sigma^{-1}(n)} \\ &\overset{\rho=\sigma^{-1}}{=} \sum_{\rho\in S_n} \operatorname{sgn}(\rho) (A^T)_{\rho(1)1} \dots (A^T)_{\rho(n)n} = \det\left(A^T\right). \end{aligned} \tag{4.14}$$

□

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 $(AB)^j = \sum_k B_{kj}\mathbf{A}^k$. Hence

$$
\begin{aligned}
\det(AB) = \det\big((AB)^1, \ldots, (AB)^n\big) &= \det\left(\sum_{k_1} B_{k_1 1}\mathbf{A}^{k_1}, \ldots, \sum_{k_n} B_{k_n n}\mathbf{A}^{k_n}\right) \\
&\overset{(J1)}{=} \sum_{k_1,\ldots,k_n} B_{k_1 1}\ldots B_{k_n n}\det\big(\mathbf{A}^{k_1}, \ldots, \mathbf{A}^{k_n}\big) \\
&\overset{k_a=\sigma(a)}{=} \sum_{\sigma\in S_n} B_{\sigma(1)1}\ldots B_{\sigma(n)n}\det\big(\mathbf{A}^{\sigma(1)}, \ldots, \mathbf{A}^{\sigma(n)}\big) \\
&\overset{(J2)}{=} \underbrace{\sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) B_{\sigma(1)1}\ldots B_{\sigma(n)n}}_{\det(B)} \underbrace{\det\big(\mathbf{A}^1, \ldots, \mathbf{A}^n\big)}_{\det(A)} = \det(A)\det(B).
\end{aligned}
\tag{4.15}
$$

□

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

$$
A \text{ is bijective (i.e., } A \text{ has an inverse)} \iff \det(A) \neq 0 \tag{4.16}
$$

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 linaer map, since

$$
\det\big(PAP^{-1}\big) = \det(P)\det(A)\det(P)^{-1} = \det(A). \tag{4.17}
$$

4.3 Laplace's Expansion of Determinants

The co-factor of a $n \times n$ matrix A is defined as

$$
C_{ij} \equiv \det\Big(\tilde{A}_{(i,j)}\Big) = (-1)^{i+j}\det\big(A_{(i,j)}\big), \tag{4.18}
$$

where $\tilde{A}_{(i,j)}$ is an associated matrix of A with i^{th} row and j^{th} column changed to all zeros except for the overlapped entries where it is changed to 1.

$$
\tilde{A}_{(i,j)} = \begin{pmatrix}
 & & 0 & & \leftarrow j^{\text{th}}\, col. & \\
 & \text{“}A\text{”} & \vdots & & \text{“}A\text{”} & \\
 & & 0 & & & \\
0 & \ldots\ 0 & 1 & 0 & 0 & 0 \\
 & & 0 & & & \\
 & \text{“}A\text{”} & \vdots & & \text{“}A\text{”} & \\
 & & 0 & & &
\end{pmatrix} \leftarrow i^{\text{th}}\ \text{row} \tag{4.19}
$$

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

Lemma 4.3.1. *For an $n\times n$ matrix A with associated co-factor matrix C, we have*

$$C^T A = \det(A)\mathbb{I}. \tag{4.20}$$

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

why the second line is true

$$\begin{aligned}(C^T A)_{ij} &= \sum_k (C^T)_{ik}A_{kj} = \sum_k A_{kj}C_{ki} = \sum_k A_{kj}\det\left(\tilde{A}_{(k,i)}\right)\\ &= \sum_k A_{kj}\det\left(\mathbf{A}^1,\dots,\mathbf{A}^{i-1},\mathbf{e}_k,\mathbf{A}^{i+1},\dots,\mathbf{A}^n\right)\\ &= \det\left(\mathbf{A}^1,\dots,\mathbf{A}^{i-1},\sum_k A_{kj}\mathbf{e}_k,\mathbf{A}^{i+1},\dots,\mathbf{A}^n\right)\\ &= \det\left(\mathbf{A}^1,\dots,\mathbf{A}^{i-1},\mathbf{A}^j,\mathbf{A}^{i+1},\dots,\mathbf{A}^n\right) \overset{(5.1)}{=} \delta_{ij}\det(A) = (\det(A)\mathbb{I})_{ij}\end{aligned} \tag{4.21}$$

□

An immediate conclusion from lemma 4.3.1 is

$$\det(A) = (C^T A)_{jj} = \sum_i (C^T)_{ji}A_{ij} = \sum_i C_{ij}A_{ij} = \sum_i (-1)^{i+j}A_{ij}\det\left(A_{(i,j)}\right). \tag{4.22}$$

This is known as the Laplace expansion of the determinant. Note that we can choose any column j and computet 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 4.3.1 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

$$A^{-1} = \frac{C^T}{\det(A)}. \tag{4.23}$$

4.4 Cramer's Rule

To solve the equation $A\mathbf{x} = \mathbf{b}$ (A here must be a square matrix) with determinants, we first define the matrices

$$B_{(i)} \equiv (\mathbf{A}^1,\dots,\mathbf{A}^{i-1},\mathbf{b},\mathbf{A}^{i+1},\dots,\mathbf{A}^n), \tag{4.24}$$

where their determinants are

$$\begin{aligned}\det\big(B_{(i)}\big) &= \det\big(\mathbf{A}^1,\ldots,\mathbf{A}^{i-1},\mathbf{b},\mathbf{A}^{i+1},\ldots,\mathbf{A}^n\big) = \det\left(\mathbf{A}^1,\ldots,\mathbf{A}^{i-1},\sum_j x_j\mathbf{A}^j,\mathbf{A}^{i+1},\ldots,\mathbf{A}^n\right)\\ &= \sum_j x_j \det\big(\mathbf{A}^1,\ldots,\mathbf{A}^{i-1},\mathbf{A}^j,\mathbf{A}^{i+1},\ldots,\mathbf{A}^n\big) = x_i \det\big(\mathbf{A}^1,\ldots,\mathbf{A}^{i-1},\mathbf{A}^i,\mathbf{A}^{i+1},\ldots,\mathbf{A}^n\big)\\ &= x_i \det(A),\end{aligned} \tag{4.25}$$

where we have used eq. (3.23) to express $\mathbf{b} = \sum_j x_j\mathbf{A}^j$.

Thus

$$x_i = \frac{\det\big(B_{(i)}\big)}{\det(A)} = \frac{\det(\mathbf{A}^1,\ldots,\mathbf{A}^{i-1},\mathbf{b},\mathbf{A}^{i+1},\ldots,\mathbf{A}^n)}{\det(A)}. \tag{4.26}$$

Example: Gram Determinant.

Question: The Gram determinant of three real vectors $\mathbf{a},\mathbf{b}$ and $\mathbf{c}$ is defined as

$$G(\mathbf{a},\mathbf{b},\mathbf{c}) = \begin{vmatrix}\mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} & \mathbf{a}\cdot\mathbf{c}\\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{c}\\ \mathbf{c}\cdot\mathbf{a} & \mathbf{c}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c}\end{vmatrix}. \tag{4.27}$$

Show that

$$G(\mathbf{a},\mathbf{b},\mathbf{c}) = \begin{pmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{pmatrix} \tag{4.28}$$

and that if $\mathbf{a},\mathbf{b}$ and $\mathbf{c}$ are linearly independent their Gram determinant must be strictly positive.

Solution: We notice that

$$\begin{pmatrix}\mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} & \mathbf{a}\cdot\mathbf{c}\\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{c}\\ \mathbf{c}\cdot\mathbf{a} & \mathbf{c}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c}\end{pmatrix} = \begin{pmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{pmatrix}\begin{pmatrix}a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{pmatrix}. \tag{4.29}$$

Therefore

$$G(\mathbf{a},\mathbf{b},\mathbf{c}) = \begin{vmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{vmatrix}\begin{vmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{vmatrix}^T = \begin{pmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{pmatrix}. \tag{4.30}$$

If $\mathbf{a},\mathbf{b}$ and $\mathbf{c}$ are linearly independent then $V = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \det(\mathbf{a},\mathbf{b},\mathbf{c}) \neq 0$, therefore $G > 0$.

5 Scalar Products

5.1 Real and Hermitian Scalar Products

Definition 5.1.1 (Scalar Products)**.** *A real (hermitian) scalar product on a vector space V over $F = \mathbb{R}$ ($F = \mathbb{C}$) is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ ($\mathbb{C}$) 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 α, $\beta \in F$.

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 $F = \mathbb{C}$).

A real or hermitian scalar product is also referred to as an inner product on V and a vector space V with such a scalar product is also called an inner product (vector) space.

Definition 5.1.2 (Norms and Normed Vector Spaces)**.** *A norm $\|\cdot\|$ on a vector space V over the field $F = \mathbb{R}$ or $\mathbb{C}$ is a map $\|\cdot\| : V \to \mathbb{R}$ which satisfies*

(L1) $\|\mathbf{v}\| > 0$ *for all non-zero* $\mathbf{v} \in V$,

(L2) $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$ *for all* $\alpha \in F$ *and all* $\mathbf{v} \in V$,

A vector space V with a norm is called a normed vector space.

Note that the notation $|\alpha|$ in $(L2)$ refers to the simple absolute value for $F = \mathbb{R}$ and complex modulus for $F = \mathbb{C}$. In fact, the complex modulus in $\mathbb{C}$ is simply a generalization of the absolute value in $\mathbb{R}$.

Since properties $(K1)$ and $(K3)$ imply that $\langle \mathbf{v}, \mathbf{v} \rangle$ is real and non-negative, respectively, we can sensibly define the norm (or length) of a vector as

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}. \tag{5.1}$$

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

$$\cos(\sphericalangle(\mathbf{v},\mathbf{w})) \equiv \frac{\langle \mathbf{v},\mathbf{w}\rangle}{\|\mathbf{v}\|\|\mathbf{w}\|}. \tag{5.2}$$

The Cauchy-Schwarz inequality and the triangle inequality are

$$|\langle \mathbf{v},\mathbf{w}\rangle| \le \|\mathbf{v}\|\|\mathbf{w}\| \quad \text{and} \quad |\mathbf{v}+\mathbf{w}| \le \|\mathbf{v}\| + \|\mathbf{w}\|, \tag{5.3}$$

More equality from riley

where the first inequality comes directly from the definition of angle above as the cosine function cannot exceed 1, and the second from the fact that $(\mathbf{v}+\mathbf{w})^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2\|\mathbf{v}\|\|\mathbf{w}\|$.

