Eigenphysics I

Contents

List of Figures

Chapter 1

Vector Spaces

1.1 Introduction

Examples:

(1) 3-D space that we know and ‘love’

(2) $4-D$ space-time in relativity

(3)Fourier series that we learn and ‘aint so bad’

(4) Wavefunctions in quantum mechanics

We tend to think in terms of ‘finite dimensions’, but we often work with ‘infinite dimensional’ spaces. Most things are analogous.

Formal Definition:

A vector space, V say, is a set of objects, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}\ldots$ or $|x>$, $|y>$, $|z> \ldots$, satisfying some axioms:

1.Addition, "+", must be defined, and if $\mathbf{x}, \mathbf{y} \in V$, then $\mathbf{x}+\mathbf{y} \in V$ [must have all space, cannot restrict to a sphere only].

2. Commutivity: $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$.

3. Associativity: $(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{y}+(\mathbf{x}+\mathbf{z})$ [do not need brackets, and can reorder at will].

4. Set V must contain a ‘null vector’, ‘0’, with the property that $\mathbf{x}+\mathbf{0} = \mathbf{0}+\mathbf{x} = \mathbf{x}$.

5. Set V must contain an ‘inverse’, $(-\mathbf{x})$, with the property, $\mathbf{x}+(-\mathbf{x}) = \mathbf{0}$.

6. Multiplication by a scalar, λ (real or complex), must be defined, and $\lambda\mathbf{x} \in V$ [scale factor].

7. $0\mathbf{x} = \mathbf{0}$.

NB: It is useful to check $\lambda\mathbf{x} + \mu\mathbf{y} \in V$, for arbitrary λ and μ, which tests most of the requirements in one go.

Example: Addittion applied to all real numbers.

Example: Multiplication applied to positive definitereal numbers, but with power-law scale-factors.

$$\begin{aligned} x + y &\mapsto xy \\ 0 &\mapsto 1 \\ (-x) &\mapsto 1/x \\ \lambda x &\mapsto x^\lambda \end{aligned} \tag{1.1}$$

(1) $xy \in V$

(2) $xy = yx$

(3) $(xy)z = x(yz)$

(4) $x1 = 1x = x$

(5) $(1/x)x = x(1/x) = 1$

(6) $x^\lambda \in V$

(7) $x^0 = 1$.

Chapter 2

Linear Dependence and Independence

Linear dependence is the most useful concept for thinking about vector spaces.

n vectors are said to be linearly dependent if there exist n non-vanishing scalars, $\lambda_1, \lambda_2, \ldots, \lambda_n$, such that:

$$\lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n} = \mathbf{0}$$

n vectors are said to be linearly independent otherwise. Linear independence is powerful, since for n linearly independent vectors:

$$\lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n} = \mathbf{0}$$

implies $\lambda_1 = 0,\ \lambda_2 = 0, \ldots\ \lambda_n = 0$

one vector equation contains n independent scalar equations usually.

The maximum number of linearly independent vectors is known as the dimension of the vector space, and a particular set of such linearly independent vectors is known as a basis for the vector space.

All vectors in a vector space can be represented as a linear sum of basis vectors:

A basis is said to *span* the vector space.

Proof:

Given $\mathbf{v}$ and a basis $\mathbf{x_1}$, $\mathbf{x_2}, \ldots, \mathbf{x_n}$, we have $n+1$ vectors must be linearly dependent, so:

$$\lambda_v \mathbf{v} + \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n} = 0$$

for non vanishing scalars. If $\lambda_v = 0$, then $\lambda_i = 0$, because the $\mathbf{x_i}$ are linearly independent, so $\lambda_v \neq 0$. Therefore

$$\mathbf{v} = -(\lambda_1/\lambda_v)\mathbf{x_1} - (\lambda_2/\lambda_v)\mathbf{x_2} \cdots - (\lambda_n/\lambda_v)\mathbf{x_n}$$

The coefficients which represent a chosen vector in a given basis *uniquely* define that vector.

Proof:

If:

$$\mathbf{v} = \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} \cdots + \lambda_n \mathbf{x_n} = \lambda'_1 \mathbf{x_1} + \lambda'_2 \mathbf{x_2} \cdots + \lambda'_n \mathbf{x_n}$$

then

$$(\lambda_1 - \lambda'_1)\mathbf{x_1} + (\lambda_2 - \lambda'_2)\mathbf{x_2} \cdots + (\lambda_! - \lambda'_n)\mathbf{x_n} = 0$$

and so $\lambda'_i = \lambda_i$ from the linear independence of the basis vectors $\mathbf{x_i}$.

The scalars which represent a vector in a given basis $\{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$ are known as the *components* of the vector in the basis:

$$\mathbf{v} = \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n}.$$

They are usually denoted by $(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

In particular we always work with components. Vectors are physical objects which are represented by symbols $\mathbf{x_i}$. The components are 'numbers' that we understand how to deal with. Even when we try to represent $\mathbf{x_i}$, we often use another basis to do it, eg.

$$\mathbf{x_1} = (1,1,1) \quad \mathbf{x_2} = (0,1,1) \quad \mathbf{x_3} = (0,0,1).$$

There is an implied basis in terms of which the $\mathbf{x_i}$ are represented as components.

This leads to the first very natural style of problem: Find scalars, $\lambda_1, \lambda_2, \ldots, \lambda_n$, which satisfy:

$$\mathbf{v} = \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n}$$

for given vectors, $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ and $\mathbf{v}$.

This is a set of *simultaneous linear equations*, and you should know how to solve this type of problem.

Example: $\mathbf{x_1} = (1,1,1)$ $\mathbf{x_2} = (0,1,1)$ $\mathbf{x_3} = (0,0,1)$, find λ_1, λ_2, λ_3, such that

$$\mathbf{v} = \lambda_1\mathbf{x_1} + \lambda_2\mathbf{x_2} + \lambda_3\mathbf{x_3}$$

$$\lambda_1\begin{pmatrix}1\\1\\1\end{pmatrix} + \lambda_2\begin{pmatrix}0\\1\\1\end{pmatrix} + \lambda_3\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}3\\2\\1\end{pmatrix}$$

so

$$\begin{aligned}\lambda_1 &= 3 \\ \lambda_1 + \lambda_2 &= 2 \quad \textit{implies} \quad \lambda_2 = -1 \\ \lambda_1 + \lambda_2 + \lambda_3 &= 1 \quad \textit{implies} \quad \lambda_3 = -1\end{aligned}$$

Example: Find a basis for the vector space spanned by:

$$V = \{(1,1,1,2),(2,-1,2,-2),(3,2,3,4),(2,0,2,0)\}.$$

What is the dimension of this vector space? Assess whether $\mathbf{v} = (1,-1,1,-2)$ is in the space, and if it is find its components in your chosen basis.

One can take sums and linear combinations of any of the vectors to try to make the problem simpler. The last vector looks the simplest so set:

$$\mathbf{x_1} = (1,0,1,0,2) = \frac{1}{2}(2,0,2,0) = \frac{1}{2}\mathbf{v_4}$$

subtracting this vector from the first vector gives:

$$\mathbf{x_2} = (0,1,0,2) = \mathbf{v_1} - \mathbf{x_1}$$

and in fact all the vectors can be generated from this basis:

$$\begin{aligned} \mathbf{v_1} &= \mathbf{x_1} + \mathbf{x_2} \\ \mathbf{v_2} &= 2\mathbf{x_1} - \mathbf{x_2} \\ \mathbf{v_3} &= 3\mathbf{x_1} + 2\mathbf{x_2} \\ \mathbf{v_4} &= 2\mathbf{x_1} \end{aligned}$$

The dimension is two and the vector $\mathbf{v}$ satisfies:

$$\mathbf{v} = \mathbf{x_1} - \mathbf{x_2}$$

and consequently the components chosen basis are $\lambda_1 = 1$, $\lambda_2 = -1$.

Solving simultaneous linear equationsn is equivalent to being able to solve two related problems:

(1) Are $\mathbf{x_1}$, $\mathbf{x_2}$, $\ldots, \mathbf{x_n}$, linearly independent?
(2) If $\mathbf{x_1}$, $\mathbf{x_2}$, $\ldots, \mathbf{x_n}$, linearly independent, find scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$, which satisfy:

$$\lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n} = \mathbf{0}$$

'Proof': Find scalars, $\lambda_1, \lambda_2, \ldots, \lambda_n$, which satisfy

$$\mathbf{v} = \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n}$$

(1) If $\mathbf{x_1}$, $\mathbf{x_2}$, $\ldots, \mathbf{x_n}$, linearly dependent, then non-vanishing scalars, $\mu_1, \mu_2, \ldots, \mu_n$ exist such that:

$$\mu_1 \mathbf{x_1} + \mu_2 \mathbf{x_2} + \cdots + \mu_n \mathbf{x_n} = \mathbf{0}$$

we can ignore one of the $\mathbf{x_i}$ with $\mu_i \neq 0$ in the original problem and then if we find a solution, λ_i^* say, then $\lambda_i^* + \kappa\mu_i$ is a solution for *arbitrary* κ. Obviously we can repeat this argument until $\mathbf{x_1}$, $\mathbf{x_2}$, $\ldots, \mathbf{x_n}$, can be assumed linearly independent.

(2) If $\mathbf{x_1}$, $\mathbf{x_2}$, $\ldots, \mathbf{x_n}$ and $\mathbf{v}$ linearly dependent, then there are no solutions. If they are linearly dependent, then

$$\lambda_v \mathbf{v} + \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n} = \mathbf{0}$$

with $\lambda_v \neq 0$ (otherwise all are zero), and the problem is solved.

Example:

Solve

$$\begin{aligned} \lambda_1 + 2\lambda_2 + 3\lambda_3 + 2\lambda_4 &= 1 \\ \lambda_1 - \lambda_2 + 2\lambda_3 &= -1 \\ \lambda_1 + 2\lambda_2 + 3\lambda_3 + 2\lambda_4 &= 1 \\ 2\lambda_1 + 2\lambda_2 + 4\lambda_3 &= -2 \end{aligned}$$

This problem is equivalent to finding the components of a particular vector in a particular basis:

$$\mathbf{v} = (1, -1, 1, -2) = \lambda_1(1, 1, 1, 2) + \lambda_2(2, -1, 2, -2) + \lambda_3(3, 2, 3, 4) + \lambda_4(2, 0, 2, 0).$$

We list the vectors

$$\begin{aligned} \mathbf{x_1} &= (1, -1, 1, -2) \\ \mathbf{x_2} &= (2, -1, 2, -2) \\ \mathbf{x_3} &= (3, 2, 3, 4) \\ \mathbf{x_4} &= (2, 0, 2, 0). \end{aligned}$$

The four vectors are linearly dependent and a particular dependence is:

$$2(1, 1, 1, 2) + 2(2, -1, 2, -2) - 3(2, 0, 2, 0) = \mathbf{0}$$

so we can ignore $\mathbf{x_2}$ and allow $\lambda_2 = 0$. The remaining three vectors are *still* linearly depndent although now the dependence is unique:

$$4(1, 1, 1, 2) - 2(3, 2, 3, 4) + 1(2, 0, 2, 0) = \mathbf{0}.$$

We can ignore $\mathbf{x_3}$ and allow $\lambda_3 = 0$. We are left with:

$$(1, -1, 1, -2) = \lambda_1(1, 1, 1, 2) + \lambda_4(2, 0, 2, 0)$$

and once more there is a linear dependence:

$$(1,-1,1,-2)+(1,1,1,2)-(2,0,2,0)=0$$

and so there is a solution:

$$\begin{aligned}(1,-1,1,-2) = -(1,1,1,2)+(2,0,2,0)+\kappa_1[2(1,1,1,2)+2(2,-1,2,-1)-3(2,0,2,0)] \\ +\kappa_2[4(1,1,1,2)-2(3,2,3,4)+(2,0,2,0)]\end{aligned} \tag{2.1}$$

or

$$(1,-1,1,-2)=(-1+2\kappa_1+4\kappa_2)(1,1,1,2)+2\kappa_1(2,-1,2,-1)-2\kappa_2(3,2,3,4)+(1-3\kappa_1+\kappa_2)(2,0,2,0)$$

and the general solution is:

$$\begin{aligned}\lambda_1 &= -1+2\kappa_1+4\kappa_2 \\ \lambda_2 &= 2\kappa_1 \\ \lambda_3 &= -2\kappa_2 \\ \lambda_4 &= 1-3\kappa_1+\kappa_2\end{aligned}$$

where the κ_i can take any values.

□

The best method of solving this type of problem is *controlled gaussian elimination*.

The other technique of use is that of determinant evaluation, which is useful for 'square' problems. If you evaluate the determinant made from the chosen vectors as columns, then the linear dependence is equaivalent to the vanishing of the determinant.

Example:

Three vectors take the form:

$$\mathbf{x_1}=\begin{pmatrix}1\\1\\1\end{pmatrix} \qquad \mathbf{x_2}=\begin{pmatrix}1\\-1\\0\end{pmatrix} \qquad \mathbf{x_3}=\begin{pmatrix}1\\\alpha\\2\end{pmatrix}.$$

For what values of α are $\{\mathbf{x_1},\mathbf{x_2},\mathbf{x_3}\}$ linearly dependent and for each case find the linear dependence.

This problem is square so we canuse the determinant idea:

$$\begin{aligned}
\det\begin{pmatrix}1 & 1 & 1\\ 1 & -1 & \alpha\\ 1 & 0 & 2\end{pmatrix} &= \det\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & \alpha-2\\ 1 & 0 & 0\end{pmatrix}\\
&= \det\begin{pmatrix}1 & -1\\ -1 & \alpha-2\end{pmatrix}\\
&= \det\begin{pmatrix}1 & -1\\ 0 & \alpha-3\end{pmatrix}\\
&= \alpha-3.
\end{aligned}$$

When $\alpha = 3$ the three vectors are linearly dependent:

$$\mathbf{x_3} = \begin{pmatrix}1\\3\\2\end{pmatrix} = 2\begin{pmatrix}1\\1\\1\end{pmatrix} = -\begin{pmatrix}1\\-1\\0\end{pmatrix} = 2\mathbf{x_1} - \mathbf{x_2}$$

and so

$$2\mathbf{x_1} - \mathbf{x_2} - \mathbf{x_3} = 0.$$

Eigenvalues and eigenvectors can be interpreted in this way.

$$H\lambda = \epsilon\lambda.$$

If we set $H - \epsilon I = [\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}]$ with the columns of $H - \epsilon I$ being interpreted as vectors, then the eigenvalue equation reads:

$$\lambda_1\mathbf{x_1} + \lambda_2\mathbf{x_2} + \cdots + \lambda_n\mathbf{x_n} = \mathbf{0}$$

which is once again the same problem. Eigenvalues occur when the vectors are linearly dependent, and the eigenvectors are the coefficients obtained by gaussian elimination.

□

Chapter 3

Inner Product

The next most useful concept for a vector space is an inner product, if it exists. An inner product leads quite directly to a concept akin to 'length', and is diectly related to measurement, and hence physics.

3.1 Formal Definition

A mapping from a pair of vectors to either a real number or a compex number is an inner product acting on a vector space V, providing certain axioms are obeyed. Denoted by $(\mathbf{x}, \mathbf{y})$, $< \mathbf{x}, \mathbf{y} >$, $\mathbf{x} \cdot \mathbf{y}$ (very poor), or $< \mathbf{x}|\mathbf{y} >$:

1. $(\mathbf{x}, \mathbf{y_1} + \mathbf{y_2}) = (\mathbf{x}, \mathbf{y_1}) + (\mathbf{x}, \mathbf{y_2})$

2. $(\mathbf{x_1} + \mathbf{x_2}, \mathbf{y}) = (\mathbf{x_1}, \mathbf{y}) + (\mathbf{x_2}, \mathbf{y})$

3. $(\mathbf{x}, \lambda\mathbf{y}) = \lambda(\mathbf{x}, \mathbf{y})$

4. $(\mathbf{x}, \mathbf{y})^* = (\mathbf{y}, \mathbf{x})$

and provide a physical 'magnitude' or 'length'

5. $(\mathbf{x}, \mathbf{x}) \geq 0$ is real and positive, with strict equality only when $\mathbf{x} = \mathbf{0}$.

A norm can be defined from an inner product satisfying 5., by:

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}. \tag{3.1}$$

This is the analogue of 'length'.

Note: $(\lambda\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \lambda\mathbf{x})^* = \lambda^*(\mathbf{y}, \mathbf{x})^* = \lambda^*(\mathbf{x}, \mathbf{y})$

3.1.1 Example: Complex numbers:

Define inner product by:

$$(x, y) \mapsto x^* y$$

(1) $x^*(y_1 + y_2) = x^* y_1 + x^* y_2$

(2) $(x_1 + x_2)^* y = x_1^* y + x_2^* y$

(3) $x^*(\lambda y) = \lambda x^* y$

(4) $(x^* y)^* = y^* x$

(5) $x^* x = 0 \quad \Rightarrow \quad x = 0$

$$\|\mathbf{x}\| = |x|.$$

□

3.1.2 Example: The Dot Product on Three-Dimensional Real Vectors

$$(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} \cdot \mathbf{y}$$

(1) $\mathbf{x} \cdot (\mathbf{y_1} + \mathbf{y_2}) = \mathbf{x} \cdot \mathbf{y_1} + \mathbf{x} \cdot \mathbf{y_2}$

(2) $(\mathbf{x_1} + \mathbf{x_2}) \cdot \mathbf{y} = \mathbf{x_1} \cdot \mathbf{y} + \mathbf{x_2} \cdot \mathbf{y}$

(3) $\mathbf{x} \cdot (\lambda \mathbf{y}) = \lambda \mathbf{x} \cdot \mathbf{y}$

(4) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$

(5) $\mathbf{x} \cdot \mathbf{x} = 0 \quad \Rightarrow \quad \mathbf{x} = \mathbf{0}$

$$\|\mathbf{x}\| = |\mathbf{x}|.$$

□

3.1.3 Relation Between Norm and Inner Product for real Inner Product

For a real inner product, norm and inner product are interchangeable:

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|^2 = (\mathbf{x}+\mathbf{y},\mathbf{x}+\mathbf{y}) &= (\mathbf{x},\mathbf{x})+(\mathbf{x},\mathbf{y})+(\mathbf{y},\mathbf{x})+(\mathbf{y},\mathbf{y}) \\
\|\mathbf{x}-\mathbf{y}\|^2 = (\mathbf{x}-\mathbf{y},\mathbf{x}-\mathbf{y}) &= (\mathbf{x},\mathbf{x})-(\mathbf{x},\mathbf{y})-(\mathbf{y},\mathbf{x})+(\mathbf{y},\mathbf{y})
\end{aligned} \tag{3.2}$$

since $(\mathbf{x},\mathbf{y})^* = (\mathbf{x},\mathbf{y}) = (\mathbf{y},\mathbf{x})$ then:

$$(\mathbf{x},\mathbf{y}) = \frac{1}{4}\left[\|\mathbf{x}+\mathbf{y}\| - \|\mathbf{x}-\mathbf{y}\|\right] \tag{3.3}$$

For complex inner products, this is not in general true.

3.1.4 Schwartz's Inequality and the Triangle Inequality

Schwartz's inequality is:

$$|(\mathbf{x},\mathbf{y})| \leq \|\mathbf{x}\|\ \|\mathbf{y}\|. \tag{3.4}$$

Proof: 'Completing the square'

$$\begin{aligned}
\|\lambda\mathbf{x}+\mu\mathbf{y}\|^2 &= (\lambda\mathbf{x}+\mu\mathbf{y},\lambda\mathbf{x}+\mu\mathbf{y}) \\
&= |\lambda|^2\|\mathbf{x}\|^2 + |\mu|^2\|\mathbf{y}\|^2 + \lambda^*\mu(\mathbf{x},\mathbf{y}) + \lambda\mu^*(\mathbf{y},\mathbf{x}) \geq 0
\end{aligned} \tag{3.5}$$

choose

$$|\lambda| = \|\mathbf{y}\|, \quad |\mu| = \|\mathbf{x}\| \qquad \text{and} \qquad \lambda\mu^* = -\|\mathbf{x}\|\ \|\mathbf{y}\|\frac{(\mathbf{x},\mathbf{y})}{|(\mathbf{x},\mathbf{y})|}$$

then

$$2\|\mathbf{x}\|\ \|\mathbf{y}\| - 2|(\mathbf{y},\mathbf{x})|\ \|\mathbf{x}\|\ \|\mathbf{y}\| \geq 0$$

and hence $|(\mathbf{x},\mathbf{y})| \leq \|\mathbf{x}\|\ \|\mathbf{y}\|$.

Triangle Inequality

The triangle inequality is:

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| \ \|\mathbf{y}\|. \tag{3.6}$$

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}) \\ &= \|\mathbf{x}\| + \|\mathbf{y}\| + (\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{y})^* \\ &= \|\mathbf{x}\| + \|\mathbf{y}\| + 2Re\ (\mathbf{x}, \mathbf{y}) \\ &\leq \|\mathbf{x}\| + \|\mathbf{y}\| + 2|(\mathbf{x}, \mathbf{y})| \\ &\leq \|\mathbf{x}\| + \|\mathbf{y}\| + \|\mathbf{x}\| \ \|\mathbf{y}\| \\ &= [\|\mathbf{x}\| + \|\mathbf{y}\|]^2 \end{aligned} \tag{3.7}$$

where we used the Schwartz's inequality.

Chapter 4

Orthogonality

Two vector are said to be orthogonal if their inner product vanishes, $(\mathbf{v}, \mathbf{v}')$. A basis is said to be orthogonal if each and every pair is independently orthogonal, $(\mathbf{x_i}, \mathbf{x_j}) = 0$ for $i \neq j$. A vector is said to be normal if its norm is unity, $(\mathbf{v}, \mathbf{v}) = 1$ and $\|\mathbf{v}\| = 1$. A basis is said to be orthonormal if it is orthogonal and each basis vector is normal. Choosing an orthonormal basis simplifies an inner product.

Let us think about the inner product interms of components. If:

$$\mathbf{v} = \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n} = \sum_{m=1}^{n} \lambda_m \mathbf{x}_m$$

$$\mathbf{v}' = \lambda'_1 \mathbf{x_1} + \lambda'_2 \mathbf{x_2} + \cdots + \lambda'_n \mathbf{x_n} = \sum_{m=1}^{n} \lambda'_m \mathbf{x}_m$$

then

$$\begin{aligned}
(\mathbf{v}, \mathbf{v}') &= \sum_{m=1}^{n} \sum_{m'=1}^{n} (\lambda_m \mathbf{x}_m, \lambda'_{m'} \mathbf{x}_{m'}) \\
&= \sum_{m=1}^{n} \sum_{m'=1}^{n} \lambda^*_m \lambda'_{m'} (\mathbf{x}_m, \mathbf{x}_{m'}) \\
&= \sum_{m=1}^{n} \sum_{m'=1}^{n} \lambda^*_m O_{mm'} \lambda'_{m'} = \lambda^\dagger O \lambda' \qquad (4.1)
\end{aligned}$$

where we have introduced the overlap matrix, $O_{ij} = (\mathbf{x}_i, \mathbf{x}_j)$.

$$
\begin{aligned}
& [\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*] \begin{bmatrix} O_{11} & O_{12} \cdots & O_{1n} & \\ O_{21} & O_{22} \cdots & O_{2n} & \\ \cdot & \cdot & \cdots & . \\ \cdot & \cdot & \cdots & . \\ O_{n1} & O_{n2} \cdots & O_{nn} & \end{bmatrix} \begin{bmatrix} \lambda_1' \\ \lambda_2' \\ \cdot \\ \cdot \\ \lambda_n' \end{bmatrix} \\
= & [\lambda_1^*, \lambda_2^*, \ldots, \lambda_n^*] \begin{bmatrix} \sum_{m'=1}^n O_{1m'} \lambda_{m'}' \\ \sum_{m'=1}^n O_{2m'} \lambda_{m'}' \\ \cdot \\ \cdot \\ \sum_{m'=1}^n O_{nm'} \lambda_{m'}' \end{bmatrix} \\
= & \sum_{m=1}^n \sum_{m'=1}^n \lambda_m^* O_{mm'} \lambda_{m'}'. \qquad (4.2)
\end{aligned}
$$

If we are dealing with an orthogonal basis, then $O_{ij} = (\mathbf{x}_i, \mathbf{x}_j) = 0$ if $i \neq j$ and O_{ij} is diagonal.

$$(\mathbf{v}, \mathbf{v}') = \sum_{m=1}^n \lambda_m^* \lambda_m' O_{mm}.$$

If we are dealing with an orthonormal basis, then additionally $O_{mm} = (\mathbf{x}_m, \mathbf{x}_m) = 1$ and so:

$$(\mathbf{v}, \mathbf{v}') = \sum_{m=1}^n = \lambda_m^* \lambda_m' = \lambda_1^* \lambda_1' + \lambda_2^* \lambda_2' + \cdots + \lambda_n^* \lambda_n'.$$

This is the ‘usual’ representation of ‘length’ for real and complex numbers. [Viz the dot product].

4.1 Physical Space

For physical space, the ‘norm’ $\|\mathbf{x}\|$ is just the length of the vector $\mathbf{x}$. $(\mathbf{x}, \mathbf{y}) = x_1 y_1 + x_2 y_2 + x_3 y_3$ because we are dealing with real numbers.

For physical space-time, we use an inner product whcih does not have an immediate norm,

$$
\begin{aligned}
\mathbf{x} = (ct, x_1, x_2, x_3) \quad \mathbf{x}' = (ct', x_1', x_2', x_3') \\
(\mathbf{x}, \mathbf{x}') = c^2 tt' - x_1 x_1' - x_2 x_2' - x_3 x_3'. \qquad (4.3)
\end{aligned}
$$

This obeys all the axioms except 5. Due to the fact that we have a real vector space, we can use $\|\mathbf{x}\|^2$ interchangeably with the inner product, but this quantity can take positive (time-like) or negative (space-like) values. When $\|\mathbf{x}\|^2$ is positive, we can define $\tau = \|\mathbf{x}\|/c$, known as proper time, and measuring the time spent by an observer traveling at uniform velocity from the origin to the point $\mathbf{r}$ (three-dimensional) at time t, $\mathbf{r} = \mathbf{v}t$ and so:

$$c^2\tau^2 = c^2t^2 - t^2\mathbf{v}\cdot\mathbf{v} = c^2t^2\left(1 - \frac{v^2}{c^2}\right) \tag{4.4}$$

and $\tau = \sqrt{[1 - \frac{v^2}{c^2}]}t$, the correctly time-dilated time. It is this quantity which is a physically measurable constant in all inertial frames. Note that there is a horrible problem with $\tau = 0$ and the sign to choose for τ. This is a real physical problem and is tied to the fundamentals of special relativity.

The laws of physics are usually simpler in an orthonormal basis, and so we often restrict attention to such a basis, how can you orthogonal basis. The decision leads to a problem: If you have an arbitrary non-orthogonal basis, how can you orthogonalise it?

Chapter 5

Gram-Schmidt Orthogonalisation

This is a method for selecting an orthogonal basis for a vector-space spanned by a given collection of vectors. The basic idea is simple, order the vectors and then orthogonalise each vector in turn with all the preceding vectors.

If the given vectors are , $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$, then set

$$\mathbf{n_1} = \mathbf{v_1} \tag{5.1}$$

followed by

$$\mathbf{n_2} = \mathbf{v_2} - a_{21}\mathbf{n_1} \tag{5.2}$$

and then choose a_{21} so that $(\mathbf{n_1}, \mathbf{n_2}) = 0$, so:

$$a_{21} = \frac{(\mathbf{n_1}, \mathbf{v_2})}{(\mathbf{n_1}, \mathbf{n_1})} \tag{5.3}$$

then set

$$\mathbf{n_3} = \mathbf{v_3} - a_{31}\mathbf{n_1} - a_{32}\mathbf{n_2} \tag{5.4}$$

and choose a_{31} so that $(\mathbf{n_1}, \mathbf{n_3}) = 0$, and $_{32}$ so that $(\mathbf{n_2}, \mathbf{n_3}) = 0$. Hence

$$a_{31} = \frac{(\mathbf{n_1}, \mathbf{v_3})}{(\mathbf{n_1}, \mathbf{n_1})} \tag{5.5}$$

and

$$a_{31} = \frac{(\mathbf{n_2}, \mathbf{v_3})}{(\mathbf{n_2}, \mathbf{n_2})} \tag{5.6}$$

and repeat the process.

In general,

$$\mathbf{n}_m = \mathbf{v}_m - \sum_{r=1}^{m-1} a_{mr}\mathbf{n_r} \tag{5.7}$$

and

$$a_{mr} = \frac{(\mathbf{n_r}, \mathbf{v}_m)}{(\mathbf{n_r}, \mathbf{n_r})}. \tag{5.8}$$

You get an orthogonal basis, but the basis depends on the order of the original vectors.

There is one minor problem, $(\mathbf{n_r}, \mathbf{n_r}) = 0$ for some r. This possibility corresponds to $\|\mathbf{n_r}\| = 0$, which means that $\mathbf{n_r} = 0$, and hence that $\mathbf{v_r}$ is linearly dependent on the previous vectors, i.e. $\mathbf{v_r} = \sum_{p=1}^{r-1} a_{rp}\mathbf{n_p}$, and hence should not be included in the basis.

You can orthonormalise if desired by setting, $\mathbf{n_i} \mapsto \mathbf{n_i}/\|\mathbf{n_i}\|$. In practice, it is useful to note that:

$$\begin{aligned}
(\mathbf{n_m}, \mathbf{n_m}) &= ([\mathbf{v}_m - \sum_{r=1}^{m-1} a_{mr}\mathbf{n_r}], \mathbf{n_m}) \\
&= (\mathbf{v}_m, \mathbf{n_m}) - \sum_{r=1}^{m-1} a^*_{mr}(\mathbf{n_r}, \mathbf{n_m}) \\
&= (\mathbf{v}_m, \mathbf{n_m}) \\
&= (\mathbf{v}_m, \mathbf{v}_m - \sum_{r=1}^{m-1} a_{mr}\mathbf{n_r}) \\
&= (\mathbf{v}_m, \mathbf{v}_m) - \sum_{r=1}^{m-1} a_{mr}(\mathbf{v}_m, \mathbf{n_r}) \\
&= (\mathbf{v}_m, \mathbf{v}_m) - \sum_{r=1}^{m-1} \frac{(\mathbf{v}_m, \mathbf{n_r})(\mathbf{n_r}, \mathbf{v}_m)}{(\mathbf{n_r}, \mathbf{n_r})}
\end{aligned} \tag{5.9}$$

(where we used (5.7) and (5.8)) so we only need $(\mathbf{v}_m, \mathbf{n_r})$ and the previous $(\mathbf{n_r}, \mathbf{n_r})$.

5.1 Diagonalise the Overlap Matrix

A second method for producing an orthogonal collection of vectors is to diagonalise the overlap matrix. A new orthogonal basis can be constructed from the eigenvectors corresponding to the non-vanishing eigenvalues.

$O_{ij} = (\mathbf{v}_i, \mathbf{v}_j)$ diagonalise via,

$$\sum_{j=1}^{n} O_{ij} \lambda_j^{(m)} = o^{(m)} \lambda_i^{(m)} \tag{5.10}$$

and then

$$\mathbf{n}^{(m)} = \sum_{j=1}^{n} \lambda_j^{(m)} \mathbf{v_j}, \qquad o^{(m)} \neq 0 \tag{5.11}$$

Proof:

$$\begin{aligned}
(\mathbf{n}^{(m)}, \mathbf{n}^{(m')}) &= \sum_{i=1}^{n} \sum_{j=1}^{n} (\lambda_i^{(m)} \mathbf{v_i}, \lambda_j^{(m')} \mathbf{v_j}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{(m)*} \lambda_j^{(m')} (\mathbf{v_i}, \mathbf{v_j}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{(m)*} O_{ij} \lambda_j^{(m')} \\
&= \sum_{i=1}^{n} \lambda_i^{(m)*} o^{(m')} \lambda_i^{(m')} \\
&= o^{(m)} \delta_{mm'} \sum_{i=1}^{n} \lambda_i^{(m)*} \lambda_i^{(m)} \qquad (5.12)
\end{aligned}$$

where we have used that for a Hermitian matrix eigenvectors are orthogonal, and $(\mathbf{v_i}, \mathbf{v_j})^* = (\mathbf{v_j}, \mathbf{v_i})$ is an axiom for the inner product.

Orthonormal states can be defined as

$$|\tilde{\mathbf{n}}^{(m)}\rangle = \frac{1}{\sqrt{o^{(m)} \sum_{i=1}^{N} \lambda_i^{(m)*} \lambda_i^{(m)}}} \sum_{i=1}^{N} \lambda_i^{(m)} |\mathbf{v}_i\rangle.$$

5.1.1 A list of vectors is linearly independent if and only if its overlap matrix is invertible

The following result proves that if the original states $\{|\mathbf{v}_i\rangle\}_{i=1,\ldots,n}$ are linearly independent then the eigenvalues are all non-zero, and hence $\{|\mathbf{n}^{(m)}\rangle\}_{m=1,\ldots,n}$ form an orthogonal basis and $\{|\tilde{\mathbf{n}}^{(m)}\rangle\}_{m=1,\ldots,n}$ form an orthonormal basis. This follows from the fact that the overlap matrix is invertible if and only if the eigenvalues are all non-zero.

Theorem: A list of vectors is linearly independent if and only if its overlap matrix is invertible.

Proof:

First we prove that invertibility of the Overlap matrix implies linear independence. Suppose that $\sum_{j=1}^{N} c_j|\mathbf{v}_j\rangle = 0$. Take the inner product with $\langle\mathbf{v}_i|$. This gives

$$\sum_{j=1}^{N} O_{ij}c_j = 0$$

for each i. Since the matrix is invertible, it follows that the coefficients c_j are all zero. This proves linear independence.

Then we prove that linear independence implies that the overlap matrix is invertible. Consider a coordinate vector of c_j that is in the null space of the matrix O, that is, such that

$$\sum_{j=1}^{N} O_{ij}c_j = 0$$

for each i. It follows that

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i^* O_{ij}c_j = 0$$

or

$$\langle\sum_{i=1}^{N} c_i\mathbf{v}_i|\sum_{j=1}^{N} c_j\mathbf{v}_j\rangle = 0$$

or

$$\|\sum_{i=1}^{N} c_i|\mathbf{v}_i\rangle\|^2 = 0$$

or

$$\sum_{i=1}^{N} c_i|\mathbf{v}_i\rangle = 0.$$

By linear dependence the coefficients c_j are all zero. This proves that the null space of O is trivial - i.e. it consists of the zero vector only, meaning there are no eigenvectors with zero eigenvalue. Therefore O is invertible.

□

5.1.2 Note on Hermitian matrices

In this section we assume all the eigenvalues are distinct. In appendix C we treat the general case where eigenvalues are degenerate.

For a matrix, Hermitian means:

$$H^*_{ji} = H_{ij} \tag{5.13}$$

and then

$$\sum_j H_{ij}\lambda_j^{(m)} = o^{(m)}\lambda_i^{(m)} \tag{5.14}$$

Taking the complex conjugate of this and using the Hermitian property implies

$$\sum_j \lambda_j^{(m)*} H_{ji} = o^{(m)*}\lambda_i^{(m)*}$$

We also have the eigenvector equation

$$\sum_j H_{ij}\lambda_j^{(m')} = o^{(m')}\lambda_i^{(m')}$$

with label m' instead of m. These can be combined to show that:

$$\sum_{ij} \lambda_i^{(m)*} H_{ij} \lambda_j^{(m')} = o^{(m)*} \sum_i \lambda_i^{(m)*} \lambda_i^{(m')} = o^{(m')} \sum_i \lambda_i^{(m)*} \lambda_i^{(m')}$$

and consequently:

$$\left(o^{(m)*} - o^{(m')} \right) \sum_i \lambda_i^{(m)*} \lambda_i^{(m')} = 0. \tag{5.15}$$

This means that when $m' = m$ we have:

$$o^{(m)*} = o^{(m)} \tag{5.16}$$

and so the eigenvalues are real, whereas $o^{(m)} \neq o^{(m')}$ leads to:

$$\sum_i \lambda_i^{(m)*} \lambda_i^{(m')} = 0 \tag{5.17}$$

and the eigenvectors are automatically orthogonal for different eigenvalues. We will encounter this basic idea several more times in these notes.

5.1.3 Example

In an orthonormal basis, three vectors are $\mathbf{v_1} = (1, 1, 1)$, $\mathbf{v_2} = (0, 1, 1)$ and $\mathbf{v_3} = (0, 0, 1)$.

□

Gram-Schidt orthogonalisation:

$$\mathbf{n_1} = (1, 1, 1)$$

$$\mathbf{n_2} = \mathbf{v_2} - a_{21}\mathbf{n_1} = (0, 1, 1) - a_{12}(1, 1, 1)$$

from $(\mathbf{n_1}, \mathbf{n_2}) = 0$ we have $2 = 3a_{21}$, so $a_{21} = 2/3$ and

$$\mathbf{n_2} = \frac{1}{3}(-2, 1, 1)$$

$$\mathbf{n_3} = \mathbf{v_3} - a_{31}\mathbf{n_1} - a_{32}\mathbf{n_2} = (0, 0, 1) - a_{31}(1, 1, 1) - a_{32}\frac{1}{3}(-2, 1, 1)$$

from $(\mathbf{n_1}, \mathbf{n_3}) = 0$ we have $1 = 3a_{31}$, so $a_{31} = 1/3$. From $(\mathbf{n_2}, \mathbf{n_3}) = 0$ we have $1/3 = a_{32}6/9$ so $a_{32} = 1/2$. So

$$\mathbf{n_3} = \frac{1}{2}(0, -1, 1)$$

If we normalise, then

$$\hat{\mathbf{n}}_\mathbf{1} = \frac{1}{\sqrt{3}}(1, 1, 1), \quad \hat{\mathbf{n}}_\mathbf{2} = \frac{1}{\sqrt{6}}(-2, 1, 1), \quad \hat{\mathbf{n}}_\mathbf{3} = \frac{1}{\sqrt{2}}(0, -1, 1).$$

The overlap matrix

The overlap matrix $O_{ij} = (\mathbf{v_i}, \mathbf{v_j})$ is

$$O_{ij} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{5.18}$$

is very difficult to diagonalise.

If $\mathbf{v_1} = (1, 1, 1)$, $\mathbf{v_2} = (0, 1, 1)$ and $\mathbf{v_3} = (0, 0, 1)$, then

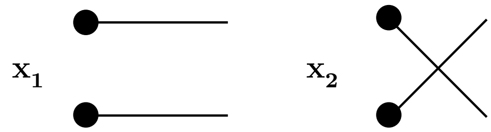
$$\begin{aligned} \mathbf{n_1} &= (0, 0, 1) \\ \mathbf{n_2} &= (0, 1, 1) - (0, 0, 1) = (0, 1, 0) \\ \mathbf{n_3} &= (1, 1, 1) - (0, 0, 1) - (0, 1, 0) = (1, 0, 0) \end{aligned} \tag{5.19}$$

a quite different result.

5.1.4 Example: Valence-bond states

Valence-bond states for many spin-half atoms. When many spin-half atoms are added together, there are a range of possible values for the total spin. One way to restrict attention to total-spin singlets, viz $\mathbf{S_1} + \mathbf{S_2} + \cdots + \mathbf{S_n} = \mathbf{0}$, is to pair up spins into singlet pairs, $\mathbf{S_i} + \mathbf{S_j} = \mathbf{0}$. Such a basis is non-orthogonal. If we split the spins into two halves and only pair between halves, then it is possible to show that the overlap between states is $2^{\#\{Cluster\} - \#\{Singlets\}}$ where $\#\{Cluster\}$ is the number of disconnected clusters found by cross-connecting the two types of pairing,and $\#\{Singlets\}$ is the number of singlets.

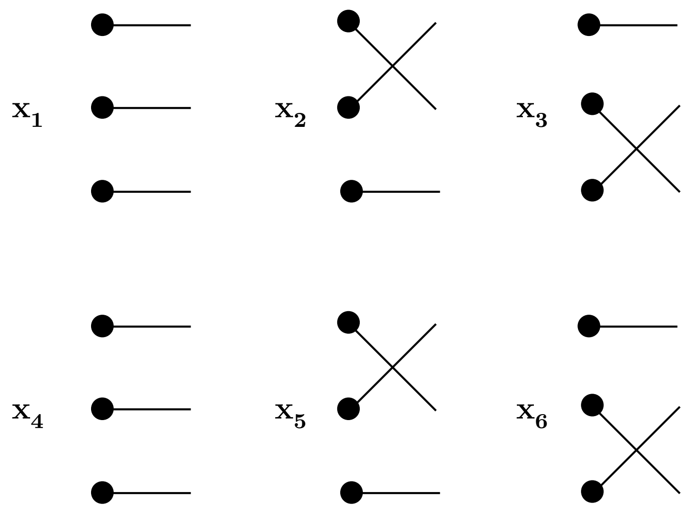
(i) Four spins



$$(\mathbf{x_i}, \mathbf{x_i}) = 1 \qquad (\mathbf{x_1}, \mathbf{x_2}) = \frac{1}{2}$$

$$\mathbf{n_1} = \mathbf{x_1} \quad \mathbf{n_2} = \mathbf{x_2} - \frac{1}{2}\mathbf{n_1} \quad (\mathbf{n_2}, \mathbf{n_2}) = 1 - \frac{1}{4} = \frac{3}{4}.$$

(ii) Six spins



$$\mathbf{n_1} = \mathbf{x_1} \quad \mathbf{n_2} = \mathbf{x_2} - \frac{1}{2}\mathbf{n_1} \quad \mathbf{n_3} = \mathbf{x_3} - \frac{1}{2}\mathbf{n_1} \quad \mathbf{n_4} = \mathbf{x_4} - \frac{1}{2}\mathbf{n_1}$$

since $(\mathbf{n_1}, \mathbf{x_i}) = \frac{1}{2}$ for $i = 1, 2, 3, 4$, $(\mathbf{x_i}, \mathbf{x_j}) = \frac{1}{4}$ for $i \neq j \in \{2, 3, 4\}$, and:

$$(\mathbf{n_2}, \mathbf{n_2}) = (\mathbf{x_2}, \mathbf{x_2}) - \frac{(\mathbf{x_2}, \mathbf{n_1})(\mathbf{n_1}, \mathbf{x_2})}{(\mathbf{n_1}, \mathbf{n_1})} = \frac{3}{4}$$

$(\mathbf{n_2}, \mathbf{x_5}) = \frac{1}{4}$ and $(\mathbf{n_i}, \mathbf{x_5}) = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$ for $i = 2, 3, 4,$, and so:

$$\mathbf{n_5} = \mathbf{x_5} - \frac{1}{4}\mathbf{n_1} - \frac{1}{2}(\mathbf{n_2} + \mathbf{n_3} + \mathbf{n_4})$$

$$(\mathbf{n_5}, \mathbf{n_5}) = 1 - \frac{1}{16} - 3 \times \left[\frac{3}{8}\right]^2 \times \frac{4}{3} = \frac{3}{8}$$

and since $(\mathbf{n_1}, \mathbf{x_6}) = \frac{1}{4}$ and $(\mathbf{n_i}, \mathbf{x_6}) = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}$ for $i = 2, 3, 4$, and $(\mathbf{n_5}, \mathbf{x_6}) = \frac{1}{4} - \frac{1}{4} \times \frac{1}{4} - \frac{3}{2} \times \frac{3}{8} = -\frac{3}{8}$, so:

$$\mathbf{n_6} = \mathbf{x_6} - \frac{1}{4}\mathbf{n_1} - \frac{1}{2}(\mathbf{n_2} + \mathbf{n_3} + \mathbf{n_4}) + \mathbf{n_5}$$

$$(\mathbf{n_6}, \mathbf{n_6}) = 1 - \frac{1}{16} - 3 \times \left[\frac{3}{8}\right]^2 \times \frac{4}{3} - \left[\frac{-3}{8}\right]^2 \times \frac{8}{3} = 0$$

$\mathbf{n_6} = 0$ and we have a linear dependence!

$$\mathbf{x_6} = \frac{1}{4}\mathbf{n_1} + \frac{1}{2}(\mathbf{n_2} + \mathbf{n_3} + \mathbf{n_4}) - \mathbf{n_5}$$

$$\mathbf{x_1} + \mathbf{x_5} + \mathbf{x_6} = \mathbf{x_2} + \mathbf{x_3} + \mathbf{x_4}$$

Note:

$$O_{ij} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & 1 \end{pmatrix}$$

with eigenvalues and eigenvectors:

$$E = \frac{3}{4} \quad \mathbf{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \begin{pmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ -2 \\ 0 \\ 0 \end{pmatrix} \quad E = 3 \quad \mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad E = 0 \quad \mathbf{v} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ -1 \\ 1 \\ 1. \end{pmatrix}$$

A matrix which is easy to diagonalise.

Chapter 6

Linear Operators and Matrices

Formal Definition: A linear operator, $\hat{L}(\mathbf{x})$ say, is a mapping from the vector space back into the vector space, satisfying some axioms,

1. $\hat{L}(\mathbf{x}+\mathbf{y}) = \hat{L}(\mathbf{x}) + \hat{L}(\mathbf{y})$

2. $\hat{L}(\lambda\mathbf{x}) = \lambda\hat{L}(\mathbf{x})$.

Linear operators are mappings of the space onto itself, including transformations of the vector space. They include rotations, reflections, dilitations, shears and projections.

In a particular basis, any linear operator can be represented by a matrix. The basis vectors are mapped onto vectors which are this matrix 'multiplied' by the original basis vectors, and the components of a vector are matrix multiplied by the transpose of this matrix.

Proof:

Given a basis $\{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$, we can consider the action of $\hat{L}$ on this basis. $\hat{L}(\mathbf{x_i})$ is a vector in the vector space, and so can be represented with the basis:

$$\hat{L}(\mathbf{x_i}) = \mathbf{x_1}L_{1i} + \mathbf{x_2}L_{2i} + \cdots + \mathbf{x_n}L_{ni} = \sum_{m=1}^{n} \mathbf{x}_m L_{mi} \tag{6.1}$$

where we have labeled the components with a first label for the first basis vector and a second label for the initial basis vector. The action of $\hat{L}$ on our basis can be represented by the matrix L_{ij}.