Examples of scalar products include:

1. The standard scalar product for orthogonal basis in $\mathbb{R}^n$: $\langle \mathbf{v},\mathbf{w}\rangle = \mathbf{v}^T\mathbf{w} = \sum_{i=1}^n v_i w_i$.
2. The standard scalar product for orthogonal basis in $\mathbb{C}^n$: $\langle \mathbf{v},\mathbf{w}\rangle = \mathbf{v}^\dagger\mathbf{w} = \sum_{i=1}^n v_i^* w_i$.
3. The standard scalar product for general basis in $\mathbb{R}^n$: $\langle \mathbf{v},\mathbf{w}\rangle = \mathbf{v}^T G\mathbf{w} = \sum_{i=1}^n \sum_{j=1}^n v_i G_{ij}\mathbf{w}_j$, where $G_{ij} = \langle \mathbf{e}_i,\mathbf{e}_j\rangle$, which equals to $\mathbb{I}$ for orthonormal basis.
4. The standard scalar product for general basis in $\mathbb{C}^n$: $\langle \mathbf{v},\mathbf{w}\rangle = \mathbf{v}^\dagger G\mathbf{w} = \sum_{i=1}^n \sum_{j=1}^n v_i^* G_{ij}\mathbf{w}_j$, where $G_{ij} = \langle \mathbf{e}_i,\mathbf{e}_j\rangle$, which equals to $\mathbb{I}$ for orthonormal basis.
5. The Minkowski product in $\mathbb{R}^4$: $\langle \mathbf{v},\mathbf{w}\rangle = \mathbf{v}^T\eta\mathbf{w} = -v_0w_0 + v_1w_1 + v_2w_2 + v_3w_3$ where $\eta = \mathrm{diag}(-1,1,1,1)$. Note that the Minkowski product does not satisfies property $(K3)$, for example, for $\mathbf{v} = \begin{pmatrix}-1 & 1 & 1 & 1\end{pmatrix}^T$, $\langle \mathbf{v},\mathbf{v}\rangle = -1$.
6. The scalar product for (real or complex-valued) functions $f : [a,b] = \mathbb{R}$ or $\mathbb{C}$ on an interval $[a,b] \subset \mathbb{R}$: $\langle f,g\rangle \equiv \int_a^b f(x)^* g(x)dx$.
7. The scalar product for real matrices: $\langle A,B\rangle \equiv \mathrm{Tr}\big(A^T B\big) = \sum_{i,j} A_{ij}B_{ij}$.
8. The scalar product for complex matrices: $\langle A,B\rangle \equiv \mathrm{Tr}\big(A^\dagger B\big) = \sum_{i,j} A_{ij}^* B_{ij}$.

5.2 Orthonormal Basis

Lemma 5.2.1. *Any two vectors* $\mathbf{v}$ *and* $\mathbf{w}$ *are called orthogonal if* $\langle \mathbf{v},\mathbf{w}\rangle = 0$. *Pairwise orthogonal and non-zero vectors* $\mathbf{v}_1,\ldots,\mathbf{v}_k$ *are linearly independent.*

Proof. Start with $\sum_{i=1}^k \alpha_i\mathbf{v}_i = 0$ and take the scalar product of this equation with one of the vectors, say $\mathbf{v}_j$. Since $\langle \mathbf{v}_i,\mathbf{v}_j\rangle = 0$ for all $i \ne j$ it follows that $\alpha_j|\mathbf{v}_j|^2 = 0$. Since $\mathbf{v}_j \ne 0$ its norm is positive, $|\mathbf{v}_j| > 0$, so $\alpha_j = 0$.

□

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

$$\langle \boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j \rangle = \delta_{ij}, \tag{5.4}$$

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

Examples of orthonormal basis include:

1. The basis of standard unit vectors, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of $\mathbb{R}^n$ with respect to the standard scalar product in $\mathbb{R}^n$.
2. The basis of standard unit vectors, $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of $\mathbb{C}^n$ with respect to the standard scalar product in $\mathbb{C}^n$.
3. The vectors $\boldsymbol{\epsilon}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{\epsilon}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ with respect to the standard scalar product in $\mathbb{R}^n$.
4. The vectors $\boldsymbol{\epsilon}_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ i \end{pmatrix}$ and $\boldsymbol{\epsilon}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2i \end{pmatrix}$ with respect to the standard scalar product in $\mathbb{C}^n$.
5. The matrices $E_{(ij)}$ defined in eq. (1.35) of the fector space of real $n \times n$ matrices with respect to the scalar product for real matrices.

An orthonormal 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 orthonormal basis $\{\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n\}$. 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 orthonormal basis, we can just take the scalar product of this equation with $\boldsymbol{\epsilon}_j$, leading to

$$\langle \boldsymbol{\epsilon}_j, \mathbf{v} \rangle = \left\langle \boldsymbol{\epsilon}_j, \sum_{i=1}^n \alpha_i \boldsymbol{\epsilon}_i \right\rangle = \sum_{i=1}^n \alpha_i \langle \boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_i \rangle = \sum_{i=1}^n \alpha_i \delta_{ij} = \alpha_j. \tag{5.5}$$

5.3 Gram-Schmidt Procedure

Theorem 5.3.1 (Gram-Schmidit Procedure)**.** *If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of vector space V, then there exists an orthonormal basis $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ of V such that Span$(\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_k)$ = Span$(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ for all $k = 1, \ldots, n$.*

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

$$\boldsymbol{\epsilon}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}. \tag{5.6}$$

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

$$\mathbf{v}'_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{\epsilon}_i, \mathbf{v}_k \rangle \boldsymbol{\epsilon}_i \implies \boldsymbol{\epsilon}_k = \frac{\mathbf{v}'_k}{|\mathbf{v}'_k|}. \tag{5.7}$$

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

$$\langle \boldsymbol{\epsilon}_j, \mathbf{v}'_k \rangle = \langle \boldsymbol{\epsilon}_j, \mathbf{v}_k \rangle - \sum_{i=1}^{k-1} \langle \boldsymbol{\epsilon}_i, \mathbf{v}_k \rangle \langle \boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_i \rangle = \langle \boldsymbol{\epsilon}_j, \mathbf{v}_k \rangle - \langle \boldsymbol{\epsilon}_j, \mathbf{v}_k \rangle = 0. \tag{5.8}$$

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

□

Example: Gram-Schmidt Procedure for Polynomials.

Question: Consider the vector space of quadratic polynomials in one variable $x \in [1, -1]$ with real coefficients and a scalar product defined by $\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) dx$. Find an orthonomral basis for this vector space.

Solution: We start with the standard monomial basis $\mathbf{v}_1 = 1, \mathbf{v}_2 = x, \mathbf{v}_3 = x^2$ (which is not orthonormal since, say, $\int_{-1}^{1} (1)(x) dx \neq 0$). To find an orthonormal basis, we apply the Gram-Schmidt procedure

1. To find $\boldsymbol{\epsilon}_1$:
$$\langle \mathbf{v}_1, \mathbf{v}_1 \rangle = \int_{-1}^{1} dx = 2, \quad \boldsymbol{\epsilon}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \frac{1}{\sqrt{2}}. \tag{5.9}$$
2. To find $\boldsymbol{\epsilon}_2$ first compute $\mathbf{v}'_2$
$$\langle \boldsymbol{\epsilon}_1, \mathbf{v}_2 \rangle = \int_{-1}^{1} dx \, \frac{x}{\sqrt{2}} = 0, \quad \mathbf{v}'_2 = \mathbf{v}_2 - \langle \boldsymbol{\epsilon}_1, \mathbf{v}_2 \rangle \boldsymbol{\epsilon}_1 = x, \tag{5.10}$$
and then normalize
$$\langle \mathbf{v}'_2, \mathbf{v}'_2 \rangle = \int_{-1}^{1} dx \, x^2 = \frac{2}{3}, \quad \boldsymbol{\epsilon}_2 = \frac{\mathbf{v}'_2}{|\mathbf{v}'_2|} = \sqrt{\frac{3}{2}}\, x. \tag{5.11}$$
3. To find $\boldsymbol{\epsilon}_3$ first compute $\mathbf{v}'_3$
$$\begin{gathered} \langle \boldsymbol{\epsilon}_1, \mathbf{v}_3 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^{1} dx \, x^2 = \frac{\sqrt{2}}{3}, \quad \langle \boldsymbol{\epsilon}_2, \mathbf{v}_3 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^{1} dx \, x^3 = 0, \\ \mathbf{v}'_3 = \mathbf{v}_3 - \langle \boldsymbol{\epsilon}_1, \mathbf{v}_3 \rangle \boldsymbol{\epsilon}_1 - \langle \boldsymbol{\epsilon}_2, \mathbf{v}_3 \rangle \boldsymbol{\epsilon}_2 = x^2 - \frac{1}{3} \end{gathered} \tag{5.12}$$
and normalize
$$\langle \mathbf{v}'_3, \mathbf{v}'_3 \rangle = \int_{-1}^{1} dx \left(x^2 - \frac{1}{3} \right)^2 = \frac{8}{45}, \quad \boldsymbol{\epsilon}_3 = \frac{\mathbf{v}'_3}{|\mathbf{v}'_3|} = \sqrt{\frac{5}{8}}(3x^2 - 1). \tag{5.13}$$

So, in summary, the orthonormal polynomial basis is

$$\boldsymbol{\epsilon}_1 = \frac{1}{\sqrt{2}}, \quad \boldsymbol{\epsilon}_2 = \sqrt{\frac{3}{2}}\, x, \quad \boldsymbol{\epsilon}_3 = \sqrt{\frac{5}{8}}(3x^2 - 1). \tag{5.14}$$

These are the first three of an infinite family of orthonormal polynomials, referred to as Legendre polynomials.

5.4 Matrix Elements of Linear Maps

To see how an orthonormal basis helps simplify things, we start with two vectors $\mathbf{v} = \sum_i \alpha_i \boldsymbol{\epsilon}_i, \alpha_i = \langle \boldsymbol{\epsilon}_i, \mathbf{v} \rangle$ and $\mathbf{w} = \sum_i \beta_i \boldsymbol{\epsilon}_i, \beta_i = \langle \boldsymbol{\epsilon}_i, \mathbf{w} \rangle$ and compute their scalar product

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i,j} \alpha_i^* \beta_j \langle \boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j \rangle = \sum_i \alpha_i^* \beta_i = \sum_i \langle \mathbf{v}, \boldsymbol{\epsilon}_i \rangle \langle \boldsymbol{\epsilon}_i, \mathbf{w} \rangle. \tag{5.15}$$

Suppose we would like to compute the representing matrix A of a linear map $f : V \to V$ relative to an orthonormal basis $\{\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n\}$ of V. Following lemma 3.5.1, the entries a_{ij} of the matrix A can be obtained from

$$f(\boldsymbol{\epsilon}_j) = \sum_i a_{ij} \boldsymbol{\epsilon}_i. \tag{5.16}$$

Taking the scalar product of this equation with $\boldsymbol{\epsilon}_k$, we thus obtain a simple formula for converting a linear map f to its representing matrix A

$$a_{ij} = \langle \boldsymbol{\epsilon}_i, f(\boldsymbol{\epsilon}_j) \rangle, \tag{5.17}$$

which are the matrix elements of the linear map f.

Lemma 5.4.1. *If two linear maps $f : V \to V$ and $g : V \to V$ 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 $f = g$, i.e., a linear map is uniquely determined by its matrix elements.*[1]

Example: Matrix Elements of a Linear Map

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

$$f(\mathbf{v}) = (\mathbf{n} \cdot \mathbf{v})\mathbf{n}. \tag{5.18}$$

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

[1] The proof for this trivial fact is given in section A.12.

Solution: Using eq. (5.16), we have

$$A_{ij} = \mathbf{e}_i \cdot f(\mathbf{e}_j) = (\mathbf{n} \cdot \mathbf{e}_i)(\mathbf{n} \cdot \mathbf{e}_j) = n_i n_j. \tag{5.19}$$

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

$$A = \begin{pmatrix} n_1^2 & n_1 n_2 & n_1 n_3 \\ n_1 n_2 & n_2^2 & n_2 n_3 \\ n_1 n_3 & n_2 n_3 & n_3^2 \end{pmatrix}. \tag{5.20}$$

5.5 Perpendicular Spaces

The perpendicular space $W^\perp$ is defined as

$$W^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{w}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{w} \in W\}. \tag{5.21}$$

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.

Lemma 5.5.1. *For a sub vector space $W \subset V$ of a finite dimensional vector space V with a scalar product:*[2]

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

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

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

5.6 Adjoint Linear Maps

Definition 5.6.1. *For a linear map $f : V \to V$ on a vector space V with scalar product, an adjoint linear map, $f^\dagger : V \to V$ is a map satisfying*

$$\langle \mathbf{v}, f\mathbf{w} \rangle = \langle f^\dagger \mathbf{v}, \mathbf{w} \rangle \text{ for all } \mathbf{v}, \mathbf{w} \in V. \tag{5.22}$$

Lemma 5.6.1 (Properties of Adjoint Linear Maps)**.** *For the adjoint $f^\dagger$ of the linear map f:*[3]

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

[2]The proofs of these obvious properties are given in section A.13.