Any vector, $\mathbf{v}$ say, can be represented in the basis

$$\mathbf{v} = \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n} = \sum_{m=1}^{n} \lambda_m \mathbf{x}_m. \tag{6.2}$$

The action of $\hat{L}$ on $\mathbf{v}$ is then

$$\begin{aligned}
\hat{L}(\mathbf{v}) &= \hat{L}(\lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n}) \\
&= \lambda_1 \hat{L}(\mathbf{x_1}) + \lambda_2 \hat{L}(\mathbf{x_2}) + \cdots + \lambda_n \hat{L}(\mathbf{x_n}) \\
&= \sum_{m=1}^{n} \lambda_m \hat{L}(\mathbf{x}_m) \\
&= \sum_{m=1}^{n} \lambda_m \sum_{m'=1}^{n} \mathbf{x}_{m'} L_{m'm} \\
&= \sum_{m'=1}^{n} \mathbf{x}_{m'} (\sum_{m=1}^{n} L_{m'm} \lambda_m)
\end{aligned} \tag{6.3}$$

so the coefficients of $\hat{L}(\mathbf{v})$ are

$$(\hat{L}(\mathbf{v}))_m = \sum_{m'=1}^{n} L_{mm'} \lambda_{m'} \tag{6.4}$$

the matrix L_{ij} multiplied by the original components.

Note: The matrix L post-multiplies by the original components, but pre-multiplies the components. This corresponds to a transpose.

In a chosen basis, linear operators act on the components as matrices under matrix multiplication.

If we have an inner product on our space, then we can find the matrix L_{ij}

$$\hat{L}(\mathbf{x_i}) = \sum_{m=1}^{n} \mathbf{x}_m L_{mi} \tag{6.5}$$

applying the inner product

$$
\begin{aligned}
(\mathbf{x_j}, \hat{L}(\mathbf{x_i})) \equiv (\mathbf{x_j}, \hat{L}\mathbf{x_i}) &= \\
&= (\mathbf{x_j}, \sum_{m=1}^{n} \mathbf{x}_m L_{mi}) \\
&= \sum_{m=1}^{n} L_{mi} (\mathbf{x_j}, \mathbf{x}_m) \\
&= \sum_{m=1}^{n} O_{jm} L_{mi} \qquad (6.6)
\end{aligned}
$$

which is the matrix multiplication of the overlap matrix with L_{ij}. If the basis is orthonormal, then

$$
L_{ij} = (\mathbf{x_i}, \hat{L}\mathbf{x_j}). \qquad (6.7)
$$

Chapter 7

Multiple Maps and Matrices

Let us consider the application of two linear oerators, $\mathbf{A}(\mathbf{x})$ and $\mathbf{B}(\mathbf{x})$, sequentially. We need only consider the action on a basis, and hence the matrix representation:

$$\mathbf{A}(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_j} A_{ji} \qquad \text{and} \qquad \mathbf{B}(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_j} B_{ji} \tag{7.1}$$

Composition leads to

$$\begin{aligned} \mathbf{B}(\mathbf{A}(\mathbf{x_j})) &= \mathbf{B}\left(\sum_{k=1}^{n} \mathbf{x_k} A_{kj}\right) \\ &= \sum_{k=1}^{n} \mathbf{B}(\mathbf{x_k}) A_{kj} \\ &= \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \mathbf{x_i} B_{ik}\right) A_{kj} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{n} \mathbf{x_i} B_{ik} A_{kj} \end{aligned} \tag{7.2}$$

and the matrix representation for the linear operator resulting from the original pair is:

$$C_{ij} = \sum_{k=1}^{n} B_{ik} A_{kj}. \tag{7.3}$$

We can envisage this using square arrays:

$$\begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1k} & \dots & B_{1n} \\ B_{21} & B_{22} & \cdots & B_{2k} & \dots & B_{2n} \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ B_{i1} & B_{i2} & \cdots & B_{ik} & \dots & B_{in} \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ B_{n1} & B_{n2} & \cdots & B_{nk} & \dots & B_{nn} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} & \dots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2k} & \dots & A_{2n} \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ A_{i1} & A_{i2} & \cdots & A_{ik} & \dots & A_{in} \\ \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ A_{n1} & A_{n2} & \cdots & A_{nk} & \dots & A_{nn} \end{bmatrix} \tag{7.4}$$

This matrix is the matrix multiplication of the original matrices.

Chapter 8

Hermitian Operators and Matrices

A linear operator is said to be Hermitian if

$$(\hat{H}(\mathbf{x}), \mathbf{y}) = (\mathbf{x}, \hat{H}(\mathbf{y})) \quad \left[(\hat{H}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \hat{H}\mathbf{y})\right] \tag{8.1}$$

i.e. the operator can act on either of the two vectors in the inner product.

$$H_{ij} = (\mathbf{x_i}, \hat{H}\mathbf{x_j}) = (\hat{H}\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_j}, \hat{H}\mathbf{x_i})^* = H^*_{ji} \tag{8.2}$$

for an orthonormal basis, i.e. the matrix is Hermitian. This is not so in with a general basis.

Note that in a general basis (remember $\hat{H}\mathbf{x_j} = \sum_{m=1}^{n} \mathbf{x_m} H_{mj}$)

$$\begin{aligned} \sum_{m=1}^{n} O_{im} H_{mj} = (\mathbf{x_i}, \hat{H}\mathbf{x_j}) &= (\hat{H}\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_j}, \hat{H}\mathbf{x_i})^* \\ &= \sum_{m=1}^{n} O^*_{jm} H^*_{mi} \end{aligned} \tag{8.3}$$

and in matrix notation,

$$OH = (OH)^{*T} = (OH)^\dagger = H^\dagger O^\dagger = H^\dagger O,$$

since $(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_j}, \mathbf{x_i})^*$. This is complicated to work with.

With a Hermitian operator, notation can be include the idea that the operator can act either side, leading to a reflection-symmetric notation:

$$(\mathbf{x}, \hat{H}\mathbf{y}) \equiv < \mathbf{x}|\hat{H}|\mathbf{y} > = < \mathbf{y}|\hat{H}|\mathbf{x} >^* . \tag{8.4}$$

8.1 Basis

Any linear operator can be uniquely represented by its action on a basis, $\{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}\}$.

Proof:

If $\mathbf{L_i} = \hat{L}(\mathbf{x_i})$ and $\mathbf{x} = \sum_{i=1}^{n} \lambda_i \hat{L}(\mathbf{x_i}) = \sum_{i=1}^{n} \lambda_i \mathbf{L_i}$

and $\hat{L}(\mathbf{x})$ has been represented in terms of $\hat{L}(\mathbf{x_i})$.

Chapter 9

Transformation and a Change of Basis

A transformation is a linear operator mapping the vector space onto itself, but with the additional requirement that every vector is mapped onto. This additional requirement ensures that the transformation is invertible.

Proof:

Firstly, it is clear that $T_i = \hat{T}(\mathbf{x_i})$ spans the space mapped to, since:

$$\hat{T}(\sum_{m=1}^{n} \lambda_m \mathbf{x_m}) = \sum_{m=1}^{n} \lambda_m \hat{T}(\mathbf{x_m}) = \sum_{m=1}^{n} \lambda_m \mathbf{T}_m. \tag{9.1}$$

If the $\mathbf{T_i}$ are linearly dependent, then the space mapped to is 'smaller' than the original (of reduced dimension - not every vector is mapped onto). If the $\mathbf{T_i}$ are linearly independent then they form a basis and so any vector can be represented by the $\mathbf{T_i}$. The linear transformation which uniquely maps the $\mathbf{T_i}$ onto $\mathbf{x_i}$, $\hat{\mathbf{T}}^{-1}$ say, is the inverse of $\hat{\mathbf{T}}$, i.e. $\hat{\mathbf{T}}^{-1}(\hat{T}_i) = \mathbf{x_i}$.

If $\hat{\mathbf{T}}(\mathbf{x}) = \mathbf{x}'$ and $\mathbf{x} = \sum_{m=1}^{n} \lambda_m \mathbf{x}_m$, $\mathbf{x}' = \sum_{m=1}^{n} \lambda'_m \mathbf{x}_m$, then

$$\mathbf{x}' = \sum_{m=1}^{n} \lambda_m \mathbf{T}_m$$

$$\hat{\mathbf{T}}^{-1}(\mathbf{x}') = \sum_{m=1}^{n} \lambda_m \hat{\mathbf{T}}^{-1}(\mathbf{T}_m) = \sum_{m=1}^{n} \lambda_m \mathbf{x}_m = \mathbf{x}$$

A transformation is just a change of basis, and can be represented by a non-singular matrix.

9.0.1 Example:

An original non-orthogonal basis:

$$\mathbf{v_1} = (1,1,1) \qquad \mathbf{v_2} = (0,1,1) \qquad \mathbf{v_3} = (0,0,1)$$

is Gram-Schidt orthogonalised to an orthonormal basis:

$$\mathbf{n_1} = \frac{1}{\sqrt{3}}(1,1,1) \qquad \mathbf{n_2} = \frac{1}{\sqrt{6}}(-2,1,1) \qquad \mathbf{n_3} = \frac{1}{\sqrt{2}}(0,-1,1).$$

Find the matrix representation for this transformation.

We see that:

$$\mathbf{n_1} = \frac{1}{\sqrt{3}}\mathbf{v_1} \qquad \mathbf{n_2} = \frac{3}{\sqrt{6}}\left[\mathbf{v_2} - \frac{2}{3}\mathbf{v_1}\right] \qquad \mathbf{n_3} = \frac{2}{\sqrt{2}}\left[\mathbf{v_3} - \frac{1}{2}\mathbf{v_2}\right]$$

and hence:

$$\mathbf{n_i} = \sum_{j=1}^{3} \mathbf{v_j} N_{ji} \tag{9.2}$$

leads to

$$N_{ji} = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 & \sqrt{2} \end{bmatrix} \tag{9.3}$$

The inverse transformation satisfies:

$$\mathbf{v_i} = \sum_{j=1}^{3} \mathbf{n_j} N_{ji}^{-1} \tag{9.4}$$

with

$$N_{ji}^{-1} = \begin{bmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{9.5}$$

Chapter 10

Unitary Transformations

In order to simplify the inner product, we usually use an orthonormal basis. There is a special class of transformations which preserve the orthonormality of a basis, the unitary transformations.

A transformation is said to be unitary if it preserves the inner product:

$$(\hat{T}(\mathbf{x}), \hat{T}(\mathbf{y})) = (\mathbf{x}, \mathbf{y}). \tag{10.1}$$

This need only be tested on a basis, since:

$$(\hat{T}(\mathbf{x}), \hat{T}(\mathbf{y})) - (\mathbf{x}, \mathbf{y}) = \sum_{m=1}^{n} \sum_{m'=1}^{n} \lambda_m^* \lambda_{m'} \left[(\hat{T}(\mathbf{x}_m), \hat{T}(\mathbf{x}_m)) - (\mathbf{x}_m, \mathbf{x}_{m'}) \right] \tag{10.2}$$

and so if it vanishes on the basis, then it vanishes on all vectors.

In terms of the matrix representation of the transformation,

$$\begin{aligned} (\hat{T}(\mathbf{x_i}), \hat{T}(\mathbf{x_j})) &= \sum_{i'=1}^{n} \sum_{j'=1}^{n} (\mathbf{x_{i'}} T_{i'i}, \mathbf{x_{j'}} T_{j'j}) \\ &= \sum_{i'=1}^{n} \sum_{j'=1}^{n} T_{i'i}^* T_{j'j} (\mathbf{x_{i'}}, \mathbf{x_{j'}}) \end{aligned} \tag{10.3}$$

and so to be unitary, the matrix representation of the operator $\hat{T}$ must satisfy:

$$\sum_{i'=1}^{n} \sum_{j'=1}^{n} T_{i'i}^* T_{j'j} O_{i'j'} = O_{ij} \tag{10.4}$$

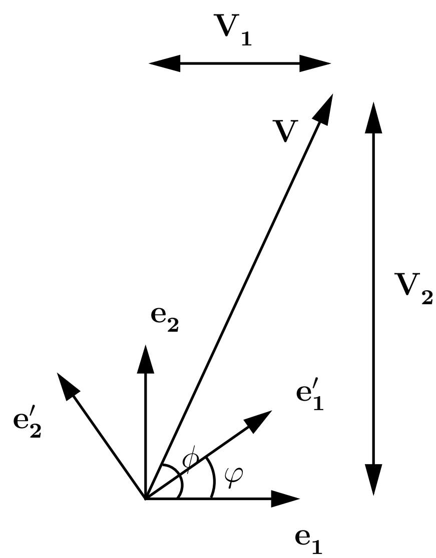
and in matrix notation, $T^{*T}OT = T^{\dagger}OT = O$.

For an orthonormal basis $O = I$ and $T^{\dagger}T = I$, the usual matrix definition of unitary, viz $T^{\dagger} = T^{-1}$.

Chapter 11

Rotations

We have a fixed vector $\mathbf{v}$ in a two diemnsional space with basis $\hat{\mathbf{e}}_\mathbf{1}$ and $\hat{\mathbf{e}}_\mathbf{2}$, $\mathbf{v} = v_1 \hat{\mathbf{e}}_\mathbf{1} + v_2 \hat{\mathbf{e}}_\mathbf{2}$ components (v_1, v_2).



Let us consider a new basis $\hat{\mathbf{e}}_\mathbf{1}{}'$, $\hat{\mathbf{e}}_\mathbf{2}{}'$, at angle ϕ with respect to $\hat{\mathbf{e}}_\mathbf{1}$, $\hat{\mathbf{e}}_\mathbf{2}$.

$$\begin{aligned} \hat{\mathbf{e}}_\mathbf{1}{}' &= \hat{\mathbf{e}}_\mathbf{1} \cos\phi + \hat{\mathbf{e}}_\mathbf{2} \sin\phi \\ \hat{\mathbf{e}}_\mathbf{2}{}' &= \hat{\mathbf{e}}_\mathbf{1}(-\sin\phi) + \hat{\mathbf{e}}_\mathbf{2} \cos\phi \end{aligned} \tag{11.1}$$

$$T_{ij} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \qquad T_{ij}^{-1} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{11.2}$$

The components of $\mathbf{v}$ with respect to the new basis are obtained from the components with respect to the original basis via

$$\begin{bmatrix} v_1' \\ v_2' \end{bmatrix} = T^{-1} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} \cos\phi v_1 + \sin\phi v_2 \\ -\sin\phi v_1 + \cos\phi v_2 \end{bmatrix} \tag{11.3}$$

If we observe that the vector $\mathbf{v}$ can be written

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{11.4}$$

(where v is the length of $\mathbf{v}$) then

$$\begin{bmatrix} v_1' \\ v_2' \end{bmatrix} = v \begin{bmatrix} \cos\phi\cos\theta + \sin\phi\sin\theta \\ -\sin\phi\cos\theta + \cos\phi\sin\theta \end{bmatrix} = v \begin{bmatrix} \cos(\theta-\phi) \\ \sin(\theta-\phi) \end{bmatrix} \tag{11.5}$$

as we would immediately deduce from geometric considerations.

In fact, this is the only type of unitary transformation in a real two dimensional vector space.

$$T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad T^\dagger = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \tag{11.6}$$

assume T is unitary, then

$$TT^\dagger = \begin{bmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{11.7}$$

so $a^2+b^2 = 1$, and $a = \cos\theta$, $b = \sin\theta$ for some θ, also $c^2+d^2 = 1$, and $c = \sin\phi$, $d = \cos\phi$ for some ϕ. Finally, $ac+bd = 0 = \cos\theta\sin\phi + \sin\theta\cos\phi = \sin(\theta+\phi)$, implying $\theta = -\phi$. Then

$$T = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \qquad \text{or} \qquad T = \begin{bmatrix} -\cos\phi & \sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \tag{11.8}$$

which is the previous rotation, or this rotation with reflection in the $y-$axis:

$$\begin{bmatrix} -\cos\phi & \sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

11.1 Example:

Find a rotation matrix that maps a Cartesian basis, $\{\mathbf{e_i}\}$, onto a new basis defined by; one vector, $\mathbf{n_1} = \frac{1}{3}(2\mathbf{e_1} + 2\mathbf{e_2} + \mathbf{e_3})$, and a second vector in the plane spanned by $\mathbf{n_1}$ and $\mathbf{e_3}$.

The first vector is:

$$\mathbf{n_1} = \frac{1}{3}(2\mathbf{e_1} + 2\mathbf{e_2} + \mathbf{e_3})$$

and the second vector is:

$$\mathbf{n_2} = \frac{\alpha}{3}(2\mathbf{e_1} + 2\mathbf{e_2} + \mathbf{e_3}) + \beta\mathbf{e_3}$$

subject to

$$0 = \mathbf{n_1} \cdot \mathbf{n_2} = \alpha + \frac{\beta}{3} \qquad 1 = \mathbf{n_2} \cdot \mathbf{n_2} = \alpha^2 + \frac{2}{3}\alpha\beta + \beta^2$$

and so:

$$1 = \alpha^2 + 2\frac{\beta}{3}\alpha + (3\alpha)^2 = 8\alpha^2$$

and so

$$\alpha = \frac{1}{2\sqrt{2}} \qquad \beta = -\frac{3}{2\sqrt{2}}$$

and

$$\mathbf{n_2} = \frac{1}{3\sqrt{2}}(\mathbf{e_1} + \mathbf{e_2} - 4\mathbf{e_3}).$$

The final vector is perpendicular to the other two:

$$
\begin{aligned}
\mathbf{n_3} &= \mathbf{n_1} \times \mathbf{n_2} \\
&= \frac{1}{9\sqrt{2}}(2\mathbf{e_1} + 2\mathbf{e_2} + \mathbf{e_3}) \times (\mathbf{e_1} + \mathbf{e_2} - 4\mathbf{e_3}) \\
&= \frac{1}{9\sqrt{2}}[(2\mathbf{e_1} + 2\mathbf{e_2}) \times (\mathbf{e_1} + \mathbf{e_2}) - 8(\mathbf{e_1} + \mathbf{e_2}) \times \mathbf{e_3} + \mathbf{e_3} \times (\mathbf{e_1} + \mathbf{e_2}) + \mathbf{e_3} \times \mathbf{e_3}] \\
&= \frac{1}{\sqrt{2}}\mathbf{e_3} \times (\mathbf{e_1} + \mathbf{e_2}) \\
&= \frac{1}{\sqrt{2}}(\mathbf{e_2} - \mathbf{e_1}).
\end{aligned}
$$

The rotation is defined by

$$\mathbf{n_i} = \sum_{j=1}^{3} \mathbf{e_j} R_{ji}$$

and hence

$$R_{ji} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{2}{3} & \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{3} & -\frac{4}{3\sqrt{2}} & 0 \end{pmatrix}.$$

Chapter 12

Special Relativity

In the current vector space language, special relativity is about four-dimensional vector space, (ct, x_1, x_2, x_3), with an inner product which has overlap matrix

$$(\hat{\mathbf{e}}_i, \hat{\mathbf{e}}_j) = O_{ij} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \tag{12.1}$$

and

$$c^2\tau^2 = (\mathbf{v}, \mathbf{v}) = c^2t^2 - x_1^2 - x_2^2 - x_3^2 \tag{12.2}$$

is the analogue of the 'norm'.

The transformations which preserve this inner product are known as Lorentz transformations and their matrices satisfy

$$L^\dagger O L = O. \tag{12.3}$$

One such example is

$$L = L^\dagger = \begin{bmatrix} \gamma & \frac{v}{c}\gamma & 0 & 0 \\ \frac{v}{c}\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12.4}$$

calculating $L^\dagger O L$,

$$
\begin{aligned}
L^\dagger O L &= (L^\dagger O)L \\
&= \begin{bmatrix} \gamma & -\frac{v}{c}\gamma & 0 & 0 \\ \frac{v}{c} & -\gamma & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \gamma & \frac{v}{c}\gamma & 0 & 0 \\ \frac{v}{c}\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \gamma^2(1-\frac{v^2}{c^2}) & 0 & 0 & 0 \\ 0 & -\gamma^2(1-\frac{v^2}{c^2}) & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}
\end{aligned} \tag{12.5}
$$

and so provided $\gamma^2 = \left(1 - \frac{v^2}{c^2}\right)^{-1}$ we have a Lorentz transformation.

Lorentz transformations preserve the inner product. If we consider a trajectory starting at $\mathbf{x} = \mathbf{0}$ (both space and time), and travelling to $(ct, \mathbf{r})$ at uniform velocity, then this is described by $(cs, \mathbf{r}s/t)$ where $s \in (0, t)$ parameterises the trajector, and also described by $(cs, \mathbf{v}s)$ in terms of the velocity $\mathbf{v}$. The inner product (for $s = t$) is given by

$$
c^2\tau^2 = c^2t^2(1 - \frac{\mathbf{v}\cdot\mathbf{v}}{c^2}). \tag{12.6}
$$

In the special frame where $\mathbf{v} = \mathbf{0}$, viz $\mathbf{r} = \mathbf{0}$, then $\tau = t$, and in the other frames, $\tau = t\sqrt{1 - \frac{v^2}{c^2}}$ and so t is correspondingly longer, i.e. we have time dilation. As $v \to c$, $\tau \to 0$ and no proper time passes. This is related to how the speed of light appears the same in all reference frames.

An observer with his clock ticking is described by

$$
\begin{bmatrix} c\tau \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{12.7}
$$

and his time is proper time. In a boosted frame he is described by

$$
\begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \gamma & \frac{v}{c}\gamma & 0 & 0 \\ \frac{v}{c}\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\tau \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma c\tau \\ \gamma v\tau \\ 0 \\ 0 \end{bmatrix} \tag{12.8}
$$

and an observer in this new frame can calculate the proper time, $c^2\tau^2 = c^2t^2 - x^2$ which is invariant, or can compare the two times:

$$t = \gamma\tau \qquad \Rightarrow \qquad \tau = \left[1 - \frac{v^2}{c^2}\right]^{\frac{1}{2}} t \tag{12.9}$$

and notice time dilation.

12.1 Ordinary Rotations as a Subset of Lorentz Transformations

For example a rotation in the $x - y$ plane preserves the inner product

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi & 0 \\ 0 & \sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad L^{\dagger} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & \sin\phi & 0 \\ 0 & -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12.10}$$

calculating $L^{\dagger}OL$

$$\begin{aligned} L^{\dagger}OL &= (L^{\dagger}O)L \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -\cos\phi & -\sin\phi & 0 \\ 0 & \sin\phi & -\cos\phi & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi & 0 \\ 0 & \sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -\cos^2\phi - \sin^2\phi & 0 & 0 \\ 0 & 0 & -\ sin^2\phi - \cos^2\phi & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \end{aligned} \tag{12.11}$$

Ordinary rotations introduce no new physics.

12.2 Full Lorentz Transformation

A full Lorentz transformation can be decomposed into an ordinary spatial rotation, followed by a boost, followed by a further ordinary rotation. New physics comes from the boost and so without loss of generality we my restrict attention to a special Lorentz transformation.

Chapter 13

Tensors

Tensors are representations of objects and operators in terms of arrays of 'componenets' for a given basis, combined with a law of transformation which yields the components in another basis.

Tensors can be one-dimensional, a line of numbers (Rank 1) a vector;

two-dimensional, a square of numbers (Rank 2) a matrix;

three-dimensional, a cube of numbers (Rank 3);

and so on.