[3]The proofs of these not very trivial properties are given in section A.14.

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

Let us now consider two matrices A and B describing f and $f^\dagger$ relative to an orthonormal basis $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ of V, so A and B are given by

$$A_{ij} = \langle \boldsymbol{\epsilon}_i, f(\boldsymbol{\epsilon}_j)\rangle \text{ and } B_{ij} = \langle \boldsymbol{\epsilon}_i, f^\dagger(\boldsymbol{\epsilon}_j)\rangle = \langle f^\dagger(\boldsymbol{\epsilon}_j), \boldsymbol{\epsilon}_i\rangle^* = \langle \boldsymbol{\epsilon}_j, f(\boldsymbol{\epsilon}_i)\rangle^* = A^*_{ji} = A^\dagger_{ij}. \quad (5.23)$$

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

$$\langle \mathbf{v}, A\mathbf{w}\rangle = \mathbf{v}^\dagger A\mathbf{w} = (A^\dagger \mathbf{v})^\dagger \mathbf{w} = \langle A^\dagger \mathbf{v}, \mathbf{w}\rangle. \quad (5.24)$$

Definition 5.6.2. *A linear map* $f : V \to V$ *on a vector space* V *with scalar product is called self-adjoint (or hermitian) if and only if* $f = f^\dagger$ *(or* $A^\dagger = A$*).*

For an abstract example of a self-adjoint linear map, consider the vector space consisting of all differentiable functions $\varphi : [a, b] \to \mathbb{C}$, satisfying $\varphi(a) = \varphi(b)$, with the scalar product $\langle \varphi, \psi\rangle = \int_a^b dx\, \varphi(x)^* \psi(x)$ and the derivative operator $D = -i\frac{d}{dx}$ which defines a linear map on this vector space. Peforming an integration by parts, we find

$$\begin{aligned} \langle \varphi, D\psi\rangle &= -i \int_a^b dx\, \varphi(x)^* \frac{d\psi}{dx}(x) = -i\,[\varphi(x)^*\psi(x)]_a^b + i\int_a^b dx\, \frac{d\varphi}{dx}(x)^*\psi(x) \\ &= \int_a^b dx\, (D\varphi)(x)^*\psi(x) = \langle D\varphi, \psi\rangle. \end{aligned} \quad (5.25)$$

Hence, D is indeed self-adjoint. 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.

5.7 Unitary Maps

5.7.1 Definitions of Unitary Maps

Definition 5.7.1. *Let* V *be a vector space with a real (hermitian) scalar product. A linear map* $f : V \to V$ *is called orthogonal (unitary) if and only if*

$$\langle f(\mathbf{v}), f(\mathbf{w})\rangle = \langle \mathbf{v}, \mathbf{w}\rangle \text{ for all } \mathbf{v}, \mathbf{w} \in V, \quad (5.26)$$

i.e., The scalar product is an invariant before and after the linear map action.

Lemma 5.7.1 (Properties of Unitary Maps)**.** *For any unitary map* f*:*

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

(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. (5.26) can be rewritten as $\langle \mathbf{v}, f^\dagger \circ f(\mathbf{w})\rangle = \langle \mathbf{v}, \mathrm{id}_V(\mathbf{w})\rangle$. It follows that $f^\dagger \circ f = \mathrm{id}_V$.

$(N2)$ 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.

□

In the language of matrices, condition $(N1)$ turns into

$$A^T A = \mathbb{I} \quad \Longleftrightarrow \quad A^{-1} = A^T \quad \Longleftrightarrow \quad \mathbf{A}^i \cdot \mathbf{A}^j = \delta_{ij}, \tag{5.27}$$

where the last step can be shown by $(A^T A)_{ij} = A_{ki} A_{kj} = \mathbf{A}^i \cdot \mathbf{A}^j = \mathbb{I} = \delta_{ij}$.

Taking the determinant of the LHS condition in eq. (5.27) gives

$$1 = \det(\mathbb{I}) = \det\big(AA^T\big) = \det(A)\det\big(A^T\big) = \det(A)^2 \implies \det(A) = \pm 1. \tag{5.28}$$

As we will see, an orthogonal matrix can be used to describe a rotation plus and reflection(s), *i.e.,* $A = R\sum_i F_i$, where $\det(F_i) = -1$ and $\det(R) = 1$.

$$A = FR \tag{5.29}$$

of a rotation R and a reflection F.

In copmlex case, the characterizing conditions becomes

$$A^\dagger A = 1 \iff A^{-1} = A^\dagger \iff (\mathbf{A}^i)^\dagger \cdot \mathbf{A}^j = \delta_{ij} \tag{5.30}$$

Taking the determinant we conclude that

$$1 = \det(\mathbb{I}) = \det\big(A^\dagger A\big) = \det(A)^* \det(A) = |\det(A)|^2 \implies |\det(A)| = 1. \tag{5.31}$$

Example: Two Sets of Orthonormal Basis.

Question: The two sets of orthonormal basis $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$ and $(\mathbf{e}'_1, \cdots, \mathbf{e}'_n)$ are related by $\mathbf{e}'_j = \sum_{i=1}^n U_{ij}\mathbf{e}_i$. Find U_{ij} and prove that U is unitary.

Solution: The columns of j^{th} column of the matrix U, denoted by $\mathbf{U}^j$, is

just the new basis $\mathbf{e}'_j$, expressed in terms of the old basis $(\mathbf{e}_1, \cdots, \mathbf{e}_n)$, and since the old basis is orthonormal, the i^{th} component of $\mathbf{U}^j$, which is the entry for U_{ij}, is just $\langle \mathbf{e}_i, \mathbf{e}'_j \rangle$.
To prove that U is unitary, we calculate

$$(U^\dagger U)_{ij} = U^*_{ki} U_{kj} = \langle \mathbf{e}'_i, \mathbf{e}_k \rangle \langle \mathbf{e}_k, \mathbf{e}'_j \rangle = \langle \mathbf{e}'_i, \mathbf{e}'_j \rangle = \delta_{ij} \implies U = \mathbb{I}. \tag{5.32}$$

The last equality from the first relation of the above equation can be understand by regarding the RHS as a scalar product of $\mathbf{e}'_i$ and $\mathbf{e}'_j$, where $\langle \mathbf{e}'_i, \mathbf{e}'_j \rangle = \sum_{k=1}^n (\mathbf{e}'_i)^*_k (\mathbf{e}'_j)_k = \sum_{k=1}^n \langle \mathbf{e}'_i, \mathbf{e}_k \rangle \langle \mathbf{e}_k, \mathbf{e}'_j \rangle$.

Example: Exponential of Matrices (1).

Question: Let M be a Hermitian matrix. Prove that the matrix e^{iM} is unitary by writing it as a power series.

Solution: We first compute that hermitian conjugate

$$M^\dagger = \left(\sum_{n=0}^\infty \frac{(iM)^n}{n!} \right)^\dagger = \sum_{n=0}^\infty \frac{(-iM^\dagger)^n}{n!} = \sum_{n=0}^\infty \frac{(-iM)^n}{n!} = e^{-iM}, \tag{5.33}$$

it then follows directly that $M^\dagger M = e^{-iM} e^{iM} = \mathbb{I}$.

Example: Exponential of Matrices (2).

Question: Prove that if $BC = CB$ then $\exp(B)\exp(C) = \exp(B + C)$.

Solution: From the definition of exponential of matrices we have

$$\exp(B+C) = \sum_{m=0}^\infty \frac{(B+C)^m}{m!} = \sum_{m=0}^\infty \frac{1}{m!} \sum_{k=0}^m \binom{m}{k} B^k C^{m-k} = \sum_{m=0}^\infty \sum_{k=0}^m \frac{B^k C^{m-k}}{k!(m-k)!}. \tag{5.34}$$

At this point we have a double sum in the triangular region

$$m = 0, 1, 2 \ldots \quad \text{and for each } m \quad k = 0, 1, \ldots, m. \tag{5.35}$$

Then this is analogous to how we integrate a function over a traigular region bounded by the x-axis and the line $y = x$ and doing the substitution $r = k$ and $s = m - k$, then the double sum becomes

$$\sum_{m=0}^\infty \sum_{k=0}^m \frac{B^k C^{m-k}}{k!(m-k)!} = \sum_{r=0}^\infty \sum_{s=0}^\infty \frac{B^r C^s}{r!s!} = \left(\sum_{r=0}^\infty \frac{B^r}{r!} \right) \left(\sum_{s=0}^\infty \frac{C^s}{s!} \right) = e^B e^C. \tag{5.36}$$

5.7.2 Two Dimensional Rotations

To find the explicit form of two-dimensional rotation matrices, start with a general 2×2 matrix: $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where a, b, c, d are real numbers. We impose the conditions $R^T R = \mathbb{I}$ and $\det(R) = 1$. This gives:

$$R^T R = \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \det(R) = ad - bc = 1. \tag{5.37}$$

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:

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{5.38}$$

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

$$\mathbf{x}' = R\mathbf{x} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}, \tag{5.39}$$

where θ is assumed to be positive if the rotation of the new coordinate system relative to the original one is anti-clockwise. It is also easy to find the inverse of R, simply by replacing θ to $-\theta$, as the original coordiante system rotate by $-\theta$ anti-clockwisely.

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

$$\cos(\sphericalangle(\mathbf{x}', \mathbf{x})) = \frac{\langle \mathbf{x}', \mathbf{x} \rangle}{|\mathbf{x}'||\mathbf{x}|} = \frac{(x\cos\theta - y\sin\theta)x + (x\sin\theta + y\cos\theta)y}{|\mathbf{x}|^2} = \cos\theta. \tag{5.40}$$

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

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2), \tag{5.41}$$

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.

5.7.3 Three-Dimensional Rotations

Since two dimensional rotations are basically three dimensional rotations, but with the axis of rotation implicited fixed, we can represent a two-dimensional rotation about the x-axis as

$$R_1(\theta_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{pmatrix} \tag{5.42}$$

Analogously, the rotations matrices around y and z axes are

$$R_2(\theta_2) = \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{pmatrix}^{4} \text{ and } R_3(\theta_3) = \begin{pmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.43}$$

We can then 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.

Another choice that is used more often in physics is $R(\psi, \theta, \phi) = R_3(\psi)R_1(\theta)R_3(\phi)$. The angles ψ, θ, ϕ in this parametrization are known as 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.

Unlike two-dimensional rotations, rotations in three dimensions do not commute, *i.e.,* in general, $R_1(\theta_1)R_2(\theta_2) \neq R_2(\theta_2)R_1(\theta_1)$.

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 with (relative) angular velocity $\boldsymbol{\omega}$. Mathematically,

$$\mathbf{x} = R(t)\mathbf{y}, \tag{5.44}$$

where $R(t)$ is the time-dependent rotation matrix.[5]

For example, a roation around the z-axis with constant angular speed ω can be written as

$$R(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) & 0 \\ \sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5.45}$$

To understand the relation between the rotation matrix $R(t)$ and the angular velocity $\boldsymbol{\omega}$, we use the defining orthogonal properties of the rotation matrix

$$R(t)^T R(t) = \mathbb{I} \implies R^T\dot{R} + \dot{R}^T R = R^T\dot{R} + (R^T\dot{R})^T \equiv W + W^T = 0. \tag{5.46}$$

(and $\det(R(t)) = 1$) for all times t.

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

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

[5]Rotation matrix is usually written in terms of the rotation angles, however, since the angles are related to time by the angular velocity, we write the rotation matrix as a function of time.

$$W = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \quad \text{or} \quad W_{ij} = \epsilon_{ikj}\omega_k. \tag{5.47}$$

The three independent entries ω_i of this matrix define the angular velocity $\boldsymbol{\omega} = \begin{pmatrix} \omega_1 & \omega_2 & \omega_3 \end{pmatrix}^T$. To see that this makes sense let us work out the matrix W for $R(t)$ in eq. (5.45).

$$W = R^T\dot{R} = \omega \begin{pmatrix} \cos(\omega t) & \sin(\omega t) & 0 \\ -\sin(\omega t) & \cos(\omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\sin(\omega t) & -\cos(\omega t) & 0 \\ \cos(\omega t) & -\sin(\omega t) & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.48}$$

indcating an angular velocity $\boldsymbol{\omega} = \begin{pmatrix} 0 & 0 & \omega \end{pmatrix}^T$, as expected.

In eq. (3.31) 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

$$W\mathbf{b} = \boldsymbol{\omega} \times \mathbf{b} \tag{5.49}$$

This can be directly verified by the index calculation $(W\mathbf{b})_i = W_{ij}b_j = \epsilon_{ikj}\omega_k b_j = (\boldsymbol{\omega} \times \mathbf{b})_i$. This relation can be used to rewrite expressions involving W in terms of the angular velocity $\boldsymbol{\omega}$.

For example, to find the velocity of an object, with velocity $\dot{\mathbf{y}}$ in the rotating system, relative to the stationary coordinate system, we differentiating $\mathbf{x} = R(t)\mathbf{y}$ to get

$$\dot{\mathbf{x}} = R\dot{\mathbf{y}} + \dot{R}\mathbf{y} = R(\dot{\mathbf{y}} + W\mathbf{y}) = R(\dot{\mathbf{y}} + \boldsymbol{\omega} \times \mathbf{y}). \tag{5.50}$$

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

$$\mathbf{F}_R = mR^T\ddot{\mathbf{x}}, \tag{5.51}$$

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

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

$$\dot{\mathbf{x}} = R\dot{\mathbf{y}} + \dot{R}\mathbf{y} \text{ and } \ddot{\mathbf{x}} = R\ddot{\mathbf{y}} + 2\dot{R}\dot{\mathbf{y}} + \ddot{R}\mathbf{y}. \tag{5.52}$$

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

$$m\ddot{\mathbf{y}} = \mathbf{F}_R - 2mR^T\dot{R}\dot{\mathbf{y}} - mR^T\ddot{R}\mathbf{y}. \tag{5.53}$$

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

$$m\ddot{\mathbf{y}} = \mathbf{F}_R - 2mW\dot{\mathbf{y}} - mW^2\mathbf{y} - m\dot{W}\mathbf{y} = \mathbf{F}_R - 2m\boldsymbol{\omega}\times\dot{\mathbf{y}} - \boldsymbol{m}\omega\times(\omega\times\mathbf{y}) - 2m\dot{\boldsymbol{\omega}}\times\mathbf{y}. \quad (5.54)$$

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

5.8 Dual Vector Space

Definition 5.8.1 (Definition of Dual Vector Space)**.** *For a vector space V over F the dual vector space V^* is the set of all linear maps $V \to F$ (where F is seen as a one-dimensional vector space). The elements of V^* are called linear functionals.*

Examples of linear functionals include:

1. For $V = \mathbb{R}^n$ and a fixed vector $\mathbf{w} \in V$ we can define $\varphi_{\mathbf{w}} \in (\mathbb{R}^n)^*$ by

$$\varphi_{\mathbf{w}}(\mathbf{v}) = \mathbf{w}^T\mathbf{v} \in \mathbb{R}. \quad (5.55)$$

   In fact, any linaer map from $\mathbb{R}^n$ to $\mathbb{R}$ can be completely described by a dot product with a unique vector $\mathbf{w}$, this is because we can set $\mathbf{w} = \sum_{i=1}^n \varphi(\mathbf{e}_i)$, where $\mathbf{e}_i$ are the standard basis, then

$$\varphi(\mathbf{v}) = \varphi\left(\sum_{i=1}^n v_i e_i\right) = \sum_{i=1}^n v_i\varphi(\mathbf{e}_i) = \sum_{i=1}^n v_i w_i = \mathbf{w}^T\mathbf{v} = \varphi_{\mathbf{w}}(\mathbf{v}). \quad (5.56)$$

2. For the vector space of continuous functions $h : [a, b] \to \mathbb{R}$ the integral

$$I(h) = \int_a^b h(x)dx \quad (5.57)$$

   is a linear functional. Another intersting functional on the same vector space is $\delta_{x_0}(h) \equiv h(x_0)$, where $x_0 \in [a, b]$ is a fixed point, this is the famous Dirac delta function.

Since linear functionals map vector in $\mathbb{R}^n$ to $\mathbb{R}$, so they are described by $1 \times n$ matrices, that is, by row vectors. So we can think of the vector space V as consisting of column vectors and its dual V^* as consisting of row vectors.

Theorem 5.8.1 (Existence of Dual Basis)**.** *For a basis $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ of V there is a basis $\boldsymbol{\epsilon}_*^1, \ldots, \boldsymbol{\epsilon}_*^n$ of V^*, called the dual basis, such that*

$$\boldsymbol{\epsilon}_*^i(\boldsymbol{\epsilon}_j) = \delta_j^i. \quad (5.58)$$

In particular, $\dim(V^*) = \dim(V)$.

Proof. We claim that the dual basis is

$$\boldsymbol{\epsilon}_*^i \equiv \mathbf{e}_i^T \psi^{-1}(\mathbf{v}), \tag{5.59}$$

where $\psi(\boldsymbol{\alpha}) = \sum_i = \alpha_i \boldsymbol{\epsilon}_i$ is the coordinate map which assigns to a coordinate vector $\boldsymbol{\alpha} = (\alpha^1, \ldots, \alpha^n)$ the corresponding vector, relative to the chosen basis $\boldsymbol{\epsilon}_i$.

First we check

$$\boldsymbol{\epsilon}_*^i(\boldsymbol{\epsilon}_j) = \mathbf{e}_i^T \psi^{-1}(\boldsymbol{\epsilon}_j) = \mathbf{e}_i^T \mathbf{e}^j = \delta_i^j. \tag{5.60}$$

To verify that $\boldsymbol{\epsilon}_*^i$ form a basis we first check linear independence. Applying $\sum_i \beta_i \boldsymbol{\epsilon}_*^i = 0$ to $\boldsymbol{\epsilon}_j$ and using the above equations shows immediately that $\beta_j = 0$.

Next to see that they span V^* we start with an arbitrary functional $\varphi \in V^*$ and a vector $\mathbf{v} = \sum_i v^i \boldsymbol{\epsilon}_i$, then

$$\varphi(\mathbf{v}) = \varphi\left(\sum_i v^i \boldsymbol{\epsilon}_i\right) = \sum_i v^i \varphi(\boldsymbol{\epsilon}_i) = \sum_i \varphi_i v^i = \sum_i \varphi_i \boldsymbol{\epsilon}_*^i(\mathbf{v}). \tag{5.61}$$

This means $\varphi = \sum_i \varphi_i \boldsymbol{\epsilon}_*^i$ so that we have written an arbitrary functional φ as a linear combination of $\boldsymbol{\epsilon}_*^i$. □

Vector space basis elements have lower indices and their coordinates have upper indices while the situation is reversed for dual vectors. This allows us to decide the origin of coordinate vectors simply by the position of their index. The situation is summarized in section 5.8.

	Vectors in V	**Dual Vectors in V^***
Vectors	$\mathbf{v} = v^i\, \boldsymbol{\epsilon}_i$	$\varphi = \varphi_j\, \boldsymbol{\epsilon}_*^j$
Coordinates	v^i	φ_j

The action of dual vectors on vectors can be written as

$$\varphi(\mathbf{v}) = \sum_{i,j} \varphi_i v^j \boldsymbol{\epsilon}_*^i(\boldsymbol{\epsilon}_j) = \varphi_i v^i. \tag{5.62}$$

6 Eigenvectors and Eigenvalues

6.1 Definition of Eigenvectors and Eigenvalues

Definition 6.1.1 (Definition of Eigenvectors and Eigenvalues)**.** *For a linear map $f : V \to V$ on a vector space V over $\mathbb{F}$ the number $\lambda \in F$ is called an eigenvalue of f if there is a non-zero vector $\mathbf{v}$ such that*

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

In this case, $\mathbf{v}$ is called an eigenvector of f with eigenvalue λ, which is only scaled by the action of a linear map.

Example: Eigenvectors of the Inverse Matrix.

Question: A non-singular matrix A has eigenvalues λ_i and eigenvectors $\mathbf{v}_i$. Find the eigenvalues and eigenvectors of the inverse matrix A^{-1}.

Solution: The eigenvalues and eigenvectors of A satisfy

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i \implies A^{-1}A\mathbf{v}_i = \lambda_i A^{-1}\mathbf{v}_i \implies A^{-1}\mathbf{v}_i = \frac{1}{\lambda_i}\mathbf{v}_i, \tag{6.2}$$

thus A^{-1} has the same eigenvectors $\mathbf{v}_i$ as does A but the corresponding eigenvalues are $1/\lambda_i$.

Example: Eigenvalues of a 2×2 matrix.

Question: Find the fastest way to calculate the eigenvalues of a general 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Solution: The trace and the determinant of the matrix is

$$\mathrm{Tr}(A) = a + d = \lambda_1 + \lambda_2 \quad \text{and} \quad \det(A) = ad - bc = \lambda_1 \lambda_2, \tag{6.3}$$

which are conveniently the sum and product of roots of the charateristic polynomial, which we will denote them by s and p from now on. The roots are then

$$\lambda_{1,2} = \frac{s}{2} \pm \sqrt{\left(\frac{s}{2}\right)^2 - p}. \tag{6.4}$$

Lemma 6.1.1 (Properties of Characteristic Polynomials)**.** *The characteristic polynomial* $\chi_A(\lambda) \equiv \det(f - \lambda \mathrm{id}_V) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0$ *of a* $n \times n$ *matrix* A *has the following properties:*

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

(O2) The coeffiecients c_i *of the characteristic polynomial are basis-indepedent.*

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

label items

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 $1, x, x^2, \ldots$ 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

$$\chi_A(\lambda) = \prod_{i=1}^n (A_{ii} - \lambda) + \mathcal{O}(\lambda^{n-2}) = (-1)^n\lambda^n + (-1)^{n-1}\left(\sum_{i=1}^n A_{ii}\right)\lambda^{n-1} + \mathcal{O}(\lambda^{n-2}). \tag{6.5}$$

□

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 $\mathrm{Tr}(A) \equiv \sum_{i=1}^n A_{ii}$ is also basis-independent. However this can be seen more directly by noting that

$$\mathrm{Tr}(AB) = \sum_{i,j} A_{ij}B_{ji} = \sum_{i,j} B_{ji}A_{ij} = \mathrm{Tr}(BA) \implies \mathrm{Tr}\big(PAP^{-1}\big) = \mathrm{Tr}\big(P^{-1}PA\big) = \mathrm{Tr}(A). \tag{6.6}$$

Resulting from the charateristic polynomial of the matrix, there will be n eigenvalues (which are not necessarily distinct) for a $n \times n$ matrix. Distinct eigenvalues are also known as non-degenerate eigenvalues, and repeated eigenvlues are known as degenerate eigenvalues. If there happens to be a m-folded repeated root, then there will be usually m linearly indpendent eigenvector for the same eigenvalue, otherwise the matrix is called a defective matrix, and we will not be able to find n independent eigenvectors for the matrix and thus is not diagonalizable.

6.2 Diagonalization of Matrices

Definition 6.2.1. *We say a linear map $f : V \to V$ 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 a $n \times n$ 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.