13.0.1 Rank 1

The components of a vector are a rank 1 tensor, under a transformation

$$\mathbf{T_i} = \hat{\mathbf{T}}(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_j} T_{ji} \tag{13.1}$$

A vector $\mathbf{x}$ can be represented into different basis:

$$\mathbf{x} = \sum_{i=1}^{n} \lambda_i \mathbf{x_i} = \sum_{i=1}^{n} \lambda_i^{(T)} \mathbf{T_i}$$

using the transformation

$$\sum_{i=1}^{n} \lambda_i \mathbf{x_i} = \sum_{j=1}^{n} \lambda_j^{(T)} \sum_{i=1}^{n} \mathbf{x_i} T_{ij}$$

$$\sum_{i=1}^{n}\left[\lambda_i - \sum_{j=1}^{n} T_{ij}\lambda_j^{(T)}\right]\mathbf{x_i} = 0$$

As the $\mathbf{x_i}$ are linearly independent this implies $\lambda_j - \sum_{k=1}^{n} T_{jk}\lambda_k^{(T)} = 0$ so that

$$\lambda_i^{(T)} = \sum_{j=1}^{n} T_{ij}^{-1}\lambda_j \tag{13.2}$$

which is the law of transformation. In matrix notation, $\lambda^{(T)} = T^{-1}\lambda$.

13.0.2 Rank 2

The matrix representation of a linear operator is a rank 2 tensor,

$$\mathbf{L_i} = \hat{\mathbf{L}}(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_j} L_{ji}$$

and so

$$\begin{aligned}
\hat{\mathbf{L}}(\mathbf{T_i}) &= \hat{\mathbf{L}}\left(\sum_{j=1}^{n} T_{ji}\mathbf{x_i}\right) \\
&= \sum_{j=1}^{n} T_{ji}\hat{\mathbf{L}}(\mathbf{x_i}) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} \mathbf{x_k} L_{kj} T_{ji} \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \mathbf{T_l} T_{lk}^{-1} L_{kj} T_{ji}
\end{aligned}$$

and so

$$L_{il}^{(T)} = \sum_{j=1}^{n}\sum_{k=1}^{n} T_{ij}^{-1} L_{jk} T_{kl}. \tag{13.3}$$

in matrix notation, $L^{(T)} = T^{-1}LT$, which is the law of transformation. For unitary transformations

$$L_{il}^{(T)} = \sum_{j=1}^{n} \sum_{k=1}^{n} L_{jk} T_{ji}^{*} T_{kl}. \tag{13.4}$$

13.0.3 Rank 3

Linear mappings of many types all correspond to tensors. E.g. $\hat{\mathbf{M}}(\mathbf{x}, \mathbf{y})$ which maps a pair of vectors onto a third vector. For a chosen basis,

$$\hat{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{n} \mathbf{x}_k M_{kij} \tag{13.5}$$

and so

$$\begin{aligned}
\hat{\mathbf{M}}(\mathbf{T}_i, \mathbf{T}_j) &= \hat{\mathbf{M}} \left(\sum_{i_1=1}^{n} \mathbf{x}_{i_1} T_{i_1 i} \, , \, \sum_{j_1=1}^{n} \mathbf{x}_{j_1} T_{j_1 j} \right) \\
&= \sum_{i_1=1}^{n} \sum_{j_1=1}^{n} \hat{\mathbf{M}}(\mathbf{x}_{i_1}, \mathbf{x}_{j_1}) T_{i_1 i} T_{j_1 j} \\
&= \sum_{i_1=1}^{n} \sum_{j_1=1}^{n} \sum_{k=1}^{n} \mathbf{x}_{k_1} M_{k_1 i_1 j_1} T_{i_1 i} T_{j_1 j} \\
&= \sum_{i_1=1}^{n} \sum_{j_1=1}^{n} \sum_{k_1=1}^{n} \sum_{k=1}^{n} \mathbf{T}_k \left[T^{-1} \right]_{k k_1} M_{k_1 i_1 j_1} T_{i_1 i} T_{j_1 j}
\end{aligned}$$

and so

$$M_{kij}^{(T)} = \sum_{i_1=1}^{n} \sum_{j_1=1}^{n} \sum_{k_1=1}^{n} M_{i_1 j_1 k_1} \left[T^{-1} \right]_{i i_1} T_{j_1 j} T_{k_1 k} \tag{13.6}$$

and for a unitary transformation

$$M_{kij}^{(T)} = \sum_{i_1=1}^{n} \sum_{j_1=1}^{n} \sum_{k_1=1}^{n} M_{i_1 j_1 k_1} T_{i i_1}^{*} T_{j_1 j} T_{k_1 k}. \tag{13.7}$$

13.1 Totally Anti-Symmetric Mapping

A very important example is the totally anti-symmetric mapping, $A(\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n})$, which maps n vectors onto a scalar. Additionally,

$$A(\ldots, \mathbf{v_i}, \ldots, \mathbf{v_j}, \ldots) = -A(\ldots, \mathbf{v_j}, \ldots, \mathbf{v_i}, \ldots) \tag{13.8}$$

for all values of i, j, and $A(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}) = 1$ for an orthonormal basis. We write for this scalar $A(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_n}) = A_{i_1, i_2, \ldots, i_n}$. This yields an essentially unique function, and

$$A(\mathbf{T}_{i_1}, \mathbf{T}_{i_2}, \ldots, \mathbf{T}_{i_n}) = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_n=1}^{n} = A(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \ldots, \mathbf{x}_{i_n}) T_{j_1 i_1} T_{j_2 i_2} \cdots T_{j_n i_n} \tag{13.9}$$

so

$$A^{(T)}_{i_1, i_2, \ldots, i_n} = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_n=1}^{n} = A_{i_1, i_2, \ldots, i_n} T_{j_1 i_1} T_{j_2 i_2} \cdots T_{j_n i_n}. \tag{13.10}$$

Chapter 14

Choosing a Basis to Make an Operator Look Simple

This is the first major use for the 'mechanism' being set up. We have a linear operator, $\mathbf{L}$ say, and now we want to choose a linear transformation, $\mathbf{T}$ say, to 'simplify' the operator.

14.1 Simplest Example: An Operator which Maps a Vector Down to a Scalar

The simplest example is that of a linear operator which maps a vector down to a scalar, $L(\mathbf{x})$. For a given basis this operator is controlled by a rank 1 tensor, $L_i = L(\mathbf{x_i})$ and in this basis:

$$L(\mathbf{x}) = L(\sum_{i=1}^{n} \lambda_i \mathbf{x_i}) = \sum_{i=1}^{n} \lambda_i L(\mathbf{x_i}) = \sum_{i=1}^{n} \lambda_i L_i.$$

If we transform the basis using $\mathbf{T}(\mathbf{x})$, then with $\mathbf{T_i} = \mathbf{T}(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_i} T_{ji}$ we have

$$L(\mathbf{x}) = \sum_{j=1}^{n} \lambda_j^{(T)} L(\mathbf{T_i}) = \sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_j^{(T)} L_i T_{ij}.$$

and so as we already know,

$$\lambda_i^{(T)} = \sum_{j=1}^{n} (T^{-1})_{ij} \lambda_j.$$

To make $L(\mathbf{x})$ a simple object, we must make, $L_j^{(T)} = \sum_{i=1}^n L_i T_{ij}$ simple. This is an elementary inner-product, and so we can choose

$$T_{i0} = \frac{L_i^*}{\|L\|} \tag{14.1}$$

and choose T_{in} to be orthonormal (by Gram-Schmidt) with respect to T_{i0} and each other, i.e.

$$\sum_{i=1}^{n} T_{in}^* T_{in'} = I_{nn'} \equiv \delta_{nn'} \tag{14.2}$$

including

$$\sum_{i=1}^{n} T_{i0}^* T_{i0} = \sum_{i=1}^{n} \frac{L_i L_i^*}{\|L\|^2} = 1 \tag{14.3}$$

and

$$\sum_{i=1}^{n} T_{i0}^* T_{in} = \sum_{i=1}^{n} \frac{L_i T_{in}}{\|L\|^2} = 0 \tag{14.4}$$

if $n \neq 1$. For this particular choice $L_j^T = \sum_{i=1}^n L_i T_{ij} = 0$ if $j \neq 1$ and $\sum_{i=1}^n L_i L_i^* / \|L\| = \|L\|$ if $j = 1$.

We have chosen a basis for which the first vector is mapped to a non-zero scalar and all others are mapped tozero:

$$L(\mathbf{x}) = \lambda_1^{(T)} \|L\|. \tag{14.5}$$

14.1.1 Example:

There are two charges of charge Q at positions: $(0, 0, a)$ and $(0, 0, -a)$. Choose a basis to make the electric field at the point $(2a, 0, a)$ simple.

The electric potential is:

$$V(x, y, z) = \frac{Q}{[x^2 + y^2 + (z - a)^2]^{\frac{1}{2}}} + \frac{Q}{[x^2 + y^2 + (z + a)^2]^{\frac{1}{2}}} \tag{14.6}$$

and consequently the electric field is

$$\mathbf{E}(x,y,z) = -\nabla V(x,y,z) = \frac{Q(x,y,z-a)}{[x^2+y^2+(z-a)^2]^{\frac{3}{2}}} + \frac{Q(x,y,z+a)}{[x^2+y^2+(z+a)^2]^{\frac{3}{2}}} \tag{14.7}$$

leading to:

$$\begin{aligned}
\mathbf{E}(2a,0,a) &= \frac{Q(2a,0,a-a)}{[(2a)^2+(a-a)^2]^{\frac{3}{2}}} + \frac{Q(2a,0,a+a)}{[(2a)^2+(a+a)^2]^{\frac{3}{2}}} \\
&= Q\left(\frac{1}{4a^2},0,0\right) + Q\left(\frac{1}{8\sqrt{2}a^2},0,\frac{1}{8\sqrt{2}a^2}\right) \\
&= \frac{Q}{8\sqrt{2}a^2}(2\sqrt{2}+1,0,1)
\end{aligned} \tag{14.8}$$

Obviously, to make this vector easy to describe we should choose one vector parallel to its direction:

$$\mathbf{n}_1 = \frac{1}{\sqrt{[10+4\sqrt{2}]}}(2\sqrt{2}+1,0,1).$$

and the other two to be perpendicular

$$\mathbf{n}_2 = \frac{1}{\sqrt{[10+4\sqrt{2}]}}(1,0,-2\sqrt{2}-1) \qquad \mathbf{n}_3 = (0,1,0)$$

and in this new basis:

$$\mathbf{E}(2a,0,a) = \frac{Q}{8\sqrt{2}a^2}\sqrt{[10+4\sqrt{2}]}\,\mathbf{n}_1.$$

14.2 Example: A Linear Mapping of Two Vectors Down to a Scalar

A second very simple case is a linear mapping of two vectors down to a scalar. For a given basis this operator is controlled by a rank 2 tensor, $L_{ii'} = L(\mathbf{x}_i, \mathbf{x}_{i'})$. In this basis

$$L(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{n} \sum_{i'=1}^{n} \lambda_i \lambda'_{i'} L(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{i=1}^{n} \sum_{i'=1}^{n} \lambda_i \lambda'_{i'} L_{ii'}$$

If we transform with $\mathbf{T}(\mathbf{x})$, then

$$L(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{n} \sum_{i'=1}^{n} \lambda_i^{(T)} \lambda_{i'}^{'(T)} L(\mathbf{T_i}, \mathbf{T_{i'}}) = \sum_{i=1}^{n} \sum_{i'=1}^{n} \sum_{j=1}^{n} \sum_{j'=1}^{n} \lambda_i^{(T)} \lambda_{i'}^{'(T)} L_{jj'} T_{ji} T_{i'j'}$$

and so

$$L_{ii'}^{(T)} = \sum_{j=1}^{n} \sum_{j'=1}^{n} L_{jj'} T_{ji} T_{i'j'}$$

and as we know $\lambda_i^{(T)} = \sum_{j=1}^{n} (T^{-1})_{ij} \lambda_j$.

In matrix language $L^{(T)} = T^T L T$. For a real system this quite natural, whereas for a complex system this is not.

If we consider the case of a *mixed* linear, anti-linear operator, $L(\lambda \mathbf{x}, \mathbf{x}') = \lambda^* L(\mathbf{x}, \mathbf{x}')$ and $L(\mathbf{x}, \lambda \mathbf{x}') = \lambda L(\mathbf{x}, \mathbf{x}')$, like an inner product, then

$$L_{ii'}^{(T)} = \sum_{j=1}^{n} \sum_{j'=1}^{n} L_{jj'} T_{ji}^* T_{i'j'}$$

and in the language, $L^{(T)} = T^\dagger L T$, which is quite natural.

The simplest that $L^{(T)}$ can become is *diagonal* if $L(\mathbf{x}, \mathbf{x}')$ is *hermitian*. This is the second major mathematical task encountered that you have previously met; *diagonal* a matrix.

If we let the columns ofT_{ij} be the eigenvectors of L_{ij}, i.e. $T_{ij} = u_i^{(j)}$ with $\sum_{j=1}^{n} L_{ij} u_j^{(n)} = l_n u_i^{(n)}$, then

$$L_{ii'}^{(T)} = \sum_{j=1}^{n} \sum_{j'=1}^{n} u_j^{(i)*} L_{jj'} u_{j'}^{(i')} = \sum_{j=1}^{n} u_j^{(i)*} u_j^{(i')} l_{i'} = I_{ii'} l_i$$

and

$$L(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{n} l_i \lambda_i^{(T)*} \lambda_i^{'(T)}.$$

Example:

A curve is defined by:

$$8x^2 - 12xy + 17y^2 = 20$$

where (x, y) are components in a Cartesian basis. Classify/Describe this curve.

A curve such as this is represented mathematically in terms of a linear mapping:

$$L(\mathbf{x}, \mathbf{x}') \qquad \text{implies} \qquad L(\mathbf{x}, \mathbf{x}) = \mathit{constant}$$

of two vectors onto a scalar.

$$L(\mathbf{x}, \mathbf{x}') = \sum_{ij} \lambda_i L(\mathbf{x_i}, \mathbf{x_j}) \lambda_j = [x, y] \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^2 L_{11} + xy(L_{12} + L_{21}) + y^2 L_{22}.$$

In our Cartesian basis:

$$L_{ij} = L(\mathbf{x_i}, \mathbf{x_j}) = \begin{pmatrix} \frac{8}{20} & -\frac{6}{20} \\ -\frac{6}{20} & \frac{17}{20} \end{pmatrix} \qquad \text{and} \qquad L(\mathbf{x}, \mathbf{x}) = 1.$$

We can simplify this operator by choosing a basis in which this matrix is diagonal:

$$\det \begin{pmatrix} \frac{8}{20} - \epsilon & -\frac{6}{20} \\ -\frac{6}{20} & \frac{17}{20} - \epsilon \end{pmatrix} = \epsilon^2 - \frac{5}{4}\epsilon \frac{1}{4} = \left(\epsilon - \frac{1}{4}\right)(\epsilon - 1)$$

together with eigenvectors:

$$\begin{pmatrix} \frac{8}{20} & -\frac{6}{20} \\ -\frac{6}{20} & \frac{17}{20} \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \qquad \begin{pmatrix} \frac{8}{20} & -\frac{6}{20} \\ -\frac{6}{20} & \frac{17}{20} \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

The transformation is effected with:

$$T = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \qquad T^{-1} = T^{\dagger} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$$

and:

$$L^{(T)} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$$

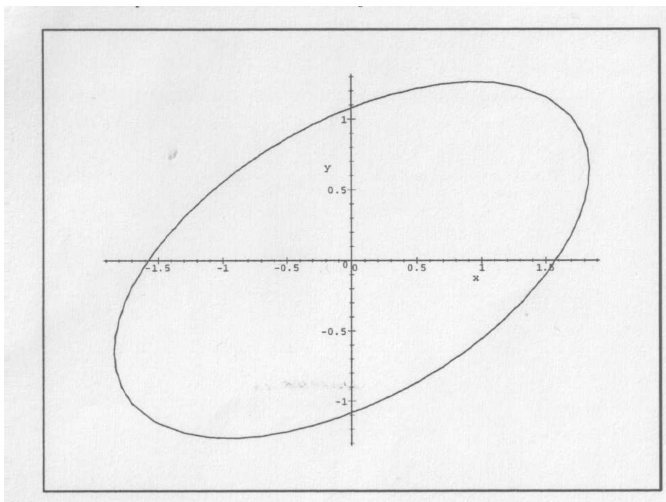
and consequently:

$$\left[x_1^{(T)}\right]^2 + \frac{1}{4}\left[x_2^{(T)}\right]^2 = 1$$

with:

$$\mathbf{T_1} = \frac{1}{\sqrt{5}}[\mathbf{x_1} - 2\mathbf{x_2}] \qquad \mathbf{T_2} = \frac{1}{\sqrt{5}}[2\mathbf{x_1} + \mathbf{x_2}]$$

We find an *ellipse* with minor axis 1 and major axis 2.



Example:

A surface is defined:

$$x^2 + y^2 + z^2 + yz + zx + xy = 2$$

where (x, y, z) are the components in a Cartesian basis. Describe this surface.

Once again we are dealing with a mapping of two vectors onto a scalar. The matrix representation is:

$$L_{ij} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

and then:

$$T_{ij} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \end{pmatrix} \qquad L^{(T)}_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$$

which leads to:

$$\left[x^{(T)}_1\right]^2 + \frac{\left[x^{(T)}_2\right]^2 + \left[x^{(T)}_3\right]^2}{4} = 1$$

and an ellipsoid with a minor axis of 1 paraellel to:

$$\mathbf{T_1} = \frac{1}{\sqrt{3}}(\mathbf{x_1} + \mathbf{x_2} + \mathbf{x_3})$$

and two equal major axes of 2 in the other two perpendicular directions.

The most important example for us is linear transformation mapping the vector space onto itself, $\mathbf{L}(\mathbf{x})$. For a given basis this operator is controlled by rank 2 tensor, L_{ij} with $\mathbf{L}(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_j} L_{ji}$ and in this basis

$$\mathbf{L}(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i \mathbf{L}(\mathbf{x_i}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \mathbf{x_j} L_{ji}.$$

If we transform the basis using $\mathbf{T}(\mathbf{x})$, then

$$\begin{aligned}
\mathbf{L}(\mathbf{x}) &= \sum_{i=1}^{n} \lambda_i^{(T)} \mathbf{L}(\mathbf{T_i}) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i^{(T)} \mathbf{L}(\mathbf{x_j}) T_{ji} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_i^{(T)} \mathbf{x_k} L_{kj} T_{ji} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \mathbf{T_l} (T^{-1})_{lk} L_{kj} T_{ji} \lambda_i^{(T)}
\end{aligned}$$

and so

$$L_{ij}^{(T)} = \sum_{i'=1}^{n} \sum_{j'=1}^{n} (T^{-1})_{ii'} L_{i'j'} T_{j'j}$$

in matrix language, $L^{(T)} = T^{-1} L T$, identical to the previous case for unitary transformations. Once again the 'simplest' we can make $L^{(T)}$ is diagonal, using $T_{in} = u_i^{(n)}$ and $\sum_{j=1}^{n} L_{ij} u_j^{(n)} = l_n u_i^{(n)}$, then

$$L_{ij}^{(T)} = \sum_{i'=1}^{n} \sum_{j'=1}^{n} u_{i'}^{(i)*} L_{i'j'} u_{j'}^{(j)} = \sum_{i'=1}^{n} u_{i'}^{(i)*} u_{i'}^{(j)} l_j = I_{ij} l_i.$$

The new basis vectors $\mathbf{T_i} = \sum_{j=1}^{n} \mathbf{x_j} T_{ji} = \sum_{j=1}^{n} \mathbf{x_j} u_j^{(i)}$ satisfy

$$\mathbf{L}(\underset{\sim}{T}_i) = \sum_{j=1}^{n} \mathbf{T_j} L_{ji}^{(T)} = l_i \mathbf{T_i}$$

and are mapped onto scalar multiples of themselves.

Example:

A physical sample has its magnetic properties examined. A field is applied in each of three Cartesian directions and the induced magnetism is measured to be:

$$
\begin{aligned}
\mathbf{M}[B\mathbf{e_1}] &= B\alpha[3\mathbf{e_1} + \sqrt{2}(\mathbf{e_2} + \mathbf{e_3})] \\
\mathbf{M}[B\mathbf{e_2}] &= B\alpha[\sqrt{2}\mathbf{e_1} + 2\mathbf{e_2} + \mathbf{e_3}] \\
\mathbf{M}[B\mathbf{e_3}] &= B\alpha[\sqrt{2}\mathbf{e_1} + \mathbf{e_2} + 2\mathbf{e_3}]
\end{aligned}
$$

and is observed (and assumed) to be *linear* in the apllied field, B. Find tyhe natural basis in which to describe the induced magnetic moment.

We are dealing with a linear operator:

$$
\mathbf{M}[\mathbf{e_i}] = \sum_{j=1}^{n} \mathbf{e_j} M_{ji}
$$

with:

$$
M_{ji} = \alpha \begin{pmatrix} 3 & \sqrt{2} & \sqrt{2} \\ \sqrt{2} & 2 & 1 \\ \sqrt{2} & 1 & 2 \end{pmatrix}
$$

and thisis diagonalised by:

$$
T_{ji} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{\sqrt{2}} \end{pmatrix} \quad \text{implies} \quad M_{ji}^{(T)} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}
$$

and we find times the moment induced in the direction:

$$
\mathbf{T_1} = \frac{1}{\sqrt{2}}\mathbf{e_1} + \frac{1}{2}(\mathbf{e_2} + \mathbf{e_3})
$$

in comparison to the other two perpendicular directions.

Chapter 15

Vector Spaces: Summary

The 'conceptual' vector-space section is now at an end. There are only a few ideas to remember:

V the space itself, with vectors $\mathbf{x}$, scale factors λ, and rules to test.

$(\mathbf{x}, \mathbf{y})$ the inner product, corresponding to the measurements.

Operators acting on the space:

$L(\mathbf{x})$ a vector down to a scalar: $V \mapsto S$
$L(\mathbf{x}, \mathbf{y})$ two vectors down to a scalar: $(V, V) \mapsto S$
$\mathbf{L}(\mathbf{x})$ a vector down to a vector: $V \mapsto V$

and beyond...

$(\mathbf{x}, \hat{H}\mathbf{y}) = (\hat{H}\mathbf{x}, \mathbf{y})$: Hermitian
$(\hat{T}\mathbf{x}, \hat{T}\mathbf{y}) = (\mathbf{x}, \mathbf{y})$: Unitary

When we describe a vector space, we usually use a basis. In terms of a basis, all of the above objects are represented by arrays of numbers known as tensors. We often consider changes in basis, which are called transformations. All tensors have a transformation law, which for unitary transformation law, which for a unitary transformation has the form:

$$O^{(T)}_{i_1 i_2 \ldots i_n} = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \cdots \sum_{j_n=1}^{n} O_{j_1 j_2 \ldots j_n} T^{x_1}_{j_1 i_1} T^{x_2}_{j_2 i_2} \ldots T^{x_n}_{j_n i_n} \tag{15.1}$$

where x_i denote either complex-conjugate or not.

Using

$$\mathbf{T_i} = \sum_{j=1}^{n} \mathbf{x_j} T_{ji} \tag{15.2}$$

to denote the change of basis, and:

$$\mathbf{L}(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_j} L_{ji} \tag{15.3}$$

for example, to represent an operator, you should be able to derive such a law.

There are three basic techniques:

(1) Solving Simultaneous Linear Equations: Where am I?
(2) Gram-Schmidt Orthogonalisation: Find me a basis
(3) Diagonalisation of a matrix: Choose a basis to make an operator simple

You should appreciate that $V \mapsto S$ involves a single direction mapping down to a scalar, all others mapping to zero, and $V \mapsto V$ involves independent directions which map onto themselves but with a scale factor (the eigenvalue).