Lemma 6.2.1. *A linear map $f : V \to V$ can be diagonalised if and only if 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 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V are obtained from $f(\mathbf{v}_j) = \sum_i a_{ij}\mathbf{v}_i$ as discussed in lemma 3.5.1. From this equation, if $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal, then $f(\mathbf{v}_j) = \lambda_j \mathbf{v}_j$ and the basis vectors $\mathbf{v}_j$ are eigenvectors with eigenvalues λ_j. □

Lemma 6.2.2. *The $n \times n$ matrix A with entries in F can be diagonalized if and only if A has n eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ which form a basis of F^n. In this case, if we define the matrix*

$$P = (\mathbf{v}_1, \ldots, \mathbf{v}_n) \tag{6.7}$$

whose columns are the eigenvectors of A, it follows that

$$P^{-1}AP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \tag{6.8}$$

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

Proof. We assume that we have a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of eigenvectors with eigenvalues λ_i so that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Then

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

where in second equality we used the concept that a matrix is simply a collection of column vectors and matrix multiplication is simply acting the linear map on each of them separately. □

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

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

Theorem 6.2.3 (Simultaneous diagonalization)**.** *Let A, B be two diagonalizable $n \times n$ matrices. Then we have*

$$A, B \text{ can be diagonalized simultaneously} \quad \Longleftrightarrow \quad [A, B] = 0. \tag{6.11}$$

Proof. " $\Longrightarrow$ ": Suppose A, B are simultaneously diagonalizable. Then there exists an invertible P such that

$$\widehat{A} = P^{-1}AP, \tag{6.12}$$

$$\widehat{B} = P^{-1}BP \tag{6.13}$$

are both diagonal. Hence

$$[A, B] = AB - BA = P\widehat{A}\,P^{-1}P\widehat{B}\,P^{-1} - P\widehat{B}\,P^{-1}P\widehat{A}\,P^{-1} = P\big[\widehat{A}, \widehat{B}\big]P^{-1} = 0, \tag{6.14}$$

since diagonal matrices commute.

" $\Longleftarrow$ ": Conversely, assume $[A, B] = 0$ and that A has nondegenerate eigenvalues $\lambda_1, \ldots, \lambda_n$. Since A is diagonalizable, there is a basis of eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ with

$$A\,\mathbf{v}_i = \lambda_i\,\mathbf{v}_i, \quad i = 1, \ldots, n. \tag{6.15}$$

Multiplying B on the left of both sides gives

$$B(A\mathbf{v}_i) = A(B\mathbf{v}_i) = \lambda_i\,(B\mathbf{v}_i), \tag{6.16}$$

so each $B\mathbf{v}_i$ is again an eigenvector of A with eigenvalue λ_i. By nondegeneracy of λ_i, $B\mathbf{v}_i$ must lie in the one-dimensional eigenspace for λ_i, hence

$$B\mathbf{v}_i = \mu_i\,\mathbf{v}_i \tag{6.17}$$

for some scalar μ_i. Thus each v_i is also an eigenvector of B, and $\{\mathbf{v}_i\}$ is a basis of common eigenvectors. It follows that A and B are simultaneously diagonalizable. □

Example: AB and BA share the same eigenvalues.

Question: Prove that λ is an eigenvalue of AB if and only if λ is an eigenvalue of BA.

Solution: If λ is an eigenvalue of AB, then

$$AB\mathbf{v} = \lambda\mathbf{v}. \tag{6.18}$$

Multiplying B on both sides we get

$$BA(B\mathbf{v}) = \lambda B\mathbf{v}, \tag{6.19}$$

so λ is also an eigenvalue of BA.

6.3 Eigenvectors and Eigenvalues of Normal Matrices

Theorem 6.3.1. *Let V be a vector space over $\mathbb{R}(\mathbb{C})$ with real (hermitian) scalar product $\langle\cdot,\cdot\rangle$. If $f : V \to V$ is normal then:*[1]

[1] A more abstract proof for just the hermitian case is given in section A.15.

(P1) *$f^\dagger$ has the same set of eigenvectors as f but the eigenvalues of $f^\dagger$ are the complex conjugates of that of f.*

(P2) *Eigenvectors for different eigenvalues are orthogonal.*

Proof. $(P1)$ We first show that if A is normal then $A - \lambda\mathbb{I}$ is also normal:

$$(A-\lambda\mathbb{I})^\dagger(A-\lambda\mathbb{I}) = (A^\dagger-\lambda^*\mathbb{I})(A-\lambda\mathbb{I}) = A^\dagger A-\lambda A^\dagger\mathbb{I}-\lambda^* A^\dagger\mathbb{I}+\lambda\lambda^*\mathbb{I}^2 = (A-\lambda\mathbb{I})(A-\lambda\mathbb{I})^\dagger. \tag{6.20}$$

Since $(A - \lambda\mathbb{I})\mathbf{v} = 0$, the norm of this vector must be zero

$$((A - \lambda\mathbb{I})\mathbf{v})^\dagger((A - \lambda\mathbb{I})\mathbf{v}) = \mathbf{v}^\dagger(A - \lambda\mathbb{I})^\dagger(A - \lambda\mathbb{I})\mathbf{v} = 0. \tag{6.21}$$

Using the normality of $A - \lambda\mathbb{I}$, we swap the middle terms to get

$$\mathbf{v}^\dagger(A - \lambda\mathbb{I})(A - \lambda\mathbb{I})^\dagger\mathbf{v} = ((A - \lambda\mathbb{I})^\dagger\mathbf{v})^\dagger((A - \lambda\mathbb{I})^\dagger\mathbf{v}) = 0. \tag{6.22}$$

The norm of the vector $(A - \lambda\mathbb{I})^\dagger\mathbf{v}$ is zero, so the vector itself must be zero

$$(A - \lambda\mathbb{I})^\dagger\mathbf{v} = (A^\dagger - \lambda^*\mathbb{I})\mathbf{v} = 0. \tag{6.23}$$

Therefore λ^* are the eigenvalues of the matrix $A^\dagger$.

$(P2)$ For two eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ with eigenvalues λ_1 and λ_2 resepctively, we take the inner product of $\mathbf{v}_2$ with $A\mathbf{v}_1$. For a normal matrix this can be evaluated in two ways:

$$\mathbf{v}_2{}^\dagger(A\mathbf{v}_1) = \lambda_1\mathbf{v}_2{}^\dagger\mathbf{v}_1, \tag{6.24}$$

and

$$\mathbf{v}_2{}^\dagger(A\mathbf{v}_1) = (A^\dagger\mathbf{v}_2)^\dagger\mathbf{v}_1 = (\lambda_2^*\mathbf{v}_2)^\dagger\mathbf{v}_1 = \lambda_2\mathbf{v}_2{}^\dagger\mathbf{v}_1. \tag{6.25}$$

By comparing the two equations we conclude that if $\lambda_1 \neq \lambda_2$, then $\mathbf{v}_2{}^\dagger\mathbf{v}_1 = 0$, which means that $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.

□

For hermitian matrices $(P1)$ implies that all eigenvalues of f are real. For anti-hermitian matrices this means that all eigenvalues of f are imaginary.

This fact can be directly proved by

$$\mathbf{v}^\dagger(A\mathbf{v}) = \lambda\mathbf{v}^\dagger\mathbf{v} = (A^\dagger\mathbf{v})^\dagger\mathbf{v} = \pm(A\mathbf{v})^\dagger\mathbf{v} = \pm\lambda^*\mathbf{v}^\dagger\mathbf{v} \implies \lambda = \pm\lambda^*. \tag{6.26}$$

Example: Eigenvalues of a Unitary Matrix.

Question: Prove that the Eigenvalues of a unitary matrix has unit modulus.

Solution: We take the scalar product of an arbitrary vector $\mathbf{v}$ with itself and find

$$\mathbf{v}^\dagger \mathbf{v} = \mathbf{v}^\dagger A^\dagger A \mathbf{v} = \lambda^* \lambda \mathbf{v}^\dagger \mathbf{v} \implies \lambda\lambda^* = |\lambda|^2 = 1 \tag{6.27}$$

Theorem 6.3.2. *Let V be a vector space over $\mathbb{C}$ with hermitian scalar product $\langle \cdot, \cdot \rangle$. If $f : V \to V$ is normal, it has an orthonormal basis of eigenvectors $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$.*[2]

Proof. If the eigenvalues are distinct (*i.e.,* all of them are non-degenerate), then it follows immediately from theorem 6.3.1 that the eigenvectors are mutually orthogonal and can be made orthonormal by simple normalization. However, if there are degenerate eigenvalues, the degenerate eigenvectors are not necessarily orthogornal but we can choose them to be (or choose any linearly independent eigenvectors and perform the Gram-Schmidt procedure). □

In summary, it means that every normal matrix has an orthonormal basis $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ of eigenvectors with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ (which are real for hermitian matrices) and can be diagonalized through changing its basis to the eigenvectors. From lemma 3.5.1, the columns of the change of basis matrix is simply the new basis expressed in terms of the old basis, so the change of basis matrix is

$$P = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n) \implies P^{-1}AP = P^\dagger AP = \operatorname{diag}(\lambda_1, \ldots, \lambda_n). \tag{6.28}$$

This is because $A(P)$ stretch the column vectors of P by a factor of their eigenvalues, then $P^\dagger(AP)$ contains the entries of all possible combinations of the original and the scaled eigenvectros, with gives the eigenvalues on the diagonal. The result can be understood more easily since after changing the basis to eigenvectors, the transformed eigenvectors are scaled by a factor of their eigenvalues and the result follows from lemma 3.5.1.

Example: Express a Normal Matrix in terms of its Eigenvectors.

Question: Show that a $n \times n$ normal matrix A can be written in terms of its eigenvalues λ_i and orthonormal eigenvectors $\mathbf{v}_i$ as

$$A = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i {\mathbf{v}_i}^\dagger. \tag{6.29}$$

Solution: we first note that an aribtrary vector can be written as $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{v}_i$, where $v_i = \mathbf{v}_i^\dagger \mathbf{v}$, since the eigenvectors $\mathbf{v}_i$ forms an orthonormal basis.
We can thus show that when both sides of the above equation is acted on $\mathbf{v}$, we get

[2] A more abstract proof is given in section A.16.

the same result

$$A\mathbf{v} = A\sum_{i=1}^{n} v_i \boldsymbol{\epsilon_i} = \sum_{i=1}^{n} v_i \lambda_i \mathbf{v}_i = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^\dagger \sum_{i=1}^{n} v_i \mathbf{v}_i. \tag{6.30}$$

For a general square matrix, there will be n eigenvalues from the fundamental theorem of algebra, though they may not be distinct. If they are distinct, then we can show that there are n linearly independent eigenvectors, and can be orthogonalized to orthonormal basis by the Gram-Schmidt procedure. If they are not distinct (*i.e.,* some of the eigenvalues are degenerate), then it may or may not have n linearly independent eigenvectors (since two eigenvectors corresponding to the same degenerate eigenvalue can form a plane or can be collinear).

A simple way to test whether the eigenvectors are linearly independent is by calculating the determinant. If the determinant is zero, the eigenvectors are linearly dependent, and vice versa.

Example: Diagonalizing with Degenerate Eigenvalues.

Question: Diagonlize the matrix $A = \frac{1}{4}\begin{pmatrix} 2 & 3\sqrt{2} & 3\sqrt{2} \\ 3\sqrt{2} & -1 & 3 \\ 3\sqrt{2} & 3 & -1 \end{pmatrix}$.

Solution: There are two eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -1$.
For $\lambda_1 = 2$, we get

$$\begin{pmatrix} -2 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & -3 & 1 \\ \sqrt{2} & 1 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0} \implies y = z = \frac{x}{\sqrt{2}} \implies \boldsymbol{\epsilon}_1 = \frac{1}{2}\begin{pmatrix} \sqrt{2} \\ 1 \\ 1 \end{pmatrix}. \tag{6.31}$$

For $\lambda_2 = -1$, we get

$$\begin{pmatrix} 2 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & 1 & 1 \\ \sqrt{2} & 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{0} \implies z = -\sqrt{2}x - y \implies \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -\sqrt{2} \end{pmatrix} \text{ and } \mathbf{v}_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}. \tag{6.32}$$

We have found two linearly independent eigenvectors for $\lambda_2 = -1$, which are both orthogonal to $\boldsymbol{\epsilon}_1$. However $\mathbf{v}_2$ and $\mathbf{v}_3$ themselves are not orthogonal.
To find $\boldsymbol{\epsilon}_2$ and $\boldsymbol{\epsilon}_3$ we apply the Gram-Schmidt procedure. First we normalize $\mathbf{v}_2$ to get

$$\boldsymbol{\epsilon}_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 0 \\ -\sqrt{2} \end{pmatrix}. \tag{6.33}$$

Then we subtract $\mathbf{v}_3$ from its projection onto $\boldsymbol{\epsilon}_2$ and normalize to get

$$\mathbf{v}_3' = \mathbf{v}_3 - (\epsilon_2 \cdot \mathbf{v}_3)\boldsymbol{\epsilon}_2 = \frac{1}{3}\begin{pmatrix} -\sqrt{2} \\ 3 \\ -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\epsilon_3} = \frac{1}{2\sqrt{3}}\begin{pmatrix} -\sqrt{2} \\ 3 \\ -1 \end{pmatrix}. \tag{6.34}$$

Therefore we have the transformation matrix

$$P = (\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\epsilon}_3) = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{2} & 0 & \frac{\sqrt{3}}{2} \\ \frac{1}{2} & -\frac{\sqrt{6}}{3} & -\frac{1}{2\sqrt{3}} \end{pmatrix}, \tag{6.35}$$

which indeed satsfies

$$P^{-1}AP = P^T AP = \mathrm{diag}(2, -1, -1). \tag{6.36}$$

Example: Trace Formula.

Question: Prove the trace formula $\det(\exp(A)) = \exp(\mathrm{Tr}(A))$.

Solution: Let $A' = P^{-1}AP = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ be the transfomred matrix of A. Then we have

$$\begin{aligned} \det(\exp(A')) &= \exp(\mathrm{diag}(\lambda_{1,2}, \ldots, \lambda_n)) = \det\big(\mathrm{diag}(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n})\big) \\ &= \prod_{i=1}^{n} e^{\lambda_i} = \exp\left(\sum_{i=1}^{n} \lambda_i\right) = \exp(\mathrm{Tr}(A)), \end{aligned} \tag{6.37}$$

which is equals to the determinant of A, since

$$\begin{aligned} &\exp(A') = 1 + A' + \frac{A'^2}{2} + \cdots = P^{-1}\exp(A)P \\ &\quad \Longrightarrow \det(\exp(A')) = \det\big(P^{-1}\big)\det(P)\det(\exp(A)) = \det(\exp(A)). \end{aligned} \tag{6.38}$$

Example: Symmetric Unitray Tranformation Matrix.

Question: A matrix A can be transformed from one orthonormal basis to another by the unitary transformation $A' = U^\dagger AU$. State the corresponding transformation for column vector $\mathbf{x}$. For the specific examples

$$A = \begin{pmatrix} 0 & 0 & i \\ 0 & 3 & 0 \\ -i & 0 & 0 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

construct a symmetric, unitary matrix U that can be used to transform A to a basis in which it is diagonal and find transformed matrix A' and vector $\mathbf{x}'$ in this basis.

Solution: A column vector transforms as

$$\mathbf{x}' = U^\dagger \mathbf{x}. \tag{6.39}$$

The eigenvalues of A are $1, 3, -1$ and their corresponding normalized eigenvectors are

$$\mathbf{u}_1 = \tfrac{1}{\sqrt{2}} \begin{pmatrix} i \\ 0 \\ 1 \end{pmatrix} e^{i\phi_1}, \quad \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} e^{i\phi_2}, \quad \mathbf{u}_3 = \tfrac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 0 \\ 1 \end{pmatrix} e^{i\phi_3}, \tag{6.40}$$

where the phase factors are added to add extra freedoms in order to satisfy the condition that the transformation matrix has to be symmetric.
Note that the eigenvectors are still normalized since $\left|e^{i\phi}\right| = 1$.
To construct a symmetric matrix we set $\phi_1 = -\pi/2, \phi_2 = \phi_3 = 0$, to get

$$U = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 1 & 0 \\ -\frac{i}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}. \tag{6.41}$$

The transfomred matrix and the transformed vector are therefore

$$A' = U^\dagger A U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \mathbf{x}' = U^\dagger \mathbf{x} = U^* \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1+i}{\sqrt{2}} \\ 1 \\ \frac{1+i}{\sqrt{2}} \end{pmatrix}. \tag{6.42}$$

6.4 Three-Dimensional Rotations

Since the eigenvalues λ_i of an unitary matrix U has modulus 1, *i.e.,* $|\lambda| = 1$, as we have proved in eq. (6.27), there always exists a transformation matrix P such that

$$P^{-1} U P = P^\dagger U P = \mathrm{diag}(e^{i\phi_1}, \ldots, e^{i\phi_n}). \tag{6.43}$$

For an orthogonal rotation matrix R for any eigenvalue λ its copmlex conjugate λ^* is also an eigenvalue by considering the charateristic polynomial $\chi_R(\lambda)$ where all the coefficients are real. In three dimensions it folows that the eigenvalues of R must be of the form $1, e^{i\phi}, e^{-i\phi}$, since from the charateristic polynomial $\chi_R(\lambda)$ we find the product of roots

$$\lambda_1\lambda_2\lambda_3 = (-1)^3(-\det R) \implies e^{i\phi}e^{-i\phi}\lambda_3 = \det R \implies \lambda_3 = 1. \tag{6.44}$$

The eigenvector corresponding to the eigenvalue 1 is the axis of rotation $\mathbf{n}$ which satsifies

$$R\mathbf{n} = \mathbf{n}. \tag{6.45}$$

Therefore for every three-dimensional rotation R we can find an orthonormal basis where the rotation matrix $\tilde{R}$ in this basis takes the form

$$\tilde{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \tag{6.46}$$

where the first column corresponds to the transformed vector of the axis of rotation, which is the axis of rotation itself, and the second and the third column correspond to the transformed vectors of two other eigenvectors, after rotating about the axis of rotation by θ.

From the basis-independence of the trace we also have that

$$\mathrm{Tr}(R) = \mathrm{Tr}\left(\tilde{R}\right) = 1 + 2\cos\theta. \tag{6.47}$$

Note, however, the direction of rotation is not defined by the matrix itself, which is why we can only obtain $\cos\theta$, but not θ itself from the matrix. This is becuase the axis of rotation eigenvector can have opposite directions, and if we apply the right hand rule we find that a vector would rotate in opposite directions. For three general vectors there are no rules like $\hat{\mathbf{x}} \times \hat{\mathbf{y}} = \hat{\mathbf{z}}$ to keep track of the direction of the axes.

6.5 Quadratic Forms

A quadractic function in terms of the components of a vector $\mathbf{x} = (x_1, x_2, x_3)$ can be written in matrix form as

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}, \tag{6.48}$$

where A is a symmetric matrix.

If A is not symmetric, then we have

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = \frac{1}{2}(\mathbf{x}^T A \mathbf{x} + \mathbf{x}^T A^T \mathbf{x}) = \mathbf{x}^T \left(\frac{A + A^T}{2}\right) \mathbf{x}, \tag{6.49}$$

where $(A + A^T)/2$ is a symmetric matrix. Note that we have used $\mathbf{x}^T A^T \mathbf{x} = \mathbf{x}^T A \mathbf{x}$, since one is a transpose of another, and both of them are scalar.

Q is basis-independent if the change of basis matrix P is orthogonal (*i.e.,* the column vectors in P are orthogonal), as

$$Q'(\mathbf{x}') = \mathbf{x}'^T A' \mathbf{x} = (P^{-1}\mathbf{x})^T P^{-1} A P (P^{-1}\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = Q(\mathbf{x}), \tag{6.50}$$

where we have used the orthogonality of P, *i.e.,* $P^T P = \mathbb{I}$ to rewrite $(P^{-1})^T = P$.

Since eigenvectors are generally orthogonal, so if we choose them to be the basis, we have

$$Q'(\mathbf{x}') = (\mathbf{x}')^T \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)\mathbf{x}' = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \cdots + \lambda_n x_n'^2. \tag{6.51}$$

Quadratic forms can be used to define quadratic curves in two dimensions or quadratic surfaces in three dimensions by the set of all points $\mathbf{x}$ satisfying

$$Q(\mathbf{x}) = Q'(\mathbf{x}') = \mathbf{x}^T A \mathbf{x} = \mathbf{x}'^T A' \mathbf{x} = c. \tag{6.52}$$

More explicitly,

$$A_{11}x^2 + A_{22}y^2 + A_{33}z^2 + \frac{A_{12}}{2}xy + \frac{A_{13}}{2}yz + \frac{A_{23}}{2}xz = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \lambda_3 x_3'^2. \tag{6.53}$$

By diagonalizing the quadratic form the nature of the quadratic curve or surface can be immediately read off from the eigenvalues λ_i of A as indicated in the table below.

Dimension	**Eigenvalue condition**	**Quadric type**
2D	all λ_i equal and same sign as c	circle
2D	all λ_i same sign as c	ellipse
2D	λ_i of mixed sign	hyperbola
3D	all λ_i equal and same sign as c	sphere
3D	all λ_i same sign as c	ellipsoid
3D	λ_i of mixed sign	hyperboloid

Table 6.1: Classification of central quadrics by eigenvalue signature.

If the quadratic form of A is positive for all column vectors $\mathbf{x}$, then it is positive definite. From the above equation, we see that this is equivalent to all the eigenvalues of A are positive.

For example the kinetic energy of a rotating rigid body is given by

$$E = \frac{1}{2}\sum_{i,j} I_{ij}\omega_i\omega_j = \frac{1}{2}\boldsymbol{\omega}^T I \boldsymbol{\omega}, \tag{6.54}$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ is the angular velocity and I is the moment of inertia tensor of the rigid body. By diagonlizing the moment of inertia tensor

$$PIP^T = \mathrm{diag}(I_1, I_2, I_3), \quad \boldsymbol{\Omega} = P\boldsymbol{\omega}, \tag{6.55}$$

we can write the kinetic energy as

$$E = \frac{1}{2}\sum_{i=1}^{3} I_i \Omega_i^2. \tag{6.56}$$

Example: Equation of an Ellipse.

Question: Write the equation of the ellipse $x^2 + 3y^2 - 2xy = 1$ in quadratic form and find the length of the two axis by diagonlizing A.

Solution: Let $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, then $\mathbf{x}^T A\mathbf{x} = ax^2 + 2bxy + cy^2$. Comparing the coefficients with the expression on the LHS of the ellipse equation, we have $a = 1, b = -1$ and $c = 3$. Diagonlizing A gives the matrix $B = \begin{pmatrix} 2+\sqrt{2} & 0 \\ 0 & 2+\sqrt{2} \end{pmatrix}$, and we have

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = x^2+3y^2-2xy = Q'(\mathbf{x}') = \mathbf{x}'^T B\mathbf{x}' = (2+\sqrt{2})x'^2+(2-\sqrt{2})y'^2 = 1. \tag{6.57}$$

Thus the length of the two axes are $a = \frac{1}{\sqrt{2+\sqrt{2}}}$ and $b = \frac{1}{\sqrt{2-\sqrt{2}}}$.

Example: Quadratic Form (1).