Chapter 16

The Binomial Theorem and Leibnitz's Theorem

16.1 The Binomial Theorem

$$(x+y)^n = \sum_{r=0}^{n} \frac{n!}{(n-r)!r!} x^{n-r} y^r \tag{16.1}$$

If desired we can use proof by induction: Assume (16.1) then

$$\begin{aligned}
& (x+y)^{n+1} \\
= \; & (x+y) \sum_{r=0}^{n} \frac{n!}{(n-r)!r!} x^{n-r} y^r \\
= \; & (x+y)[C_0^n x^n + C_1^n x^{n-1} y + C_2^n x^{n-2} y^2 + \cdots + C_r^n x^{n-r} y^r + C_{r+1}^n x^{n-r-1} y^{r+1} + \ldots C_n^n y^n] \\
= \; & C_0^n x^{n+1} + (C_0^n + C_1^n) x^n y + (C_1^n + C_2^n) x^{n-1} y^2 + \cdots + (C_r^n + C_{r+1}^n) x^{n+1-r-1} y^{r+1} \\
& + \cdots + (C_{n-1}^n + C_n^n) x y^n + C_n^n y^{n+1}
\end{aligned} \tag{16.2}$$

we calculate

$$\begin{aligned}
C_r^n + C_{r+1}^n &= \frac{n!}{(n-r)!r!} + \frac{n!}{(n-r-1)!(r-1)!} \\
&= \frac{n!}{(n-r)!(r+1)!} \{(r+1) + (n-r)\} \\
&= \frac{(n+1)!}{(n+1-r-1)!(r+1)!} \\
&= C_{r+1}^{n+1}
\end{aligned} \tag{16.3}$$

therefore we have

$$\begin{aligned}(x+y)^{n+1} &= C_0^n x^{n+1} + \sum_{r=0}^{n-1} C_{r+1}^{n+1} x^{n-r} y^r + C_n^n y^{n+1} \\ &= \sum_{r=0}^{n+1} \frac{(n+1)!}{(n+1-r)!r!} x^{n+1-r} y^r \qquad (16.4)\end{aligned}$$

which is (16.1) with $n \mapsto n+1$. We then need only establish the result for $n=1$:

$$(x+y) = \frac{1!}{(1-0)!0!} x^1 y^0 + \frac{1!}{(1-1)!1!} x^0 y^1$$

where by convention $0! = 1$.

An alternative to (16.1) is

$$(x_1+x_2)^n = \sum_{r_1=0}^{n} \sum_{r_2=0}^{n} \frac{n!}{r_1!r_2!} x_1^{r_1} x_2^{r_2} \delta_{r_1+r_2,n} \qquad (16.5)$$

which is easily derived from

$$(x_1+x_2)^n = \sum_{r_1=0}^{n} \sum_{r_2=0}^{n} A(r_1,r_2)\, x_1^{r_1} x_2^{r_2} \delta_{r_1+r_2,n},$$

where $A(r_1, r_2)$ is a constant depending on r_1 and r_2, by applying

$$\frac{\partial^{r_1}}{\partial x_2^{r_1}} \frac{\partial^{r_2}}{\partial x_2^{r_2}}$$

for $r_1 + r_2 = n$ to both sides. We easily find

$$A(r_1, r_2) = \frac{n!}{r_1!r_2!}.$$

It is simply to show the generalisation

$$(x_1+x_2+\cdots+x_m)^n = \sum_{r_1=0}^{n} \sum_{r_2=0}^{n} \cdots \sum_{r_m=0}^{n} \frac{n!}{r_1!r_2!\ldots r_m!} x_1^{r_1} x_2^{r_2} \ldots x_m^{r_m} \delta_{r_1+r_2+\cdots+r_m,n} \qquad (16.6)$$

16.2 Leibnitz's Theorem

Recall the product rule

$$\frac{d}{dx}(uv) = \frac{du}{dx}\,v + u\frac{dv}{dx} \tag{16.7}$$

If we apply it twice

$$\begin{aligned}\frac{d^2}{dx^2}(uv) = \left(\frac{d}{dx}\right)^2 &= \frac{d}{dx}(\frac{du}{dx}\,v) + \frac{d}{dx}(u\frac{dv}{dx}) \\ &= \frac{d^2u}{dx^2}\,v + \frac{du}{dx}\,\frac{dv}{dx} + \frac{du}{dx}\,\frac{dv}{dx} + u\frac{d^2v}{dx^2} \\ &= \frac{d^2u}{dx^2}\,v + 2\frac{du}{dx}\,\frac{dv}{dx} + u\frac{d^2v}{dx^2} \end{aligned} \tag{16.8}$$

We use the notation

$$u^{(r)} = \frac{d^r v}{dx^r}, \quad v^{(r)} = \frac{d^r v}{dx^r}$$

the general result, known as Leibnitz's theorem, is

$$(uv)^{(n)} = \sum_{r=0}^{n} \frac{n!}{(n-r)!r!} u^{(n-r)}\,v^{(r)} \tag{16.9}$$

We can use proof by induction again. By the product rule the result istruefor $n = 1$. Now assume (16.9) and differentiate both sides

$$
\begin{aligned}
(uv)^{(n+1)} &= \frac{d}{dx}\sum_{r=0}^{n} C_r^n u^{(n-r)}\ v^{(r)} \\
&= \frac{d}{dx}(u^{(n)}v + C_1^n u^{(n-1)}v_1 + C_2^n u^{(n-2)}v^{(2)} + C_3^n u^{(n-3)}v^{(3)} + \cdots + uv^{(n)}) \\
&= (u^{(n+1)}v + u^{(n)}v^{(1)}) + C_1^n(u^{(n)}v^{(1)} + u^{(n-1)}v^{(2)}) + C_2^n(u^{(n-1)}v^{(2)} + u^{(n-2)}v^{(3)}) \\
&\quad + \cdots + uv^{(n+1)} \\
&= u^{(n+1)}v + (C_0^n + C_1^n)u^{(n)}v^{(1)} + (C_1^n + C_2^n)u^{(n-1)}v^{(2)} + (C_2^n + C_3^n)u^{(n-2)}v^{(3)} \\
&\quad + \cdots + uv^{(n+1)} \\
&= u^{(n+1)}v + \sum_{r=0}^{n-1}(C_r^n + C_{r+1}^n)u^{(n-r)}\ v^{(r+1)} + uv^{(n+1)} \\
&= \sum_{r=0}^{n+1} C_r^{n+1}u^{(n+1-r)}\ v^{(r)} \qquad (16.10)
\end{aligned}
$$

we see the obvious analogy with the proof of the Binomial Theorem.

16.3 Example: The Binomial Theorem

Given a bag with n numbered balls, what is the probability that you see each ball at least once in N random selections?

16.4 Example: Leibnitz's Theorem

Define the function

$$L_s(x) = e^x \left(\frac{d}{dx}\right)^s (x^s e^{-x}). \qquad (16.11)$$

We use Leibnitz to prove that it satisfies the following differential equation

$$x\frac{d^2L_s}{dx^2} + (1-x)\frac{dL_s}{dx} + sL_s = 0 \qquad (16.12)$$

By the product rule

$$\frac{d}{dx}(x^s e^{-x}) = sx^{s-1}e^{-x} - x^s e^{-x}$$

so

$$x\frac{d}{dx}(x^s e^{-x}) + (x-s)x^s e^{-x} = 0. \tag{16.13}$$

We have factors proportional to $x-C$ where C is a constant, in this case (i.e. $(x-C)f(x)$) Leibnitz reduces to

$$[f(x)(x-C)]^{(r)} = f^{(r)}(x)(x-C) + rf^{(r-1)}(x) + 0.$$

Applying $\left(\frac{d}{dx}\right)^{s+1}$ to (16.13) we obtain from Leibnitz

$$\begin{aligned}
&\left(\frac{d}{dx}\right)^{s+1}\left[x\frac{d}{dx}(x^s e^{-x}) + (x-s)x^s e^{-x}\right] \\
= \ & x\frac{d^2}{dx^2}\left(\frac{d}{dx}\right)^s (x^s e^{-x}) + (s+1)\frac{d}{dx}\left(\frac{d}{dx}\right)^s (x^s e^{-x}) \\
& + (x-s)\frac{d}{dx}\left(\frac{d}{dx}\right)^s (x^s e^{-x}) + (s+1)\left(\frac{d}{dx}\right)^s (x^s e^{-x}) \\
= \ & x\frac{d^2}{dx^2}[e^{-x}L_s] + (s+1)\frac{d}{dx}[e^{-x}L_s] + (x-s)\frac{d}{dx}[e^{-x}L_s] + (s+1)[e^{-x}L_s] = 0.
\end{aligned} \tag{16.14}$$

Using Leibnitz we have

$$x\frac{d^2}{dx^2}[e^{-x}L_s] = x\left(e^{-x}\frac{d^2}{dx^2}L_s - 2e^{-x}\frac{d}{dx}L_s + e^{-x}L_s\right)$$

and

$$\frac{d}{dx}[e^{-x}L_s] = e^{-x}\frac{d}{dx}L_s - e^{-x}L_s.$$

Substituting these into the last line of (16.14) and multiplying by e^x we obtain

$$x\left(\frac{d^2}{dx^2}L_s - 2\frac{d}{dx}L_s + L_s\right) + (x+1)\left(\frac{d}{dx}L_s - L_s\right) + (s+1)L_s = 0. \tag{16.15}$$

This easily simplifies to (16.12).

Chapter 17

Non-Linear Functions and Taylor's Theorem

Taylor's theorem in one-dimension is

$$f(x+h) = f(x) + hf^{(1)}(x) + \frac{h^2}{2!}f^{(2)}(x) + \cdots = \sum_{n=0}^{\infty} \frac{h^n}{n!} f^{(n)}(x) \tag{17.1}$$

where

$$f^{(n)}(x) = \left[\frac{d}{dx}\right]^n f(x) \tag{17.2}$$

are the derivatives of the function. What is the analogue in vector space?

Locally we will revert to real numbers for this discussion. If

$$\mathbf{x} = \lambda_1 \mathbf{x_1} + \lambda_2 \mathbf{x_2} + \cdots + \lambda_n \mathbf{x_n}$$

then a general function will map

$$F(\mathbf{x}) = f(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

where f is a completely geeneral function of coordinates. To obtain an analogue of Taylor's theorem, we should translate to the point of interest and consider this as the origin for our vector space. $\mathbf{x} = \mathbf{x}_0$, $\mathbf{x_0} = \lambda_1\mathbf{x_1} + \lambda_2\mathbf{x_2} + \cdots + \lambda_n\mathbf{x_n}$, and $\delta\mathbf{x} = h_1\mathbf{x_1} + h_2\mathbf{x_2} + \cdots + h_n\mathbf{x} - \mathbf{n}$ and then consider $\mathbf{h}$ as our vectors, or at least their components.

$f(\lambda_1+h_1, \lambda_2+h_2, \ldots, \lambda_n+h_n)$ is expanded using $n-$sets of independent one-dimensional Taylor's theorem

$$f(\lambda_1+h_1, \lambda_2+h_2, \ldots, \lambda_n+h_n) = \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_n=0}^{\infty} \frac{h_1^{i_1}}{i_1!}\left[\frac{\partial}{\partial\lambda_1}\right]^{i_1} \frac{h_2^{i_2}}{i_2!}\left[\frac{\partial}{\partial\lambda_2}\right]^{i_2}\cdots\frac{h_n^{i_n}}{i_n!}\left[\frac{\partial}{\partial\lambda_n}\right]^{i_n} f(\lambda) \tag{17.3}$$

From the binomial theorem:

$$\begin{aligned}
&\sum_{i=0}^{\infty}\frac{1}{i!}\left[h_1\frac{\partial}{\partial\lambda_1}+h_2\frac{\partial}{\partial\lambda_2}+\cdots+h_n\frac{\partial}{\partial\lambda_n}\right]^i f(\lambda)\\
= \;& \sum_{i=0}^{\infty}\frac{1}{i!}\sum_{i_1=0}^{i}\sum_{i_2=0}^{i}\cdots\sum_{i_n=0}^{i}\delta_{i_1+i_2+\cdots+i_n,i}\frac{i!}{i_1!i_2!\ldots i_n!}h_1^{i_1}\left[\frac{\partial}{\partial\lambda_1}\right]^{i_1} h_2^{i_2}\left[\frac{\partial}{\partial\lambda_2}\right]^{i_2}\cdots h_n^{i_n}\left[\frac{\partial}{\partial\lambda_n}\right]^{i_n} f(\lambda)\\
= \;& \sum_{i_1=0}^{\infty}\sum_{i_2=0}^{\infty}\cdots\sum_{i_n=0}^{\infty}\frac{h_1^{i_1}}{i_1!}\left[\frac{\partial}{\partial\lambda_1}\right]^{i_1}\frac{h_2^{i_2}}{i_2!}\left[\frac{\partial}{\partial\lambda_2}\right]^{i_2}\cdots\frac{h_n^{i_n}}{i_n!}\left[\frac{\partial}{\partial\lambda_n}\right]^{i_n} f(\lambda)
\end{aligned} \tag{17.4}$$

so we can write

$$f(\lambda_1+h_1, \lambda_2+h_2, \ldots, \lambda_n+h_n) = \sum_{i=0}^{\infty}\frac{1}{i!}\left[h_1\frac{\partial}{\partial\lambda_1}+h_2\frac{\partial}{\partial\lambda_2}+\cdots+h_n\frac{\partial}{\partial\lambda_n}\right]^i f(\lambda) \tag{17.5}$$

or more compactly

$$f(\lambda+\mathbf{h}) = \sum_{n=0}^{\infty}\frac{1}{n!}\left[\mathbf{h}\cdot\nabla\right]^n f(\lambda) \tag{17.6}$$

where the $\mathbf{h}$ are the components of the vector $\delta\mathbf{x}$ and ∇ is the gradient operator. Each term involves a tensor, which increases in rank term by term.

If we consider a transformation $\hat{\mathbf{T}}$, then $\mathbf{T}(\mathbf{x_i}) = \mathbf{T_i}$ and $\mathbf{T_i} = \sum_{j=1}^{n}\mathbf{x_i}T_{ji}$ so

$$\sum_{i=1}^{n} h_i^{(T)}\mathbf{T_i} = \sum_{i=1}^{n} h_i^{(T)}\sum_{j=1}^{n}\mathbf{x_i}T_{ji} = \sum_{j=1}^{n} h_j\mathbf{x_i} \tag{17.7}$$

and so $\sum_{i=1}^{n} h_i^{(T)}T_{ji} = h_j$ so that we have the transformation law

$$h_i^{(T)} = \sum_{j=1}^{n} (T^{-1})_{ij} h_j \tag{17.8}$$

And ∇ transforms as

$$\frac{\partial}{\partial h_j^{(T)}} = \sum_{i=1}^{n} \frac{\partial h_i}{\partial h_j^{(T)}} \frac{\partial}{\partial h_j} = T_{ij} \frac{\partial}{\partial h_i}$$

and so

$$\nabla_i^{(T)} = \sum_{j=1}^{n} \nabla_j \; T_{ji}. \tag{17.9}$$

Chapter 18

Optimising Functions and Plotting Graphs

Plotting functions in one-dimension is easy. The slope or gradient tells you which way the curve is going (up or down) and and 'turning points' tell you change in direction,viz maxima or minima. What is the analogue in vector space?

We can use Tayor's theorem to find the turning points just as we did in one-dimension:

$$f(\mathbf{x}+\mathbf{h}) = \sum_{n=0}^{\infty} \frac{1}{n!} F_n(\mathbf{h}, \ldots, \mathbf{h}) = F_0 + \frac{1}{2} F_1(\mathbf{h}, \mathbf{h}) + \frac{1}{6} F_3(\mathbf{h}, \mathbf{h}, \mathbf{h}) + \ldots$$

$\mathbf{h}$ is presumed small, so we need to look for the first non-zero non-trivial term.

If $F_1(\mathbf{h}) \neq 0$ for some $\mathbf{h}$, then we have previously shown that $F_1(\mathbf{h})$ can be represented in terms of a basis for which the first component yields a non-zero result but all others are mapped onto zero. $F_1(\mathbf{h}) = \lambda_1^{(T)} \|F_1\|$, where the vector $F_{1i} = F_1(\mathbf{x_i}) \frac{\partial f}{\partial x_i}$ and

$$\mathbf{T_1} = \frac{1}{\|F_1\|} \sum_{i=1}^{n} \mathbf{x_i} F_{1i} = \frac{1}{\|\nabla f\|} \sum_{i=1}^{n} \mathbf{x_i} \frac{\partial f}{\partial x_i} = \frac{\nabla f}{\|\nabla f\|}$$

where we recognise the gradient operator

$$\nabla = \sum_{i=1}^{n} \mathbf{x_i} \frac{\partial}{\partial x_i}.$$

If $F_1(\mathbf{h}) \neq 0$, then there is one direction which is 'up' and all the other orthogonal directions are flat.

If $F_1(\mathbf{h}) = 0$ for all $\mathbf{h}$, then we need to go to $F_2(\mathbf{h}, \mathbf{h})$. We have previously shown that (for a real vector), that there is a basis in which $F_2(\mathbf{x}, \mathbf{y})$ is diagonal:

$$F_2(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{n} \lambda_i^{(T)} F_2^{(i)} \lambda_i'^{(T)}$$

where $F_2^{(i)}$ are the eigenvalues of $F_{2ij} = F_2(\mathbf{x_i}, \mathbf{x_j}) = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

In this basis

$$F_2(\mathbf{h}, \mathbf{h}) = \sum_{i=0}^{n} \left[\lambda_i^{(T)}\right]^2 F_2^{(i)}$$

and each direction yields a quadratic dependence. The chosen directions are just

$$\mathbf{T_i} = \sum_{j=1}^{n} \mathbf{x_j} T_{ji} = \sum_{j=1}^{n} \mathbf{x_j} u_j^{(i)}$$

directly related to the eigenvectors of the appropriate eigenvalue.

The types of behaviour for a turning point of this variety are classified by the eigenvalues related to the curvature matrix:

If all the eigenvalues are positive, then we have a minimum.
If all the eigenvalues are negative, then we have a maximium.
If we find eigenvalues of both signs,then we have a saddle point.

Example:

Consider the fucntion of two-dimensions

$$f(x_1, x_2) = 2 - \cos x_1 - \cos x_2 + \kappa \sin x_1 \sin x_2.$$

Show that this function is a minimum for vectors parallel to the basis vectors, and find the values of κ for which it is a minimum in *all* directions.

The slope or or gradient will tell us whether the point is a turning point

$$\frac{\partial f}{\partial x_i} = (\sin x_1 + \kappa \cos x_1 \sin x_2, \sin x_2 + \kappa \sin x_1 \cos x_2)$$

and this vanishes at the origin.

The curvature matrix tells us the leading order behaviour

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} \cos x_1 - \kappa \sin x_1 \sin x_2 & \kappa \cos x_1 \cos x_2 \\ \kappa \cos x_1 \cos x_2 & \cos x_2 - \kappa \sin x_1 \sin x_2 \end{pmatrix}$$

and at the origin

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{pmatrix} 1 & \kappa \\ \kappa & 1 \end{pmatrix}$$

and so

$$\begin{aligned} F_2(\mathbf{h}, \mathbf{h}) &= \sum_{i=1}^{2} \sum_{j=1}^{2} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j} \\ &= \mathbf{h}^T C \mathbf{h} \\ &= [h_1, h_2] \begin{pmatrix} 1 & \kappa \\ \kappa & 1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \\ &= h_2^2 + h_2^2 + 2\kappa h_1 h_2 \end{aligned} \qquad (18.1)$$

Parallel to $\mathbf{x_1}$ we have $h_2 = 0$ and $F_2(\mathbf{h}, \mathbf{h}) = h_1^2$ so a mininmum.
Parallel to $\mathbf{x_2}$ we have $h_1 = 0$ and $F_2(\mathbf{h}, \mathbf{h}) = h_2^2$ so a mininmum.

To find the general behaviour we should transform to the diagonal basis.

$$\det \begin{pmatrix} 1-\epsilon & \kappa \\ \kappa & 1-\epsilon \end{pmatrix} = (\epsilon - 1)^2 - \kappa^2 = 0$$

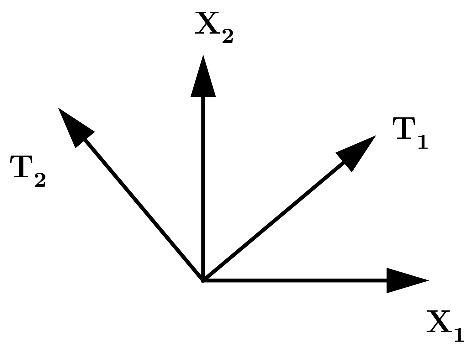
and so $\epsilon = 1 \pm \kappa$ and

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad at \quad \epsilon = 1 + \kappa \qquad \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad at \quad \epsilon = 1 - \kappa.$$

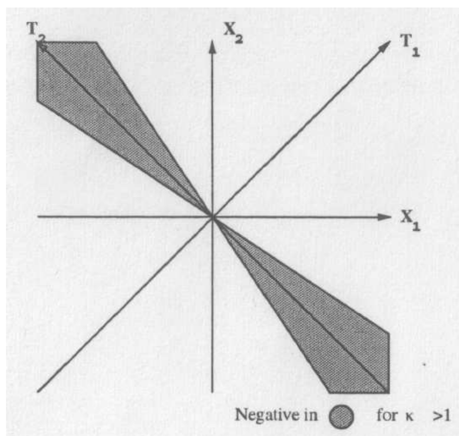
In this new basis

$$F_2(\mathbf{h}, \mathbf{h}) = \left[h_1^{(T)}\right]^2 (1+\kappa) + \left[h_2^{(T)}\right]^2 (1-\kappa)$$

Obviously this is only a minimum for $\kappa \in (-1, 1)$. The new basis is $\mathbf{T_1} = \frac{1}{\sqrt{2}}(\mathbf{x_1} + \mathbf{x_2})$ and $\mathbf{T_2} = \frac{1}{\sqrt{2}}(-\mathbf{x_1} + \mathbf{x_2})$ so



What actually happens when $\kappa > 1$ is that the function vanishes for a pair of straight lines.



(1) Old basis:

$$
\begin{aligned}
&h_1^2 + h_2^2 + 2\kappa h_1 h_2 \\
&(h_1 + \kappa h_2)^2 = (\kappa^2 - 1) h_2^2 \\
&h_1 + \kappa h_2 = \pm(\kappa^2 - 1)^{\frac{1}{2}} h_2 \\
&h_2 + \kappa h_1 = \pm(\kappa^2 - 1)^{\frac{1}{2}} h_1.
\end{aligned}
$$

(2) New basis:

$$
\begin{aligned}
&\left[h_1^{(T)}\right]^2 (1 + \kappa) = \left[h_2^{(T)}\right]^2 (\kappa - 1) \\
&h_2^{(T)} = \pm\sqrt{\frac{1+\kappa}{\kappa - 1}} h_!^{(T)}.
\end{aligned}
$$

□

Example:

Consider the function

$$f(x,y;\lambda) = \frac{1}{2}(\cos x + \cos y) + \frac{\lambda}{2}(\cos 2x + \cos 2y)$$

where (x, y) are the coordinates of a Cartesian vector space and λ is an additional positive parameter. Show that we can restrict attention to $x, y \in (-\pi, \pi]$ and find the position and value of the minimum of this function in this region, following its properties as λ is varied.