Question: The quadratic forms $Q_1(x, y, z)$ and $Q_2(x, y, z)$ are defined by

$$Q_1(x, y, z) \equiv x^2-4y^2+z^2+4yz-4zx-4xy \text{ and } Q_2(x, y, z) \equiv x^2+y^2+z^2. \tag{6.58}$$

Find the corresponding maxtrix A_1 and A_2 representing the quadratic forms Q_1 and Q_2. Find a real orthogonal transformation matrix P such that A and A' can be diagonalized simultaneously.
Sketch the surface Σ defined by the equation $Q_1(x, y, z) = -5$ and describe the shape of the intersection of Σ with the plane through the origin perpendicular to the direction $(2, 1, -2)$.

Solution: The matrix A and A' can be stated as

$$A_1 = \begin{pmatrix} 1 & -2 & -2 \\ -2 & -4 & 2 \\ -2 & 2 & 1 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbb{I}. \tag{6.59}$$

Since $A_1A_2 = A_1\mathbb{I} = IA_1 = A_2A_1$, A_1 and A_2 commute and the two matrices can be diagonalized simultaneously.
The eigenvalues for A_1 are $-1, 4, -5$. Arranging the normalized eigenvectors into a matrix we obtain the transformation matrix P

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{2}{3} \end{pmatrix}. \tag{6.60}$$

Note that for P to be a rotation matrix so as to diagonalize $A_2 = \mathbb{I}$ the column vectors must be diagonalized.

$Q_1(x, y, z) = -X^2 - 5Y^2 + 4Z^2 = -5$ is a one-sheeted hyperboloid, which can be determined by the fact that there is no Z-intercept but X- and Y-intercepts. The point $(\mathbf{r} = 2, 1, -2)$ is transfomred by P by

$$\mathbf{r}' = P^T\mathbf{r} = \left(9, -\frac{8}{3\sqrt{2}}, -\frac{7}{3}\right), \tag{6.61}$$

so the equation of the plane given in X, Y, Z basis is

$$-\frac{8}{3\sqrt{2}}Y + \frac{7Z}{3} = 0. \tag{6.62}$$

Together with $X^2 + 5Y^2 - 4Z^2 = 5$ we have

$$X^2 + \frac{117}{49}Y^2 = 5, \tag{6.63}$$

which is an ellipse from the top view. Figure 6.1 shows the intersection between the plane and the hyperboloid.

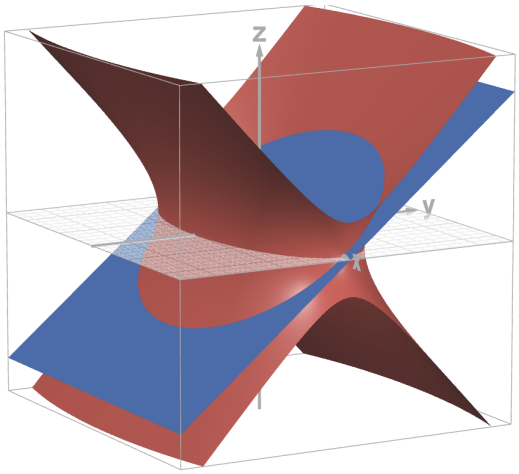


Figure 6.1

Example: Pauli Matrices.

Question: Explain the Properties of Pauli Matrices.

Solution:

1. The Pauli matrices and their algebra

The three 2×2 *Pauli matrices* are

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{6.64}$$

They span the real vector space

$$L = \mathrm{Span}_{\mathbb{R}}\{\sigma_1, \sigma_2, \sigma_3\}, \tag{6.65}$$

namely all 2×2 Hermitian, traceless matrices. Their multiplication rule is

$$\sigma_i \, \sigma_j \;=\; \delta_{ij} \, I \;+\; i \, \varepsilon_{ijk} \, \sigma_k, \tag{6.66}$$

where δ_{ij} is the Kronecker delta, ε_{ijk} is the Levi–Civita symbol, and repeated indices are summed over $1, 2, 3$. From this one obtains

$$\{\sigma_i, \sigma_j\} = \sigma_i \sigma_j + \sigma_j \sigma_i = 2 \, \delta_{ij} \, I, \tag{6.67}$$

$$[\sigma_i, \sigma_j] = \sigma_i \sigma_j - \sigma_j \sigma_i = 2 \, i \, \varepsilon_{ijk} \, \sigma_k. \tag{6.68}$$

2. Squaring a general "Pauli-vector"

For any real vector $\boldsymbol{a} = (a_1, a_2, a_3)$ define

$$\boldsymbol{a} \cdot \boldsymbol{\sigma} = a_i \, \sigma_i. \tag{6.69}$$

Then

$$(\boldsymbol{a} \cdot \boldsymbol{\sigma})^2 = a_i a_j \, \sigma_i \sigma_j = \tfrac{1}{2} \, (a_i a_i) \, I = \tfrac{1}{2} \, |\boldsymbol{a}|^2 \, I, \tag{6.70}$$

and by induction

$$(\boldsymbol{a} \cdot \boldsymbol{\sigma})^{2n} = \frac{|\boldsymbol{a}|^{2n}}{2^n} \, I, \tag{6.71}$$

$$(\boldsymbol{a} \cdot \boldsymbol{\sigma})^{2n+1} = \frac{|\boldsymbol{a}|^{2n}}{2^n} \, (\boldsymbol{a} \cdot \boldsymbol{\sigma}). \tag{6.72}$$

3. Exponentiation of $i\theta \, \boldsymbol{n} \cdot \boldsymbol{\sigma}$

Let $\boldsymbol{n}$ be a unit vector and $\theta \in \mathbb{R}$. Then

$$U = \exp\bigl(i\theta \, \boldsymbol{n} \cdot \boldsymbol{\sigma}\bigr) = \sum_{m=0}^{\infty} \frac{(i\theta)^m}{m!} \, (\boldsymbol{n} \cdot \boldsymbol{\sigma})^m. \tag{6.73}$$

Splitting into even and odd terms and using the formulas above gives

$$\exp\bigl(i\theta \, \boldsymbol{n} \cdot \boldsymbol{\sigma}\bigr) = \cos\theta \, I + i \, \sin\theta \, (\boldsymbol{n} \cdot \boldsymbol{\sigma}). \tag{6.74}$$

4. Unitarity and $\mathrm{SU}(2)$

Since each σ_i is Hermitian,

$$\bigl(i\theta \, \boldsymbol{n} \cdot \boldsymbol{\sigma}\bigr)^{\dagger} = -\, i\theta \, \boldsymbol{n} \cdot \boldsymbol{\sigma}, \tag{6.75}$$

one checks

$$U^{\dagger} U = \bigl(\cos\theta \, I - i \, \sin\theta \, \boldsymbol{n} \cdot \boldsymbol{\sigma}\bigr)\bigl(\cos\theta \, I + i \, \sin\theta \, \boldsymbol{n} \cdot \boldsymbol{\sigma}\bigr) = I, \tag{6.76}$$

so $U \in U(2)$. Moreover,

$$\det U = \cos^2\theta + \sin^2\theta = 1, \tag{6.77}$$

hence $U \in SU(2)$.

5. Parametrization of all $SU(2)$

Every $U \in SU(2)$ can be written as

$$U = \exp\bigl(i\theta\, \boldsymbol{n}\cdot\boldsymbol{\sigma}\bigr), \tag{6.78}$$

which explicitly yields

$$U = \begin{pmatrix} \cos\theta + i\, n_3\, \sin\theta & (n_2 + i\, n_1)\, \sin\theta \\ -(n_2 - i\, n_1)\, \sin\theta & \cos\theta - i\, n_3\, \sin\theta \end{pmatrix}, \tag{6.79}$$

with $\det U = 1$. Equivalently, setting

$$\alpha = \cos\theta + i\, n_3 \sin\theta, \quad \beta = (n_2 + i\, n_1)\, \sin\theta, \tag{6.80}$$

one has $|\alpha|^2 + |\beta|^2 = 1$.

7 Hibert Space

7.1 Vector Spaces for Functions

7.1.1 Classes of Functions

The set of all functions with the set S as their domain and V as their co-domain is denoted by

$$\mathcal{F}(S, V) \equiv f : S \to V. \tag{7.1}$$

For example, if $S = [a, b] \subset \mathcal{R}$ and $V = \mathbb{R}$, then we are considering all the real-valued functions whose domain is $[a, b]$. Sometimes $\mathbb{R}$ is placed in the subscript, so $\mathcal{C}_{\mathbb{R}}([a, b])$ is the space with all real-valued continuous functions of $[a, b]$.

Some commonly used sub vector spaces include $\mathcal{C}([a, b])$, which is the set of all continuous functions on the interval $[a, b]$. $\mathcal{C}^k([a, b])$ is the sset of all k times continuously differentiable functions on this interval, where $k = \infty$ is allowed. On the other hand, $\mathcal{C}_c(\mathbb{R})$ contains all the continuous functions with compact support, *i.e.,* there exist a certain radius $R > 0$ such that $f(x) = 0$ for all $|x| > R$. The vector space of all polynomials is denoted by $\mathcal{P}([a, b])$.

7.1.2 Linear Maps

Integral Operator

The map $T : \mathcal{C}([a, b]) \to C([a, b])$ is an integral operator if

$$T(f)(x) \equiv \int_a^b K(x, \tilde{x}) f(\tilde{x}) d\tilde{x}, \tag{7.2}$$

where $K : [a, b] \times [a, b] \to \mathbb{R}$ is a real-valued continuous function of two variables, called the kernel of this integral operator, which encodes important information about the operator.

Differential Operator

The map $D : \mathcal{C}^\infty([a,b]) \to \mathcal{C}^\infty([a,b])$ is an differential operator if

$$D(f)(x) \equiv \frac{df}{dx}(x). \tag{7.3}$$

Multiplication Operator

THe map $M_p : C^\infty([a,b]) \to C^\infty[a,b]$ is a multiplication operator if

$$M_p(f)(x) \equiv p(x)f(x), \tag{7.4}$$

where $p \in \mathcal{C}^\infty([a,b])$ is a fixed function.

Linear Differential Operator

We can combine the differential operator and the multiplication operator to construct a linear differential operator $T : \mathcal{C}^\infty([a,b]) \to \mathcal{C}^\infty([a,b])$, defined by

$$T \equiv p_k \frac{d^k}{dx^k} + p_{k-1}\frac{d^{k-1}}{dx^{k-1}} + \cdots + p_1 \frac{d}{dx} + p_0, \tag{7.5}$$

where p_i, for $i = 0, 1, \ldots, k$ are fixed functions in $\mathcal{C}^\infty([a,b])$.

Definition 7.1.1 (Bounded Linear Operator)**.** *A linear operator $T : V \to W$ is called bounded if there exists a positive $K \in \mathbb{R}$ such that $\|T(\mathbf{v})\|_W \leq K\|\mathbf{v}\|_V$ for all $\mathbf{v} \in V$. The smallest number K for which this condition is satisfied is called the norm, $\|T\|$, of the operator T.*

Definition 7.1.2 (Isometries)**.** *A bounded linear operator $T : V \to W$ is an isometry if and only if $\|T(\mathbf{v})\|_W = \|\mathbf{v}\|_\mathbf{V}$ for all $\mathbf{v} \in V$.*

Appendices

A

Rigorous Definitions and Proofs

A.1 Definitions of Set, Group, Field and Vector Spaces

Here we will formally define set, group, field and vector space.