Firstly: $f(x + 2\pi, y; \lambda) = f(x, y; \lambda) = f(x, y + 2\pi; \lambda)$ and so $x, y \in (-\pi, \pi]$ provides a maximum region *not* connected to itself by these periodicities. Secondly we need the turning points:

$$\frac{\partial f}{\partial x} = \frac{1}{2}[\sin x + 2\lambda \sin 2x] = -\frac{1}{2}\sin x\,[1 + 4\lambda \cos x]$$

$$\frac{\partial f}{\partial y} = \frac{1}{2}[\sin y + 2\lambda \sin 2y] = -\frac{1}{2}\sin y\,[1 + 4\lambda \cos y]\,.$$

In the region $\lambda \in \left(0, \frac{1}{4}\right)$ we have $x, y \in \{0, \pi\}$ and in the region $\lambda \in \left(\frac{1}{4}, \infty\right)$ we have $x, y \in \{0, k^*, \pi\}$ with $\cos k^* = -\frac{1}{4\lambda}$. Now since each of x and y can be any of these three roots, there are up to *nine styles* of turning point. Since the x nd y behaviour of the nine turning points directly. The curvature matrix is:

$$\frac{\partial^2 f}{\partial x^2} = -\frac{1}{2}[\cos x + 4\lambda \cos 2x], \quad \frac{\partial^2 f}{\partial x \partial y} = 0, \quad \frac{\partial^2 f}{\partial x^2} = -\frac{1}{2}[\cos y + 4\lambda \cos 2y]$$

and when $x = 0$:

$$\frac{\partial^2 f}{\partial x^2} = -\frac{1}{2}(1 + 4\lambda)$$

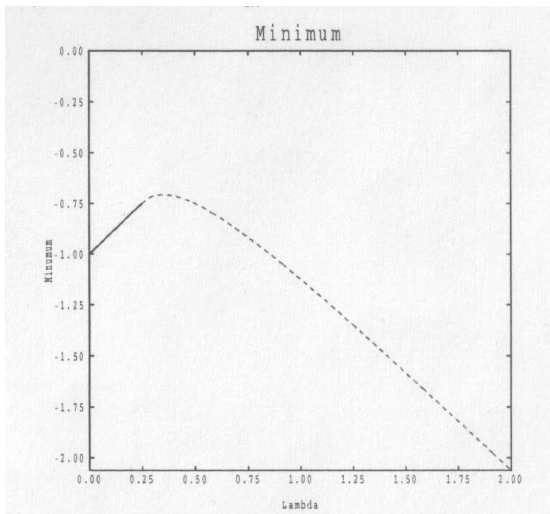
and when $x = \pi$:

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{2}(1 - 4\lambda)$$

when $x = k^*$:

$$
\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &= -\frac{1}{2}\left[-\frac{1}{4\lambda} + 4\lambda\left(2\frac{1}{(4\lambda)^2} - 1\right)\right] \\
&= 2\lambda - \frac{1}{8\lambda} \\
&= \frac{(4\lambda - 1)(4\lambda + 1)}{8\lambda}.
\end{aligned}
$$

In the region $\lambda \in \left(0, \frac{1}{4}\right)$ $x, y = \pi$ and $f = -1 + \lambda$ is the sole minimum, while when $\lambda \in \left(\frac{1}{4}, \infty\right)$ $x, = k^*$ and $f = -\lambda - \frac{1}{8\lambda}$ is the sole minimum.



Chapter 19

Symmetries of Matrices and Commutation

Two matrices are said to commute if:

$$AB = BA$$

Theorem: Two hermitian matrices, A and B, may be *simultaneouslydiagonalised* they commute.

Proof: Let ususe a basis which diagonalises the matrix A, to represent the matrix B:

$$A\mathbf{a}^{(n)} = \alpha^{(n)}\mathbf{a}^{(n)}$$

where $\alpha^{(n)}$ are the eigenvalues and $\mathbf{a}^{(n)}$ are the corresponding eigenvectors. Hermitian conjugate leads to:

$$\mathbf{a}^{(m)\dagger}A = \mathbf{a}^{(m)\dagger}\alpha^{(m)*}$$

Applying a second eigenvector yields:

$$\alpha^{(n)}\mathbf{a}^{(m)\dagger}\mathbf{a}^{(n)} = \mathbf{a}^{(m)\dagger}A\mathbf{a}^{(n)} = \alpha^{(m)*}\mathbf{a}^{(m)\dagger}\mathbf{a}^{(n)}$$

and hence:

$$(\alpha^{(n)} - \alpha^{(m)*})\mathbf{a}^{(m)\dagger}\mathbf{a}^{(n)} = 0.$$

Using $n = m$ shows that $\alpha^{(n)}$ is real. Using $m \neq n$ shows that if $\alpha^{(n)} \neq \alpha^{(m)}$ then $\mathbf{a}^{(m)\dagger}\mathbf{a}^{(n)} = 0$ and the eigenvectors with different eigenvalues are necessarily orthogonal.

An analogous argument involving B and assuming commutation with A provides:

$$\alpha^{(n)}\mathbf{a}^{(m)\dagger}B\mathbf{a}^{(n)} = \mathbf{a}^{(m)\dagger}BA\mathbf{a}^{(n)} = \mathbf{a}^{(m)\dagger}AB\mathbf{a}^{(n)} = \alpha^{(m)*}\mathbf{a}^{(m)\dagger}B\mathbf{a}^{(n)}$$

and

$$(\alpha^{(n)} - \alpha^{(m)*})\mathbf{a}^{(m)\dagger}B\mathbf{a}^{(n)} \equiv (\alpha^{(n)} - \alpha^{(m)*})B_{nm} = 0.$$

This result tellsus that the matrix representation of B is block diagonal in this representation, where the blocks can be labeled by the eigenvalues of A. Each block can then be diagonalised and a linear combination of eigenvectors with the same eigenvalue of A remains an eigenvector of A with that eigenvalue. The simultaneous diagonalisation then follows.

This theorem is often used in diagonalising matrices. If a symmetry of a matrix is observed and the corresponding symmetry matrix can be deduced, then one might use the eigenbasis of the symmetry to block diagonalise the matrix.

Example: Diagonalise:

$$M = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

It isclear that any permutation of th evectors is a symmetry:

$$S_{12} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad S_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad S_{31} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

all commute with the matrix M. Note that $S_{ij}^2 = I$ and so the eigenvalues are ± 1. The subspace with an eigenvalue -1 must also be a block diagonal subspace of M.

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \qquad \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

are all eigenvectors of M, with eigenvalue $m = 1$. Note that these vectors are linearly dependent, and hence that the subspace is only two-dimensional. The other eigenvector must be perpendicular to this subspace and hence:

$$\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$$

is an eigenvector of M with eigenvalue $m = 4$.

Example: Diagonalise:

$$M = \begin{pmatrix}3 & 2 & 1 & 2 & 1 & 1\\ 2 & 3 & 2 & 1 & 1 & 1\\ 1 & 2 & 3 & 2 & 1 & 1\\ 2 & 1 & 2 & 3 & 1 & 1\\ 1 & 1 & 1 & 1 & 3 & -2\\ 1 & 1 & 1 & 1 & -2 & 3\end{pmatrix}$$

by showing that:

$$S = \begin{pmatrix}0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix}$$

is a symmetry.

Direct application shows that $MS = SM$. S is very easy to diagonalise since:

$$\det(S - \epsilon I) = (\epsilon - 1)^2(\epsilon^4 - 1)$$

as only two terms contribute to the determinant. The eigenvectors associated with the eigenvalues $s = -1,\ i,\ -i$ are all unique and hence must be eigenvectors of the original problem:

$$\frac{1}{2}\begin{pmatrix}1\\-1\\1\\-1\\0\\0\end{pmatrix}, \qquad \frac{1}{2}\begin{pmatrix}1\\i\\-1\\-i\\0\\0\end{pmatrix}, \qquad \frac{1}{2}\begin{pmatrix}1\\-i\\-1\\i\\0\\0\end{pmatrix}$$

are eigenvectors with eigenvalues $m = 0, 2, 2$ respectively. A second clear symmetry is exchange of the final two degrees of freedom which provides:

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}$$

as an eigenvector with eigenvalue $m = 5$. The final two eigenvectors must be orthogonal to these four and so we can deduce that they must have the form:

$$\begin{pmatrix} a \\ a \\ a \\ a \\ b \\ b \end{pmatrix}$$

in terms of two unknown parameters a and b. Multiplying the matrix M onto this provides a two-by-two problem:

$$\begin{pmatrix} 8 & 2 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = m \begin{pmatrix} a \\ b \end{pmatrix}$$

which is easily diagonalised to provide:

$$\frac{1}{6}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -4 \\ -4 \end{pmatrix}, \qquad \frac{1}{3\sqrt{2}}\begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \end{pmatrix}$$

with corresponding eigenvalues $m = 0, 9$ respectively.

Chapter 20

Functions of Matrices

The basic idea is to perform 'algebraic' manipulations with matrices with matrices, e.g. e^M, M^{-1}, or even $\sqrt{M}$. The basis in which such manipulations are easiest is the diagonal basis, for which alldegrees of freedom are independent. We start out with powers of our matrix, which are associated with repeated mappings under the associated linear operator:

$$\hat{M}(\hat{M}(\ldots\hat{M}(\mathbf{x})\ldots)) \equiv \hat{M}^n(\mathbf{x})$$

which are repeated as:

$$\hat{M}^n(\mathbf{x_i}) = \sum_{j=1}^{n} \mathbf{x_j} M_{ji}^{(n)}$$

in terms of the matrix $M^{(n)}$. Simply applying the operator to this operator to this relation provides us with the next operator in sequence:

$$\hat{M}^{n+1}(\mathbf{x_i}) = \hat{M}(\hat{M}^n(\mathbf{x_i})) = \sum_{j=1}^{n} \hat{M}(\mathbf{x_j}) M_{ji}^{(n)} = \sum_{k=1}^{n}\sum_{j=1}^{n} \mathbf{x_k} M_{kj} M_{ji}^{(n)}$$

and so $M^{(n+1)} = MM^{(n)}$. Since $M^{(1)} = M$, we find that repeated applications of our original map, provides us with a new linear transformation which is represented by a matrix $M^{(n)} = M^n$, which is simply the matrix multiplication of M by itself n times.

In the original basis:

$$\hat{M}(\mathbf{T_i}) = m^{(i)}\mathbf{T_i}$$

and so:

$$\hat{M}^n(\mathbf{T_i}) = \left[m^{(i)}\right]^n \mathbf{T_i}$$

which amounts to pointing out that a diagonal matrix raised to the n'th power remains diagonal:

$$M^{(n)}_{ij} = M^n_{ij} = \delta_{ij} \left[m^{(i)}\right]^n .$$

Let us remember the general transformation to a new basis

$$\mathbf{T_i} = \sum_{j=1}^{n} \mathbf{x_j} T_{ji}$$

and the resulting transformation for the matrix representation:

$$M^{(T)}_{ij} = \sum_{i'=1}^{n} \sum_{j'=1}^{n} T^{-1}_{ii'} M_{i'j'} T_{j'j}$$

and in matrix notation $M^{(T)} = T^{-1}MT$. For a diagonalisation problem, we use the eigenvectors to make up our transformation: $t_{ij} = u^{(j)}_i$, and then $M^{(T)}$ is diagonal, D say, with $D_{ij} = \delta_{ij} m^{(i)}$. Applying the operator several times yields:

$$\left[M^{(T)}\right]^n = M^{(T)} M^{(T)} \ldots M^{(T)} = T^{-1}MTT^{-1}MT \ldots T^{-1}MT = T^{-1}M^nT$$

and so:

$$M^n = T \left[M^{(T)}\right]^n T^{-1}$$

and in the diagonal basis, $M^n = TD^nT^{-1}$.

Provided that we can represent a general fucntion, $f(x)$, in terms of powers of x, i.e. we can Taylor expand it:

$$f(x) = \sum_{n=0}^{\infty} a_n x^n$$

then we can work in the diagonal space, solely on the eigenvalues, and then transform back to the original representation:

$$f(M) = \sum_{n=0}^{\infty} a_n M^n \tag{20.1}$$

$$T^{-1} f(M) T = \sum_{n=0}^{\infty} a_n T^{-1} M^n T = \sum_{n=0}^{\infty} a_n D^n \equiv f(D)$$

and hence:

$$f(M) = T f(D) T^{-1}$$

where $f(D)$ is a diagonal matrix, with elements $f(m_i)$, the function applied to eachof the eigenvalues, $f(D)_{ij} = \delta_{ij} f(m_i)$.

In practice, we need to find the eigenvalues an eigenvectors, construct $f(D)$ as the diagonal matrix with elements $f(m_i)$ and then construct T from the eigenvectors as columns (*in the correct order*) and multiply out.

Example: If

$$M = \begin{pmatrix} 1 & 1 & 3 & 3 \\ 1 & 1 & 3 & 3 \\ 3 & 3 & 9 & 9 \\ 3 & 3 & 9 & 9 \end{pmatrix}$$

find $X = \sqrt{M}$.

We can 'guess' most of the eigenvectors from symmetry:

$$M = \begin{pmatrix} 1 & 1 & 3 & 3 \\ 1 & 1 & 3 & 3 \\ 3 & 3 & 9 & 9 \\ 3 & 3 & 9 & 9 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \frac{1}{\sqrt{(2a^2 + 2b^2)}} \begin{pmatrix} a \\ a \\ b \\ b \end{pmatrix} \begin{pmatrix} 2 & 6 \\ 6 & 18 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

and we have two eigenvectors at zero energy and a residual 2×2 problem. From the final 2×2:

$$\det \begin{pmatrix} 2 - \epsilon & 6 \\ 6 & 18 - \epsilon \end{pmatrix} = \epsilon(\epsilon - 20)$$

$$\frac{1}{\sqrt{(20)}}\begin{pmatrix}3\\3\\-1\\-1\end{pmatrix} \qquad \frac{1}{\sqrt{(20)}}\begin{pmatrix}1\\1\\3\\3\end{pmatrix}$$

at energies of zero and twenty respectively. The associated transformation is:

$$T=\begin{pmatrix}\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{(20)}} & \frac{1}{\sqrt{(20)}}\\ -\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{(20)}} & \frac{1}{\sqrt{(20)}}\\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{(20)}} & \frac{3}{\sqrt{(20)}}\\ 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{(20)}} & \frac{3}{\sqrt{(20)}}\end{pmatrix} \qquad T^{-1}=T^{\dagger}=\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0\\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{3}{\sqrt{(20)}} & \frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{(20)}} & -\frac{1}{\sqrt{(20)}}\\ \frac{1}{\sqrt{(20)}} & \frac{1}{\sqrt{(20)}} & \frac{3}{\sqrt{(20)}} & \frac{3}{\sqrt{(20)}}\end{pmatrix}$$

$$D=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&\sqrt{(20)}\end{pmatrix}$$

We can build the eventual result from $f(M)=Tf(D)T^{-1}$:

$$\sqrt{M}=T\sqrt{D}T^{-1}=\begin{pmatrix}0&0&0&1\\0&0&0&1\\0&0&0&3\\0&0&0&3\end{pmatrix}T^{\dagger}=\frac{1}{\sqrt{(20)}}\begin{pmatrix}1&1&3&3\\1&1&3&3\\3&3&9&9\\3&3&9&9\end{pmatrix}=\frac{1}{\sqrt{(20)}}M.$$

Note that there is a choice of $n-phase$ in determining a square-root, $\sqrt{D_{ij}}=\delta_{ij}\sqrt{m_i}P_i$, with $P_i=\pm1$ independently. Consequently, there are up to 2^n distinct square-roots!

Example:

$$\begin{pmatrix}2&1\\1&2\end{pmatrix}^{\frac{1}{2}}=\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}P_1\sqrt{3}&0\\0&P_2\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{pmatrix}=\frac{1}{2}\begin{pmatrix}P_1\sqrt{3}&P_2\\P_1\sqrt{3}&-P_2\end{pmatrix}\begin{pmatrix}1\sqrt{3}&1\\1&-1\end{pmatrix}$$
$$=\frac{1}{2}\begin{pmatrix}P_1\sqrt{3}+P_2 & P_1\sqrt{3}-P_2\\ P_1\sqrt{3}-P_2 & P_1\sqrt{3}+P_2\end{pmatrix} \qquad (20.2)$$

$four$ distinct roots.

Chapter 21

The Characteristic Polynomial

The *characteristic polynomial* of a matrix is:

$$P(x) = |xI - M|$$

and vanishes when evaluated at any of the eigenvalues of the matrix. Indeed, this polynomial *defines* the eigenvalues by:

$$P(x) = \prod_{i=1}^{n}(x - m_i).$$

Note that $P(M) = 0$, since:

$$T^{-1}P(M)T = P(D) = 0.$$

This is called the Cayley-Hamilton theorem. The details of the proof and the proof of the extension to matrices that are not diagonalisile (see appendix E). This fact is often very useful, since it allows one to reduce the algebra associated with the matrix M, to polynomial of degree $n-1$. For example, we can find the inverse of a matrix, since:

$$\sum_{m=0}^{n} a_m M^m = 0 \qquad \text{implies} \qquad M^{-1} = -\sum_{m=1}^{n} \frac{a_m}{a_0} M^{m-1}.$$

Example: If

$$M = \begin{pmatrix} 2 & 1 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 2 & 2 & 5 & 4 \\ 2 & 2 & 4 & 5 \end{pmatrix}$$

find M^{-1}.

One way to find the eigenvalues, using a similar technique to the last example:

$$M = \begin{pmatrix} 2 & 1 & 2 & 2 \\ 1 & 2 & 2 & 2 \\ 2 & 2 & 5 & 4 \\ 2 & 2 & 4 & 5 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix} \frac{1}{\sqrt{(2a^2+2b^2)}} \begin{pmatrix} a \\ a \\ b \\ b \end{pmatrix} \begin{pmatrix} 3 & 4 \\ 4 & 9 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

and the residual 2×2 provides:

$$\epsilon^2 - 12\epsilon + 27 - 16 = (\epsilon - 6)^2 - 25 = 0$$

and so the eigenvalues are $\epsilon = 1, 1, 1, 11$. The characteristic polynomial is $P(\epsilon) = (\epsilon - 1)^3(\epsilon - 11) = \epsilon^4 - 14\epsilon^3 + 36\epsilon^2 - 34\epsilon + 11$, and so the inverse is:

$$M^{-1} = \frac{34}{11} - \frac{36}{11}M + \frac{14}{11}M^2 - \frac{1}{11}M^3.$$

We could simply square and cube M and substitute in,but there is some natural structure to the problem: $M = 1 + A$, where $A^2 = 10A$:

$$A = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 2 & 2 & 4 & 4 \\ 2 & 2 & 4 & 4 \end{pmatrix}.$$

In terms of A:

$$\begin{aligned} M^{-1} &= \frac{34}{11} - \frac{36}{11}(1+A) + \frac{14}{11}(1+A)^2 - \frac{1}{11}(1+A)^3 \\ &= 1 - A + A^2 - \frac{1}{11}A^3 \\ &= 1 - \frac{1}{11}A \\ &= \frac{1}{11}[12 - M] \end{aligned}$$

$$M = \frac{1}{11}\begin{pmatrix} 10 & -1 & -2 & -2 \\ -1 & 10 & -2 & -2 \\ -2 & -2 & 7 & -4 \\ -2 & -2 & -4 & 7 \end{pmatrix}.$$

A more direct method is to use A at the outset. $M = 1 + A$, with $A^2 = 10A$.

Then assume:

$$M^{-1} = \alpha + \beta A \quad implies \quad M^{-1}M = 1 = (\alpha+\beta A)(1+A) = \alpha+(\alpha+\beta)A+\beta A^2 = \alpha+(\alpha+11\beta)A$$

and so $\alpha = 1$ and $\alpha + 11\beta = 0$, leading to:

$$M^{-1} = 1 - \frac{1}{11}A = \frac{1}{11}\begin{pmatrix} 10 & -1 & -2 & -2 \\ -1 & 10 & -2 & -2 \\ -2 & -2 & 7 & -4 \\ -2 & -2 & -4 & 7 \end{pmatrix}.$$

Suppose that in

$$f(M) = \sum_{n=0}^{p} a_n M^n$$

we can let $p \to \infty$, so that

$$f(M) = \sum_{n=0}^{\infty} a_n M^n. \tag{21.1}$$

We can attach meaning to $f(M)$ in this case if the matrices

$$f_p(M) = \sum_{n=0}^{p} a_n M^n \tag{21.2}$$

tend to a constant $d \times d$ matrix with finite entries in the limit as $p \to \infty$.

We start with functions of 2×2 matrices. In this case the characteristic equation

$$P(x) = x^2 + a_1 x + a_0 = 0 \tag{21.3}$$

is and it follows from the Cayley-Hamilton theorem that

$$M^2 + a_1 M + a_0 \mathbb{1}_2.$$

The significance of this result for our present purposes begins as we rearrange to give

$$M^2 = -a_1 M - a_0 \mathbb{1}_2$$

This means that M^2 can be written in terms of M and $M^0 = \mathbb{1}_2$. Moreover, multiplying by M gives

$$M^3 = -a_1 M^2 - a_0 M = -a_1(-a_1 M - a_0 \mathbb{1}_2) - a_0 M. \qquad (21.4)$$

Thus M^3 can be expressed in terms of M and $M^0 = \mathbb{1}_2$, that is in terms of powers of A that are less than $d = 2$, the order of the matrix M in this case. It is clear that we can continue the process of multiplying by A and substituting M^2. We easily see that for any integer $r \geq 2$

$$M^r = \alpha_0 \mathbb{1}_2 + \alpha_1 M \qquad (21.5)$$

where α_0 and α_1 are constants whose values depend on r. It is actually easy to determine the α_i $(i = 0, 1)$. To see how to perform the calculations, we use the characteristic equation for M itself. If we assume that the eigenvalues λ_1 and λ_2 of M are distinct then it follows from (21.3) that

$$P(\lambda_i) = \lambda_i^2 + a_1 \lambda_i + a_0 = 0 \quad (i = 1, 2)$$

Thus we can write

$$\lambda_i^2 = -a_1 \lambda_i - a_0$$

in which the constants a_1 and a_0 are the same constants as in (21.3). Then

$$\lambda_i^3 = -a_1 \lambda_i^2 - a_0 \lambda_i = -a_1(-a_1 \lambda_i - a_0) - a_0 \lambda_i$$

in complete analogy with (21.4). Proceding in this way, we deduce that for each of the eigenvalues λ_1 and λ_2 we can write

$$\lambda_i^r = \alpha_0 + \alpha_1 \lambda_i$$

with the same α_0 and α_1 as in (21.5). This therefore provides us with a procedure to calculate M^r when $r \geq 2$. The pair of distinct eigenvalues give us a pair of simultaneous equations from which we are able to evaluate α_0 and α_1 from.

What happens if the 2×2 matrix M has repeated eigenvalues so that $\lambda_1 = \lambda_2 = \lambda$, say? We can obtain a second equation by differentiating the equation $\lambda^r = \alpha_0 + \alpha_1 \lambda$ with respect to λ, to give

$$r\lambda^{r-1} = \alpha_1.$$

Having found a straightforward way of expressing any positive integer power of the 2×2 of a matrix M we see that the same process could be used for each of the terms in (21.2) for $r \geq 2$. Thus, for a 2×2 matrix M and some α_0 and α_1,

$$f(M) = \sum_{n=0}^{p} a_p M^n.$$

If, as $p \to \infty$,

$$f(M) = \lim_{p \to \infty} \sum_{n=0}^{p} a_p M^n$$

exists, that is, it is a matrix with finite entries independent of p, then we may write

$$f(M) = \sum_{n=0}^{\infty} a_n M^n = \alpha_0 \mathbb{1}_2 + \alpha_1 M. \tag{21.6}$$

Although we have worked so far with 2×2 matrices, nothing about the methodology restrics us to this case. The Cayley-Hamilton theorem allows us to express positive integer powers of any $d \times d$ square matrix M in terms of powers of M up to $n-1$. That is, if M is an $d \times d$ matrix and $p \geq d$ then

$$M^p = \sum_{n=0}^{d-1} \beta_n M^n.$$

From this we can deduce that for any $d \times d$ matrix M we may write

$$f(M) = \sum_{n=0}^{\infty} a_n M^n$$

as

$$f(M) = \sum_{n=0}^{d-1} \alpha_n M^n \tag{21.7}$$

which generalises the result (21.6). Again the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_{d-1}$ are obtained by solving the d equations

$$f(\lambda_i) = \sum_{n=0}^{d-1} \alpha_n \lambda_i^n \quad (i = 1, 2, \ldots, d) \tag{21.8}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_d$ are the eigenvalues of M. If M has repeated eigenvalues, we differentiate as before, noting that if λ_i is an eigenvalue of multiplicity m then the first $m-1$ derivatives

$$\frac{d^k}{d\lambda^k} f(\lambda_i) = \frac{d^k}{d\lambda^k} \sum_{n=0}^{d-1} \alpha_n \lambda_i^n \quad (k = 1, 2, \ldots, m-1) \tag{21.9}$$

are also satisfied by λ_i.

21.1 Distinct Eigenvalues

An explicit solution can be easily written down in the case where all the eigenvalues are distinct:

$$f(x) = \sum_{j=0}^{n-1} f(\lambda_j) \prod_{k \neq j} \frac{x - \lambda_k}{\lambda_j - \lambda_k} \tag{21.10}$$

This follows from the fact it takes the correct values at the n data points:

$$\begin{aligned} \sum_{j=0}^{n-1} f(\lambda_j) \prod_{k \neq j} \frac{\lambda_i - \lambda_k}{\lambda_j - \lambda_k} &= \sum_{j=0}^{n-1} f(\lambda_j) \delta_{ij} \\ &= f(\lambda_i) \quad \text{for } 1 \leq i \leq n. \end{aligned} \tag{21.11}$$

and is the unique polynomial having this property. To see that it is a unique polynomial suppose there is another polynomial, $g(x)$, of degree at most $n-1$ that also takes the same values at the n data points. Consider $h(x) = f(x) - g(x)$ Then $h(x)$ is a polynomial of degree at most $n-1$. At the n data points $h(\lambda_i) = f(\lambda_i) - g(\lambda_i) = 0$. Therefore $h(x)$ has n roots, but is of degree $\leq n-1$. As such $h(x)$ is the zero polynomial and hence $f(x) = g(x)$.

Which means that we can write down an explicit expression for (21.7):

$$f(M) = \sum_{j=0}^{n-1} f(\lambda_j) \prod_{k \neq j} \frac{M - \lambda_k \mathbb{1}_d}{\lambda_j - \lambda_k}. \tag{21.12}$$

Chapter 22

Function e^{At} of a Matrix A and Solving Linear System of Differential Equations

A linear system of differential equations of the form

$$\begin{aligned} x_1' &= a_{11}x_1 + \cdots + a_{1n}x_n + f_1, \\ x_2' &= a_{21}x_1 + \cdots + a_{2n}x_n + f_2, \\ &\vdots \qquad \vdots \\ x_n' &= a_{n1}x_1 + \cdots + a_{nn}x_n + f_n \end{aligned} \tag{22.1}$$

where $x_i' = dx_i/dt$. Given the functions $a_{ij}(t)$ and $f_j(t)$ on some interval $a < t < b$. The unkown $x_1(t), \ldots, x_n(t)$.

The system is called homogeneous if all $f_j = 0$, otherwise it is called inhomogeneous.

The inhomogeneous system of linear equations (22.1) can be written as vector-matrix equation:

$$\vec{x}' = A(t)\vec{x} + \vec{f}(t), \tag{22.2}$$

where

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \vec{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ldots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \tag{22.3}$$

Computing e^{At} for Distinct Eigenvalues

Let the $n \times n$ matrix A have distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ (possibly complex) and define the constant matrices $Q_1, \ldots, Q_n$ by

$$Q_j = \prod_{i \neq j} \frac{A - \lambda_k \mathbb{1}}{\lambda_j - \lambda_k}, \quad j = 1, \ldots, n. \tag{22.4}$$

Then

$$e^{At} = e^{\lambda_1 t} Q_1 + \cdots + e^{\lambda_n t} Q_n \tag{22.5}$$

Computing e^{At} for Multiple Eigenvalues

Let the $n \times n$ matrix A have distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ of algebraic multiplicities $m_1, \ldots, m_k$. Let $p(\lambda) = \det(A - \lambda \mathbb{1})$ and define the polynomials $a_1(\lambda), \ldots, a_n(\lambda)$ by the partial fraction identity

$$\frac{1}{p(\lambda)} = \frac{a_1(\lambda)}{(\lambda - \lambda_1)^{m_1}} + \cdots + \frac{a_k(\lambda)}{(\lambda - \lambda_k)^{m_k}} \tag{22.6}$$

Define the constant matrices $Q_1, \ldots, Q_k$ by

$$Q_j = a_1(A) \prod_{i \neq j} (A - \lambda_i \mathbb{1})^{m_i}, j = 1, \ldots, k. \tag{22.7}$$

We will prove that

$$e^{At} = \sum_{i=1}^{k} e^{\lambda_i t} Q_i \sum_{j=0}^{m_i - 1} (A - \lambda_i \mathbb{1})^j \frac{t^j}{j!}. \tag{22.8}$$

Proof:

Let $N_i = Q_i (A - \lambda_i \mathbb{1})$, $1 \leq i \leq k$. We prove the following:

1. $Q_1 + \cdots + Q_k = \mathbb{1}$,
2. $Q_i Q_i = Q_i$,
3. $Q_i Q_j = 0$ for $i \neq j$,
4. $N_i N_j = 0$ for $i \neq j$,

5. $N_i^{m_i} = 0$,
6. $A = \sum_{i=1}^{k} (\lambda_i Q_i + N_i)$.

To prove **1** first note that

$$0 = \sum_{i=1}^{k} a_i(\lambda) \frac{p(\lambda)}{(\lambda - \lambda_i)^{m_i}} - 1 \tag{22.9}$$

meaning that the coefficients of λ^0 on the RHS add up to zero, the coefficients of λ on the RHS add up to zero, the coefficients of λ^2 on the RHS add up to zero, etc. Therefore, if we were to write (22.9) with λ replaced by the matrix A the coefficients of A^0 on the RHS would add up to zero, the coefficients of A on the RHS would add up to zero, the coefficients of A^2 on the RHS would add up to zero, etc. Proving:

$$\mathbb{1} = Q_1 + \cdots + Q_k.$$

To prove **3** starts from observing Q_i and Q_j together contain all the factors of $p(A)$, therefore $Q_i Q_j = q(A)p(A)$ for some polynomial q. Thus $Q_i Q_j = 0$ by the Cayley-Hamiltoian theorem.

To prove **2** multiply $Q_1 + \cdots + Q_k = \mathbb{1}$ by Q_i and use **3**.

To prove **4**, write $N_i N_j = (A - \lambda_i \mathbb{1})(A - \lambda_j \mathbb{1}) Q_i Q_j$ and apply **3**.

To prove **5** use $Q_i^{m_i} = Q_i$ (by **2**) to write $N_i^{m_i} = (A - \lambda_i \mathbb{1})^{m_i} Q_i = p(A) = 0$.

Property **6** follows from **1**:

$$\begin{aligned} A &= \sum_{i=1}^{k} A Q_i \\ &= \sum_{i=1}^{k} (\lambda_i Q_i + (A - \lambda_i \mathbb{1}) Q_i) \\ &= \sum_{i=1}^{k} (\lambda_i Q_i + N_i). \end{aligned} \tag{22.10}$$

To prove (22.8), multiply $\mathbb{1}$ by e^{At} and compute as follows:

$$\begin{aligned}
e^{At} &= \sum_{i=1}^{k} Q_i e^{At} \\
&= \sum_{i=1}^{k} Q_i e^{\lambda_i \mathbb{1} t + (A-\lambda_i \mathbb{1})t} \\
&= \sum_{i=1}^{k} Q_i e^{\lambda_i \mathbb{1} t} e^{(A-\lambda_i \mathbb{1})t} \\
&= \sum_{i=1}^{k} Q_i e^{\lambda_i \mathbb{1} t} e^{Q_i(A-\lambda_i \mathbb{1})t} \quad (\text{as } Q_i^2 = Q_i) \\
&= \sum_{i=1}^{k} Q_i e^{\lambda_i \mathbb{1} t} e^{N_i t} \\
&= \sum_{i=1}^{k} Q_i e^{\lambda_i \mathbb{1} t} \sum_{j=0}^{m_i-1} N_i^j \frac{t^j}{j!} \quad (\text{as } N_i^{m_i} = 0) \\
&= \sum_{i=1}^{k} Q_i e^{\lambda_i \mathbb{1} t} \sum_{j=0}^{m_i-1} (A-\lambda_i \mathbb{1})^j \frac{t^j}{j!} \quad (\text{as } Q_i^2 = Q_i).
\end{aligned}$$

22.1 Solving Linear Systems $\vec{x}'(t) = A\vec{x}(t)$

Example:

Consider the real 2×2 system $\vec{x}'(t) = A\vec{x}(t)$. Let λ_1, λ_2 by the roots of the characteristic equation $\det(A - \lambda \mathbb{1}) = 0$. The real general solution $\vec{x}(t)$ is given bu the formula

$$\vec{x}(t) = e^{At}\vec{x}(0)$$

where the 2×2 exponential matrix e^{At} is given as follows:

(i) Real $\lambda_1 \neq \lambda_2$ we have

$$e^{At} = e^{\lambda_1 t}\mathbb{1} + \frac{e^{\lambda_2 t} - e^{\lambda_1 t}}{\lambda_2 - \lambda_1}(A - \lambda_1 \mathbb{1}). \tag{22.11}$$

(ii) Real $\lambda_1 = \lambda_2$ we have

$$e^{At} = e^{\lambda_1 t}\mathbb{1} + te^{\lambda_1 t}(A - \lambda_1 \mathbb{1}). \tag{22.12}$$

(iii) Complex $\lambda_1 = \overline{\lambda}_2$ with $\lambda_1 = a + ib$, $b > 0$, we have

$$e^{At} = e^{at}\cos(bt)\mathbb{1} + \frac{e^{at}\sin(bt)}{b}(A - a\mathbb{1}). \tag{22.13}$$

Proof:

(i) By (22.5) we have

$$\begin{aligned} e^{At} &= -e^{\lambda_1 t}\frac{A-\lambda_2\mathbb{1}}{\lambda_2-\lambda_1} + e^{\lambda_2 t}\frac{A-\lambda_1\mathbb{1}}{\lambda_2-\lambda_1} \\ &= \frac{e^{\lambda_1 t}\lambda_2 - e^{\lambda_2 t}\lambda_1}{\lambda_2-\lambda_1}\mathbb{1} + \frac{e^{\lambda_2 t}-e^{\lambda_1 t}}{\lambda_2-\lambda_1}A \\ &= \frac{e^{\lambda_1 t}\lambda_2 + (-e^{\lambda_1 t}\lambda_1 + e^{\lambda_1 t}\lambda_1) - e^{\lambda_2 t}\lambda_1}{\lambda_2-\lambda_1}\mathbb{1} + \frac{e^{\lambda_1 t}-e^{\lambda_2 t}}{\lambda_2-\lambda_1}A \\ &= e^{\lambda_1 t}\mathbb{1} + \frac{e^{\lambda_2 t}-e^{\lambda_1 t}}{\lambda_2-\lambda_1}(A-\lambda_1\mathbb{1}). \end{aligned} \tag{22.14}$$

(ii) It is easy to see that

$$\begin{aligned} \lim_{\lambda_2\to\lambda_1}\frac{e^{\lambda_2 t}-e^{\lambda_1 t}}{\lambda_2-\lambda_1} &= e^{\lambda_1 t}\lim_{\lambda_2\to\lambda_1}\frac{e^{(\lambda_2-\lambda_1)t}-1}{\lambda_2-\lambda_1} \\ &= te^{\lambda_1 t} \end{aligned}$$

and so (ii) can be seen to follow from (i). Alternatively, we can use (22.8). First $Q_1 = \mathbb{1}$. Then

$$\begin{aligned} e^{At} &= e^{\lambda_1 t}\sum_{j=0}^{1}(A-\lambda_1)^j\frac{t^j}{j!} \\ &= e^{\lambda_1 t}\mathbb{1} + te^{\lambda_1 t}(A-\lambda_1). \end{aligned} \tag{22.15}$$

(iii) The complex case is formally the real part of the distinct roots case when $\lambda_2 = \overline{\lambda}_1$:

$$\begin{aligned} e^{At} &= \Re\left(e^{at+ibt}\mathbb{1} + e^{at}\frac{e^{-ibt}-e^{ibt}}{-ib-ib}(A-(a+ib)\mathbb{1})\right) \\ &= e^{at}\cos(bt)\mathbb{1} + \frac{e^{at}\sin(bt)}{b}(A-a\mathbb{1}). \end{aligned}$$

22.2 Jordan Form

The main result of generalised eigenanalysis is Jordan's theorem

$$A = PJP^{-1}, \tag{22.16}$$

for any real or complex square matrix A.

22.2.1 Exponential of a Jordan Form

$$B = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \qquad (m \times m) \tag{22.17}$$

Define

$$N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{22.18}$$

As $(\lambda\mathbb{1})N = N(\lambda\mathbb{1})$, we have that $e^{tB} = e^{t(\lambda\mathbb{1}+N)} = e^{t\lambda\mathbb{1}}e^{tN}$. Then

$$e^{tB} = e^{t\lambda} \begin{pmatrix} 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{m-2}}{(m-2)!} & \frac{t^{m-1}}{(m-1)!} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & t \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{22.19}$$

22.2.2 Exponential of a Jordan Block Form for real matrix with two complex eigenvalues

If the Jordan chain $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ corresponds to eigenvalue $\lambda = a + ib$, then there will be a Jordan chain $\{\overline{\mathbf{u}}_1, \ldots, \overline{\mathbf{u}}_m\}$ corresponds to eigenvalue $\overline{\lambda} = a - ib$. The Jordan chain

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \overline{\mathbf{u}}_1, \ldots, \overline{\mathbf{u}}_m\}$$

has the Jordan matrix:

$$B = \begin{pmatrix} B_m & 0_m \\ 0_m & \overline{B}_m \end{pmatrix} \tag{22.20}$$

where

$$B_m = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \qquad (m \times m), \tag{22.21}$$

$\overline{B}_m$ is the complex conjugate of B_m, and 0_m is the $m \times m$ zero matrix.

If we switch from using the above Jordan chain to using the basis

$$\{\Re(\mathbf{u}_1), \Im(\mathbf{u}_1), \ldots, \Re(\mathbf{u}_m), \Im(\mathbf{u}_m)\}$$

instead, we obtain the following Jordan block matrix of dimension $2m \times 2m$:

$$B = \begin{pmatrix} \Lambda & \mathcal{I} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \Lambda & \mathcal{I} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \Lambda \end{pmatrix} \tag{22.22}$$

where

$$\Lambda = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathcal{O} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \tag{22.23}$$

This is spelt out in appendix G.6.1. We now turn to taking the exponential of the Jordan block matrix. First define

$$K = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

and note that

$$
\begin{aligned}
e^{t\Lambda} &= e^{t(a\mathcal{I}+bK)} \\
&= e^{ta\mathcal{I}}e^{tbK} \\
&= e^{ta}\left\{\left(1-\frac{t^2b^2}{2!}+\frac{t^4b^4}{4!}-\cdots\right)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}+\left(tb-\frac{t^3b^3}{3!}+\frac{t^5b^5}{5!}-\cdots\right)\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\right\} \\
&= e^{ta}\begin{pmatrix}\cos(tb) & \sin(bt)\\ -\sin(bt) & \cos(tb)\end{pmatrix}. \qquad (22.24)
\end{aligned}
$$

Write $\mathcal{R} = \begin{pmatrix}\cos(tb) & \sin(bt)\\ -\sin(bt) & \cos(tb)\end{pmatrix}$, then

$$
e^{tB} = e^{ta}\begin{pmatrix}
\mathcal{R} & t\Lambda & \frac{t^2}{2}\mathcal{R} & \cdots & \frac{t^{m-2}}{(m-2)!}\mathcal{R} & \frac{t^{m-1}}{(m-1)!}\mathcal{R} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{R} & t\mathcal{R} \\
\mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{R}
\end{pmatrix} \qquad (22.25)
$$

Solving $\mathbf{x}' = A\mathbf{x}$

The solution $\mathbf{x}(t) = e^{tA}\mathbf{x}(0)$ must be real if A is real. The real solution can be expressed as $\mathbf{x}(t) = \mathcal{P}\mathbf{y}(t)$ where $\mathbf{y}'(t) = R\mathbf{y}(t)$ and R is a real Jordan form of A, containing real Jordan blocks B down the diagonal.

We have the formula

$$
\mathbf{x}(t) = \mathcal{P}e^{Rt}\mathcal{P}^{-1}\mathbf{x}(0) \qquad (22.26)
$$

contain only real numbers, real exponentials, plus sine and cosine terms, which are possibly multiplied by polynomials in t.

22.3 Second-order Systems

Conversion between $\mathbf{x}'' = A\mathbf{x}$ and $\mathbf{u}' = C\mathbf{u}$

A second-order system can easily be rewritten as a first-order system. Consider the second-order system

$$
\mathbf{x}'' = A\mathbf{x} \qquad (22.27)
$$

where A is an $n \times n$ matrix, with initial conditions

$$\mathbf{x}(0) = \mathbf{x}_0, \quad \mathbf{x}'(0) = \mathbf{x}'_0. \tag{22.28}$$

Define the variable $\mathbf{u}$ and $2n \times 2n$ block matrix C as follows

$$\mathbf{u} = \begin{pmatrix} \mathbf{x} \\ \mathbf{x}' \end{pmatrix}, \quad C = \begin{pmatrix} 0 & \mathbb{1}_n \\ A & 0 \end{pmatrix} \tag{22.29}$$

with initial conditions

$$\mathbf{u}(0) = \begin{pmatrix} \mathbf{x}_0 \\ \mathbf{x}'_0 \end{pmatrix}. \tag{22.30}$$

The solution to $\mathbf{u}' = C\mathbf{u}$ gives the solution to $\mathbf{x}'' = A\mathbf{x}$ by the formula

$$\mathbf{x} = \mathbf{diag}(\mathbb{1}_n, 0)\mathbf{u} \quad \text{that is} \quad \mathbf{x} = \begin{pmatrix} \mathbb{1}_n & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{x}' \end{pmatrix} \tag{22.31}$$

Characteristic equation for $\mathbf{x}'' = A\mathbf{x}$

The characteristic equation for the $n \times n$ second-order system $\mathbf{x}'' = A\mathbf{x}$ can be obtained from the corresponding $2n \times 2n$ first-order system $\mathbf{u}' = C\mathbf{u}$. We prove the following identity.

Let $\mathbf{x}''A\mathbf{x}$ be given with A an $n \times n$ constant matrix and let $\mathbf{u}' = C\mathbf{u}$ be its corresponding first-order system (22.29). Then

$$\det(C - \lambda\mathbb{1}_{2n}) = (-1)^n \det(A - \lambda^2\mathbb{1}_n). \tag{22.32}$$

Proof:

$$\begin{pmatrix} -\lambda\mathbb{1}_n & \mathbb{1}_n \\ A & -\lambda\mathbb{1}_n \end{pmatrix} \begin{pmatrix} \mathbb{1}_n & 0 \\ \lambda\mathbb{1}_n & \mathbb{1}_n \end{pmatrix} = \begin{pmatrix} 0 & \mathbb{1}_n \\ A - \lambda^2\mathbb{1}_n & -\lambda\mathbb{1}_n \end{pmatrix} \tag{22.33}$$

Taking the determinant of both sides

$$\det(C - \lambda\mathbb{1}_{2n}) \det \begin{pmatrix} \mathbb{1}_n & 0 \\ \lambda\mathbb{1}_n & \mathbb{1}_n \end{pmatrix} = \det \begin{pmatrix} 0 & \mathbb{1}_n \\ A - \lambda^2\mathbb{1}_n & -\lambda\mathbb{1}_n \end{pmatrix} \tag{22.34}$$

and using that

$$\det \begin{pmatrix} \mathbb{1}_n & 0 \\ \lambda \mathbb{1}_n & \mathbb{1}_n \end{pmatrix} = 1, \quad \det \begin{pmatrix} 0 & \mathbb{1}_n \\ A - \lambda^2 \mathbb{1}_n & -\lambda \mathbb{1}_n \end{pmatrix} = (-1)^n \det(A - \lambda^2 \mathbb{1}_n) \tag{22.35}$$

we obtain (22.32).

□

22.3.1 Solving $\mathbf{u}' = C\mathbf{u}$ and $\mathbf{x}'' = A\mathbf{x}$

Eigenanalysis of A and C

Let A be a given $n \times n$ constant matrix and define the $2n \times 2n$ block matrix (**??**). Then

$$(C - \lambda \mathbb{1}_{2n}) \begin{pmatrix} \mathbf{w} \\ \mathbf{z} \end{pmatrix} = \mathbf{0} \quad \text{if and only if} \quad \begin{cases} A\mathbf{w} = \lambda^2 \mathbf{w}, \\ \mathbf{z} = \lambda \mathbf{w}. \end{cases} \tag{22.36}$$

Proof:

The result is obtained by block multiplication, because

$$C - \lambda \mathbb{1}_{2n} = \begin{pmatrix} -\lambda \mathbb{1}_n & \mathbb{1}_n \\ A & -\lambda \mathbb{1}_n \end{pmatrix}.$$

□

Eigenanalysis when A has negative eigenvalues

If all the eigenvalues μ of A are negative or zero, then, for some $\omega \geq 0$, eigenvalue μ is related to an eigenvalue λ of C by the relation $\mu = -\omega^2 = \lambda^2$. Then $\lambda = \pm i\omega$ and $\omega = \sqrt{-\mu}$. Consider an eigenpair $(-\omega^2, \mathbf{v})$ of the $n \times n$ matrix A with $\omega \geq 0$ and let

$$u(t) = \begin{cases} c_1 \cos \omega t + c_2 \sin \omega t & \omega > 0 \\ c_1 + c_2 t & \omega = 0 \end{cases} \tag{22.37}$$

Appendix A

Supplement: Interchangeability of Norm and Inner Product

A.1 Real Case

If we have a real vector space, denote it X, equipped with an inner product and let $\|\cdot\|$ denote the induced norm. Then

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \tag{A.1}$$

for all $x, y \in X$. This is called the Parallelogram law.

Conversely, if we have a vector space, X, equipped with a norm such that the Parallelogram law (A.1) holds then there exists a real inner product $\langle\cdot,\cdot\rangle$ which induces the norm $\|\cdot\|$; i.e. $\sqrt{(\mathbf{x},\mathbf{x})} = \|\mathbf{x}\|$ for each $\mathbf{x} \in X$.