1. A set is a collection of mathematical objects.
2. A set G with a binary operation that combines two elements in the set to yield another element in the set is called a group. Mathematically this can be written as $G \times G \to G$. This operation has to have certain properties (also called the axioms), and together with the set these constitute the algebraic structure of the group. Specifically, for a group these are closure, associativity, the existence of an identity element, and the existence of inverse elements. If the operation also commutes (*i.e.,* it does not depend on the order in which the elements are written in the operation), then the group is said to be a commutative or an Abelian group.
3. A set with two binary operations of addition and multiplication is called a field F. These must satisfy six field axioms, namely associativity, commutativity, the existence of additive and multiplicative identity, and the existence of additive and multiplicative inverses, as well as distributivity. We see the complexity of the structure is increasing; in particular, we note that a field is an Abelian group under both addition and multiplication, the two being connected via the requirement of distributivity. Note that the set of rationals $\mathbb{Q}$, real numbers $\mathbb{R}$, and complex numbers $\mathbb{C}$ are all fields, but the set of integers $\mathbb{Z}$ is not a field as the multiplicative inverse of every integer is not necessarily another integer (e.g., the multiplicative inverse of 2 is 1/2, not an integer). The integers are, instead, a ring. Rings generalize fields: they too have two operations, but multiplication does not have to be commutative, and multiplicative inverses need not exist.
4. A vector space over a field F is a set V together with two binary operations: vector addition ($V \times V \to V$) and scalar multiplication ($V \times F \to V$). Note the difference with the definition of a field: the notion of multiplication between vectors is not part of the fundamental algebraic structure of a vector space. In fact, as we know there are multiple ways to multiply vectors, and the multiplicative inverse (*i.e.,* vector division) is not defined. The elements of the field are called scalars to distinguish them from the elements of the vector space (vectors), and because the

definition of multiplication acts to scale the vectors. We will typically be interested in vector spaces defined over $\mathbb{R}$: this is called a real vector space. Eight axioms must be satisfied for the operations permitted on a vector space: associativity and commutativity of addition, the existence of identity elements for addition and scalar multiplication, the existence of additive inverses, compatibility of scalar multiplication with respect to field addition, and compatibility of scalar and field multiplication.

A.2 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.2.1. $0 \cdot \mathbf{v} = \mathbf{0}$ *for all* $\mathbf{v} \in V$.

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

Lemma A.2.2. $(-1)\mathbf{v} = -\mathbf{v}$ *for all* $\mathbf{v} \in V$.

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

Since $0\mathbf{w} = 0$ 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.3 Relation between Inverse and Bijective Maps

Here we provide a rigorous proof for theorem 3.1.1.

Proof. We assume that $f : X \to Y$ has an inverse $g : Y \to X$ with $f \circ g = \mathrm{id}_Y$ and $g \circ f = \mathrm{id}_X$. To show that fis 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 sujectivity 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 $f(x) = f \circ g(y) = \mathrm{id}_Y(y) = y$. In conclusion f is bijective.

To show uniqueness we consider two maps $g : Y \to X$ and $\tilde{g} : Y \to X$ 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 $g = \tilde{g}$. □

A.4 Inverse of a Composite Map

Here we provide a rigorous proof for theorem 3.1.2.

Proof. This relation follows from $(f\circ g)^{-1}\circ(f\circ g) = \mathrm{id}$ by definition and $(g^{-1}\circ f^{-1})\circ(f\circ g) = g^{-1}\circ g = \mathrm{id}$ 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) . □

A.5 Properties of a Linear Map (1)

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

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

$(F2)$ 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 $f(\mathbf{v}_1 + \mathbf{v}_2) \overset{E1}{=} f(\mathbf{v}_1) + f(\mathbf{v}_2) = \mathbf{0}$ 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 $f(\alpha\mathbf{v}) \overset{(E2)}{=} \alpha f(\mathbf{v}) = \mathbf{0}$, hence $\alpha\mathbf{v} \in \ker(f)$.

$(F3)$ If $\mathbf{w}_1, \mathbf{w}_2 \in \mathrm{Im}(f)$, 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 $\mathbf{w}_1 + \mathbf{w}_2 = f(\mathbf{v}_1) + f(\mathbf{v}_2) \overset{(E1)}{=} f(\mathbf{v}_1 + \mathbf{v}_2) \in \mathrm{Im}(f)$, hence condition (B1) is satisfied. For (B2), if $f(\mathbf{v}) = \mathbf{w}$ where $\mathbf{v} \in V$ and $\mathbf{w} \in \mathrm{Im}(f)$, then $f(\alpha\mathbf{v}) \overset{(E1)}{=} \alpha f(\mathbf{v})$

$(F4)$ 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 $f(0) = \mathbf{0}$ 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 $(g \circ f)(\alpha\mathbf{v}) = g(f(\alpha\mathbf{v})) = g(\alpha f(\mathbf{v})) = \alpha g(f(\mathbf{v})) = \alpha(g \circ f)(\mathbf{v})$.

□

A.6 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 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be the basis of $\ker(f)$ which we complete to a basis $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, v_n$ of V. We will show that $f(\mathbf{v}_{k+1}), \ldots, f(\mathbf{v}_n)$ forms a basis of $\mathrm{Im}(f)$. To satisfy condition (C2), we need to show that $\mathrm{Im}(f)$ is spanned by $f(\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$. We begin with an arbitary vector $\mathbf{w} \in \mathrm{Im}(f)$. 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

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i \tag{A.1}$$

Acting on this equation with f and using linearity we find

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

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

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

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

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

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

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

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

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

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

□

A.7 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 $\dim \mathrm{Im}(f) = \dim(W)$. Then from eq. (3.12), $\dim(V) = \dim\ker(f) + \dim \mathrm{Im}(f) = 0 + \dim(W) = \dim(W)$.

($G2$) 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.8 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

$$\mathbf{A}_1 = \sum_{j=2}^{m} \alpha_j \mathbf{A}_j \tag{A.6}$$

then if we define

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

then the column vectors of A can be written as

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

Lemma A.8.1. *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

$$\sum_{i=1}^{n} \beta_i \mathbf{b}_i = 0 \iff \beta_i = 0 \text{ for all } i \tag{A.9}$$

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 $\alpha_2, \alpha_3, \ldots, \alpha_m$ are arbitrary, thus the coeffiecients for α_i must be zero for all i. This implies that $\sum_{i=1}^{n}\gamma_i b_{ji}=0$ for all i. But we have already proved that this condition implies $\gamma_i = 0$ for all 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 linearly dependency of the column vectors. We can use the same argument if the column vectors are linearly dependent instead. □

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 column.[1] 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 quadratic[2] Therefore the row rank must be equals to the column rank. □

A.9 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.

□

[1] Since 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.

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

A.10 Rank of a Matrix in Upper Echelon Form

Here we provide a proof for lemma 3.7.2.

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

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

then by inspection $\alpha_k = 0$ for all $k < i$ 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 continue 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.11 Matrix Rank is Equals to Linear Map Rank

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

Proof. Since lemma 3.5.1 implies that $\text{Im}(A) = \psi^{-1}(\text{Im}(f))$, if we denote by $\chi \equiv \psi^{-1}\big|_{\text{Im}(f)}$ 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

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

□

A.12 Linear Map is Uniquely Determined by Matrix Elements

Here we provide a proof for lemma 1.4.1.

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 □

A.13 Properties of Perpendicular Spaces

Here we provide a proof for lemma 5.5.1.

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 orthonormal basis $\{\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_k\}$ of W and define the linear map $f : V \to V$ by $f(\mathbf{v}) = \sum_{i=1}^{k} \langle \boldsymbol{\epsilon}_i, \mathbf{v} \rangle \boldsymbol{\epsilon}_i$ (a projection onto W). Clearly $\mathrm{Im}(f) \subset W$. For $\mathbf{w} \in W$ it follows that $f(\mathbf{w}) = \mathbf{w}$ so that $\mathrm{Im}(f) = W$. Moreover, $\mathrm{Ker}(f) = W^\perp$ and the claim follows from eq. (3.12) applied to the map f.

□

A.14 Properties of Adjoint Linear Maps

Here we provide proofs for lemma 5.6.1.

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 1.4.1 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 5.4.1 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 1.4.1.

$(M4)$ $\langle (\alpha f)^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, (\alpha f)(\mathbf{w}) \rangle = \alpha \langle \mathbf{v}, f(\mathbf{w}) \rangle = \alpha \langle f^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle (\alpha^* f^\dagger)(\mathbf{v}), \mathbf{w} \rangle$ and lemma 1.4.1 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 $\mathrm{id}_V = (f \circ f^{-1})^\dagger = f^\dagger \circ (f^{-1})^\dagger$. This means $(f^{-1})^\dagger$ is the inverse of $f^\dagger$ and, hence, $(f^\dagger)^{-1} = (f^{-1})^\dagger$.

□

A.15 Eigenvectors and Eigenvalues of Hermitian Matrices

Here we provide a more abstract proof just the hermitian case of theorem 6.3.1.

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

$$\lambda\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, \lambda\mathbf{v} \rangle = \langle \mathbf{v}, f(\mathbf{v}) \rangle = \langle f(\mathbf{v}), \mathbf{v} \rangle = \langle \lambda\mathbf{v}, \mathbf{v} \rangle = \lambda^* \langle \mathbf{v}, \mathbf{v} \rangle. \tag{A.13}$$

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 $\mathbf{v} \neq 0$, $\langle \mathbf{v}, \mathbf{v} \rangle \neq 0$, it follows that $\lambda = \lambda^*$, so the eigenvalue is real.

($P2$) Consider two eigenvectors $\mathbf{v}_1, \mathbf{v}_2$, so that $f(\mathbf{v}_1) = \lambda_1 \mathbf{v}_1 \quad \text{and} \quad f(\mathbf{v}_2) = \lambda_2 \mathbf{v}_2$, with different eigenvalues, $\lambda_1 \neq \lambda_2$. Then

$$\begin{aligned}(\lambda_1 - \lambda_2)\langle \mathbf{v}_1, \mathbf{v}_2 \rangle &= \langle \lambda_1 \mathbf{v}_1, \mathbf{v}_2 \rangle - \langle \mathbf{v}_1, \lambda_2 \mathbf{v}_2 \rangle \\ &= \langle f(\mathbf{v}_1), \mathbf{v}_2 \rangle - \langle \mathbf{v}_1, f(\mathbf{v}_2) \rangle \\ &= \langle \mathbf{v}_1, f(\mathbf{v}_2) \rangle - \langle \mathbf{v}_1, f(\mathbf{v}_2) \rangle = 0.\end{aligned} \tag{A.14}$$

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

□

A.16 Self Adjoint Vector Space has Orthonomral Eigenvectors

Here we provide a proof for theorem 6.3.2.

Proof. The proof is by induction in n, the dimension of the vector space V.

The characteristic polynomial χ_f of f has at least one zero (not more since it can be repeated root), λ, over the complex numbers. Since f is self-adjoint, λ is real from the previous theorem.

Consider the eigenspace $W = \mathrm{Eig}_f(\lambda)$. Vectors $\mathbf{v} \in W^\perp$ and $\mathbf{w} \in W$ are perpendicular, so

$$\langle \mathbf{w}, f(\mathbf{v}) \rangle = \langle f(\mathbf{w}), \mathbf{v} \rangle = \langle \lambda \mathbf{w}, \mathbf{v} \rangle = \lambda \langle \mathbf{w}, \mathbf{v} \rangle = 0 \implies f(\mathbf{v}) \in W^\perp. \tag{A.15}$$

Since $\dim(W^\perp) < n$, there is an orthonormal basis $\epsilon_1, \ldots, \epsilon_k$ of $W^\perp$ consisting of eigenvectors of f by the induction assumption.

Add to this orthonormal basis of $W^\perp$ an orthonormal 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 orthonormal basis of V, consisting of eigenvectors of f.

□

B

Extra Materials

B.1 Fields

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

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

where the modulus operation is introduced such that the result of addition and multiplication is brought back to the required range $\{0, 1, \ldots, p-1\}$ 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 $\mathbb{F}_2$

$\cdot$	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.