Proof:

If we have a real (or complex) inner product space then

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y},\mathbf{x}+\mathbf{y}) + (\mathbf{x}-\mathbf{y},\mathbf{x}-\mathbf{y}) \\ &= (\mathbf{x},\mathbf{x}) + (\mathbf{x},\mathbf{y}) + (\mathbf{y},\mathbf{x}) + (\mathbf{y},\mathbf{y}) + (\mathbf{x},\mathbf{x}) - (\mathbf{x},\mathbf{y}) - (\mathbf{y},\mathbf{x}) + (\mathbf{y},\mathbf{y}) \\ &= 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \qquad (A.2) \end{aligned}$$

For the converse statement. Assume the parallelogram law (A.1) holds for all $\mathbf{x}, \mathbf{y} \in X$. Define a real $\langle\cdot,\cdot\rangle$ by

$$(\mathbf{x},\mathbf{y}) := \frac{1}{4}(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2) \tag{A.3}$$

It is clear that $(\mathbf{x},\mathbf{x}) = \|\mathbf{x}\|^2$ for each $\mathbf{x} \in X$. It is also clear that $(\cdot,\cdot)$ is symmetric: $(\mathbf{x},\mathbf{y}) = (\mathbf{y},\mathbf{x})$ as $\|\mathbf{x}-\mathbf{y}\| = \|\mathbf{y}-\mathbf{x}\|$. It remains to verify the linearity condition. We do this via the following steps:

1. $(\mathbf{x}+\mathbf{y},\mathbf{z}) = (\mathbf{x},\mathbf{z}) + (\mathbf{y},\mathbf{z})$,
2. $(\alpha\mathbf{x},\mathbf{y}) = \alpha(\mathbf{x},\mathbf{y})$ for all rational numbers α,
3. $(\alpha\mathbf{x},\mathbf{y}) = \alpha(\mathbf{x},\mathbf{y})$ for all real numbers α,

Note that

$$\begin{aligned}
4(\mathbf{x},\mathbf{z}) + 4(\mathbf{y},\mathbf{z}) &= \|\mathbf{x}+\mathbf{z}\|^2 - \|\mathbf{x}-\mathbf{z}\|^2 + \|\mathbf{y}+\mathbf{z}\|^2 - \|\mathbf{y}-\mathbf{z}\|^2 \\
&= \frac{1}{2}\left(\|\mathbf{x}+2\mathbf{z}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2\right) - \frac{1}{2}\left(\|\mathbf{x}-2\mathbf{z}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2\right) \quad \text{by } (A.1) \\
&= \frac{1}{2}\|\mathbf{x}+2\mathbf{z}+\mathbf{y}\|^2 - \frac{1}{2}\|\mathbf{x}-2\mathbf{z}+\mathbf{y}\|^2 \\
&= \frac{1}{2}\|(\mathbf{x}+\mathbf{z}+\mathbf{y})+\mathbf{z}\|^2 - \frac{1}{2}\|(\mathbf{x}-\mathbf{z}+\mathbf{y})-\mathbf{z}\|^2 \\
&= \frac{1}{2}\left(2\|\mathbf{x}+\mathbf{z}+\mathbf{y}\|^2 + 2\|\mathbf{z}\|^2 - \|\mathbf{x}+\mathbf{y}\|^2\right) \\
&\quad -\frac{1}{2}\left(2\|\mathbf{x}-\mathbf{z}+\mathbf{y}\|^2 + 2\|\mathbf{z}\|^2 - \|\mathbf{x}+\mathbf{y}\|^2\right) \quad \text{by } (A.1) \\
&= \|\mathbf{x}+\mathbf{z}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{z}+\mathbf{y}\|^2 \\
&= 4(\mathbf{x}+\mathbf{y},\mathbf{z}).
\end{aligned}$$

Hence we have proved

$$(\mathbf{x},\mathbf{z}) + (\mathbf{y},\mathbf{z}) = (\mathbf{x}+\mathbf{y},\mathbf{z}) \tag{A.4}$$

for all $\mathbf{x},\mathbf{y},\mathbf{z} \in X$.

Note that by replacing $\mathbf{x}$ with $\mathbf{x}-\mathbf{y}$ in (A.4) we obtain

$$(\mathbf{x},\mathbf{z}) - (\mathbf{y},\mathbf{z}) = (\mathbf{x}-\mathbf{y},\mathbf{z}) \tag{A.5}$$

for all $\mathbf{x},\mathbf{y},\mathbf{z} \in X$.

We now prove that $(\alpha\mathbf{x}, \mathbf{z}) = \alpha(\mathbf{x}, \mathbf{z})$ for all rational numbers α. To see this, first note that (A.4) with $\mathbf{x} = \mathbf{y}$ implies $2(\mathbf{x}, \mathbf{z}) = (2\mathbf{x}, \mathbf{z})$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X$. We prove by induction that

$$m(\mathbf{x}, \mathbf{z}) = (m\mathbf{x}, \mathbf{z}) \tag{A.6}$$

for all positive integrs m and all $\mathbf{x}, \mathbf{z} \in X$: assume $(m-1)(\mathbf{x}, \mathbf{z}) = ((m-1)\mathbf{x}, \mathbf{z})$, then using (A.4), $m(\mathbf{x}, \mathbf{z}) = (m-1)(\mathbf{x}, \mathbf{z}) + (\mathbf{x}, \mathbf{z}) = ((m-1)\mathbf{x}, \mathbf{z}) + (\mathbf{x}, \mathbf{z}) = (m\mathbf{x}, \mathbf{z})$.

However, (A.6) implies $n(\mathbf{x}/n, \mathbf{z}) = (\mathbf{x}, \mathbf{z})$ for all positive integrs m and all $\mathbf{x}, \mathbf{z} \in X$. Hence,

$$\frac{m}{n}(\mathbf{x}, \mathbf{z}) = \left(\frac{m}{n}\mathbf{x}, \mathbf{z}\right) \tag{A.7}$$

By definition every point of the real line is a limit point of a sequence of rational numbers. We now claim that, for fixed $\mathbf{x}, \mathbf{z} \in X$ and $\alpha_n \to \alpha$ we have

$$(\alpha_n\mathbf{x}, \mathbf{z}) \to (\alpha\mathbf{x}, \mathbf{z}).$$

At this point we need to prove the continuity of the norm. Let $(\mathbf{x}_n)$ be a sequence in the norm vector space X converging to $\mathbf{x}$, continuity in the norm means $\|\mathbf{x}_n\| \to \|\mathbf{x}\|$ as $(\mathbf{x}_n)$ converges to $\mathbf{x}$. The norm satisfies the triangle inequality: $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. To prove continuity in the norm we need a variation in the triangle inequality:

$$\|\mathbf{x} - \mathbf{y}\| \geq |\ \|\mathbf{x}\| - \|\mathbf{y}\|\ | \tag{A.8}$$

To prove this it sufficies to prove that

$$\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|; \tag{A.9}$$

for it follows from (A.9) that we also have

$$\begin{aligned} -(\|\mathbf{x}\| - \|\mathbf{y}\|) &= \|\mathbf{y}\| - \|\mathbf{x}\| \\ &\leq \|\mathbf{y} - \mathbf{x}\| \\ &= \|-(\mathbf{x} - \mathbf{y})\| \\ &= \|\mathbf{x} - \mathbf{y}\| \end{aligned}$$

which together with (A.9) yields (A.8). We now prove (A.9) by observing that $\|\mathbf{x}\| = \|(\mathbf{x} - \mathbf{y}) - \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\|$.

So now let $(\mathbf{x}_n)$ be a sequence in the norm vector space X converging to $\mathbf{x}$, then by (A.8)

$$|\ \|\mathbf{x}_n\| - \|\mathbf{x}\|\ | \leq \|\mathbf{x}_n - \mathbf{x}\| \to 0$$

as $n \to \infty$.

By continuity of the norm

$$(\alpha_n \mathbf{x}, \mathbf{y}) = \frac{1}{4}\left(\|\alpha_n \mathbf{x} + \mathbf{z}\|^2 - \|\alpha_n \mathbf{x} - \mathbf{z}\|^2\right) \to \frac{1}{4}\left(\|\alpha \mathbf{x} + \mathbf{z}\|^2 - \|\alpha \mathbf{x} - \mathbf{z}\|^2\right) = (\alpha \mathbf{x}, \mathbf{y})$$

$$\alpha(\mathbf{x}, \mathbf{z}) = (\alpha \mathbf{x}, \mathbf{z}) \tag{A.10}$$

for all $\alpha \geq 0$ and all $\mathbf{x} \in X$.

Finally, for $\alpha < 0$, note first that (A.5)

$$\begin{aligned}
(-\mathbf{x}, \mathbf{z}) &= (\mathbf{0} - \mathbf{x}, \mathbf{z}) \\
&= (\mathbf{0}, \mathbf{z}) - (\mathbf{x}, \mathbf{z}) \\
&= \mathbf{0} - (\mathbf{x}, \mathbf{z}) \\
&= -(\mathbf{x}, \mathbf{z})
\end{aligned}$$

and therefore, by (A.10)

$$\begin{aligned}
\alpha(\mathbf{x}, \mathbf{z}) &= -(-\alpha)(\mathbf{x}, \mathbf{z}) \\
&= -(-\alpha \mathbf{x}, \mathbf{z}) \\
&= (\alpha \mathbf{x}, \mathbf{z})
\end{aligned}$$

as required. This completes the proof.

A.2 Complex Case

If we have a complex inner product space then

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \tag{A.11}$$

The calculation is the same as in (A.2).

For the converse statement. Suppose that X is a complex normed vector space whose norm $\|\cdot\|$ satisfies the parallelogram law, then

$$(\mathbf{x},\mathbf{y}) := \frac{1}{4}\left[(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2) - i(\|\mathbf{x}+i\mathbf{y}\|^2 - \|\mathbf{x}-i\mathbf{y}\|^2)\right] \tag{A.12}$$

is an inner product on X for all $\mathbf{x},\mathbf{y}\in X$ such that $\|\mathbf{x}\| = \sqrt{(\mathbf{x},\mathbf{x})}$ for all $\mathbf{x}\in X$.

Note that

$$(\mathbf{x},\mathbf{y}) = \frac{1}{4}\sum_{k=0} i^{-k}\|\mathbf{x}+i^k\mathbf{y}\|^2.$$

It immediately follows from the definition that

$$(\mathbf{x},\mathbf{x}) = \|\mathbf{x}\|^2$$

since $\|\mathbf{x}+i\mathbf{x}\|^2 = |1+i|^2\|\mathbf{x}\|^2 = |1-i|^2\|\mathbf{x}\|^2 = \|\mathbf{x}-i\mathbf{x}\|^2$, and it immediately follows that

$$(\mathbf{y},\mathbf{x}) = (\mathbf{x},\mathbf{y})^*.$$

Using the parallelogram law, we find for any for all $\mathbf{x},\mathbf{y},\mathbf{z}\in X$ that

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}+2i^k z\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 &= \|(\mathbf{x}+i^k\mathbf{z}) + (\mathbf{y}+i^k\mathbf{z})\|^2 + \|(\mathbf{x}+i^k\mathbf{z}) - (\mathbf{y}+i^k\mathbf{z})\|^2 \\ &= 2\|\mathbf{x}+i^k\mathbf{z}\|^2 + 2\|\mathbf{y}+i^k\mathbf{z}\|^2. \end{aligned} \tag{A.13}$$

Multiplying this equation by i^{-k}, summing the result from 0 to 3, using the fact that $\sum_{k=0}^{3} i^{-k} = 0$, and using the definition of $(\cdot,\cdot)$, we get

$$(\mathbf{x}+\mathbf{y}, 2\mathbf{z}) = 2(\mathbf{x},\mathbf{z}) + 2(\mathbf{y},\mathbf{z}).$$

Setting $\mathbf{y} = \mathbf{0}$ in this equation, and using that $(\mathbf{0},\mathbf{z}) = 0$, we get $(\mathbf{x}, 2\mathbf{z}) = 2(\mathbf{x},\mathbf{z})$. It then follows that

$$(\mathbf{x}+\mathbf{y}, \mathbf{z}) = (\mathbf{x},\mathbf{z}) + (\mathbf{y},\mathbf{z}).$$

Since $(\mathbf{y},\mathbf{x}) = (\mathbf{x},\mathbf{y})^*$, we also have that

$$(\mathbf{x}, \mathbf{y} + \mathbf{z}) = (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{z}).$$

By induction we have for all positive integers m,

$$(\mathbf{x}, m\mathbf{y}).$$

It follows that for any positive integers m, n that

$$\frac{m}{n}(\mathbf{x}, \mathbf{y}) = \left(\mathbf{x}, \frac{m}{n}\mathbf{y}\right).$$

Finally, using that every point of the real line is a limit point of a sequence of rational numbers, the continuity of the norm, and the immediate properties

$$(\mathbf{x}, -\mathbf{y}) = -(\mathbf{x}, \mathbf{y}), \qquad (\mathbf{x}, i\mathbf{y}) = i(\mathbf{x}, \mathbf{y}),$$

we can conclude that

$$(\mathbf{x}, \lambda\mathbf{y}) = \lambda(\mathbf{x}, \mathbf{y})$$

for all complex numbers λ and for all $\mathbf{x}, \mathbf{y} \in X$. This completes the proof that $(\cdot, \cdot)$ defines an inner product.

Appendix B

Totally Antisymmetric symbol and Determinates

B.0.1 First Definition of Determinants

The determinant of 2 dimenional square matrices is defined as

$$\det \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = A_{11}A_{22} - A_{21}A_{12}. \tag{B.1}$$

The determinant of a 3-dimensional square matrix is defined by

$$\begin{aligned} \det \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} &= A_{11} \det \begin{pmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{pmatrix} - A_{12} \det \begin{pmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{pmatrix} \\ &\quad + A_{13} \det \begin{pmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{pmatrix} \end{aligned} \tag{B.2}$$

the elements of thee top row being taken in order from left to right with alternate signs and multiplied by the determinants of 2-diemnsional matrices which remain when the row and column through the element are deleted.

The definition can be extended to determinants of square matrices of any order iteratively in a similar way; the determinant of an $n-$dimensional matrix is expressible in terms of determinants of $(n-1)-$dimension matrices formed from the original $n-$dimensional matrix.

B.0.2 The Levi-Civita Tensor

Determinants in $n-$dimensions can be written in a form making use of the Levi-Civita tensor defined by

$$\epsilon_{ij\ldots l} = \begin{cases} +1, & \text{if } ij\ldots l \text{ is an even permutation of } 12\ldots n \\ -1, & \text{if } ij\ldots l \text{ if an odd permutation of } 12\ldots n \\ 0, & \text{if any index is repeated.} \end{cases} \tag{B.3}$$

which simplifies the proof of basic properties of determinants.

It is easy to see for 2-dimensional matrices that the determinant can be written

$$\begin{aligned} \epsilon^{ij} A_{1i} A_{2j} &= \epsilon^{12} A_{11} A_{22} + \epsilon^{21} A_{12} A_{21} \\ &= \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \end{aligned} \tag{B.4}$$

Note that if we swap 1 and 2 around we get

$$\begin{aligned} \epsilon^{ij} A_{2i} A_{1j} &= \epsilon^{ij} A_{1j} A_{2i} \\ &= -\epsilon^{ji} A_{1j} A_{2i} \\ &= -\epsilon^{ij} A_{1i} A_{2j} \\ &= -\det(A). \end{aligned} \tag{B.5}$$

So we can write

$$\begin{aligned} 2\det(A) &= \epsilon^{ij} A_{1i} A_{2j} - \epsilon^{ij} A_{2i} A_{1j} \\ &= \epsilon^{ij} (A_{1i} A_{2j} - A_{2i} A_{1j}) \\ &= \epsilon^{ij} \epsilon^{kl} A_{ki} A_{lj} \end{aligned} \tag{B.6}$$

The determinant of a 3-dimensional matrix can be written

$$\det|\mathbf{A}| = \epsilon^{ijk} A_{1i} A_{2j} A_{3k} \tag{B.7}$$

We can easily see how this agrees with the iterative definition (B.2) by writing

$$\begin{aligned}\epsilon^{ijk}A_{1i}A_{2j}A_{3k} &= A_{11}\epsilon^{1jk}A_{2j}A_{3k} + A_{12}\epsilon^{2jk}A_{2j}A_{3k} + A_{13}\epsilon^{3jk}A_{2j}A_{3k} \\ &= A_{11}\epsilon^{1jk}A_{2j}A_{3k} - A_{12}\epsilon^{j2k}A_{2j}A_{3k} + A_{13}\epsilon^{jk3}A_{2j}A_{3k}. \end{aligned} \tag{B.8}$$

From the first term $\epsilon^{1jk}A_{2j}A_{3k}$; obviously elements from the first row of the matrix A are not included in this sum, and also because ϵ^{1jk} vanishes for $j = 1$ or $k = 1$ it doesn't include terms from the first column. Summation is taken over the remaining row and column indices subject to anti-symmetrisation giving the minor of A_{11}. Similarly argument holds for the other terms in (B.8).

This kind of argument generalises to $n-$dimensional determinants and we have

$$\det|\mathbf{A}| = \epsilon^{a_1 a_2 \ldots a_n} A_{1a_1} A_{2a_2} \ldots A_{na_n}. \tag{B.9}$$

We can also generalise (B.6) to $n-$dimensions. Consider 3-dimensions first

$$\det|\mathbf{A}| = +\epsilon^{ijk}A_{1i}A_{2j}A_{3k} = \epsilon^{123}\epsilon^{ijk}A_{1i}A_{2j}A_{3k} \tag{B.10}$$

$$\det|\mathbf{A}| = -\epsilon^{ijk}A_{1i}A_{2j}A_{3k} = \epsilon^{213}\epsilon^{ijk}A_{2i}A_{1j}A_{3k} \tag{B.11}$$

and so on for the other permutations of $1, 2, 3$. Since there are 3! ways of oredring $1, 2, 3$ we have

$$\det|\mathbf{A}| = \frac{1}{3!}\epsilon^{i'j'k'}\epsilon^{ijk}A_{i'i}A_{j'j}A_{k'k} \tag{B.12}$$

In $n-$dimensions

$$\det|\mathbf{A}| = \frac{1}{n!}\epsilon^{b_1 b_2 \ldots b_n}\epsilon^{a_1 a_2 \ldots a_n}A_{b_1 a_1}A_{b_2 a_2}\ldots A_{b_n a_n} \tag{B.13}$$

Proof of basic properties of determinants

This ϵ representation makes it easy to basic prove properties of determinates.

(a) Interchanging rows and columns results in the same determinant. This is the same as saying $\det(A^T) = \det(A)$ where A^T is the transpose of the matrix A. This follows easily follows from the equivalence of (B.13) with

$$\det|\mathbf{A}| = \epsilon^{b_1 b_2 \ldots b_n} A_{b_1 1} A_{b_2} \ldots A_{b_n n} \tag{B.14}$$

(similar to proving the equivalence of (B.9) and (B.13)) which is the same as

$$\epsilon^{b_1 b_2 \ldots b_n} (A^T)_{1b_1} (A^T)_{2b_2} \ldots (A^T)_{nb_n} = \det(A^T). \tag{B.15}$$

(b) Swapping columns introduces a minus sign. We need only prove that swapping any two adjacent columns introduces a minus sign, i.e.,

$$\det \begin{pmatrix} A_{11} & \ldots & A_{1k} & A_{1k+1} & \ldots & A_{1n} \\ A_{21} & \ldots & A_{2k} & A_{2k+1} & \ldots & A_{2n} \\ \vdots & \ldots & \vdots & \ldots & \ddots & \vdots \\ A_{n1} & \ldots & A_{nk} & A_{nk+1} & \ldots & A_{nn} \end{pmatrix} = -\det \begin{pmatrix} A_{11} & \ldots & A_{1k+1} & A_{1k} & \ldots & A_{1n} \\ A_{21} & \ldots & A_{2k+1} & A_{2k} & \ldots & A_{2n} \\ \vdots & \ldots & \ldots & \vdots & \ddots & \vdots \\ A_{n1} & \ldots & A_{nk+1} & A_{nk} & \ldots & A_{nn} \end{pmatrix} \tag{B.16}$$

$$\begin{aligned} \epsilon_{ijk} A_{i1} A_{j2} A_{k3} &= \epsilon_{ijk} A_{j2} A_{i1} A_{k3} \\ &= \epsilon_{jik} A_{i2} A_{j1} A_{k3} \\ &= -\epsilon_{ijk} A_{i2} A_{j1} A_{k3} \end{aligned} \tag{B.17}$$

For $n-$dimensional determinants,

$$\begin{aligned} \epsilon_{a_1 \ldots a_k a_{k+1} \ldots a_n} A_{a_1 1} \ldots A_{a_k k} A_{a_{k+1} k+1} \ldots A_{a_n n} &= \epsilon_{a_1 \ldots a_k a_{k+1} \ldots a_n} A_{a_1 1} \ldots A_{a_{k+1} k+1} A_{a_k k} \ldots A_{a_n n} \\ &= \epsilon_{a_1 \ldots a_{k+1} a_k \ldots a_n} A_{a_1 1} \ldots A_{a_k k+1} A_{a_{k+1} k} \ldots A_{a_n n} \\ &= -\epsilon_{a_1 \ldots a_k a_{k+1} \ldots a_n} A_{a_1 1} \ldots A_{a_k k+1} A_{a_{k+1} k} \ldots A_{a_n n} \end{aligned} \tag{B.18}$$

Swapping rows introduces a minus sign. To see this first write $\det(A) = \det(A^T)$, swap adjacent columns of the matrix A^T to give another matrix, lets call it $(A')^T$, then $\det(A^T) = -\det((A')^T)$. Taking the transpose of $(A')^T$ gives a matrix A', this differs from A by the exchange of adjacent rows. We have altogether $\det(A) = -\det(A')$.

(c) Simply consequence of property (b) is that if two columns are the same the determinant vanishes. Similarly if two rows are the same then the vanishes.

(d) We have

$$
\det \begin{pmatrix} A_{11} & \dots & A_{1k} + B_{1k} & \dots & A_{1n} \\ A_{21} & \dots & A_{2k} + B_{2k} & \dots & A_{2n} \\ \vdots & \ddots & \vdots & \ddots & \\ A_{n1} & \dots & A_{nk} + B_{nk} & \dots & A_{n} \end{pmatrix} = \det \begin{pmatrix} A_{11} & \dots & A_{1k} & \dots & A_{1n} \\ A_{21} & \dots & A_{2k} & \dots & A_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ A_{N1} & \dots & A_{nk} & \dots & A_{nn} \end{pmatrix} + \det \begin{pmatrix} B_{11} & \dots & A_{1k} & \dots & A_{1n} \\ B_{21} & \dots & A_{2k} & \dots & A_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{n1} & \dots & A_{nk} & \dots & A_{nn} \end{pmatrix} \tag{B.19}
$$

which are easily seen from ϵ

$$
\epsilon_{a_1 \dots a_k \dots a_n} A_{a_1 1} \cdots (A_{a_k k} + B_{a_k k}) \cdots A_{a_n n} = \epsilon_{a_1 \dots a_k \dots a_n} A_{a_1 1} \cdots A_{a_k k} \cdots A_{a_n n} + \epsilon_{a_1 \dots a_k \dots a_n} A_{a_1 1} \cdots + B_{a_k k} \cdots A_{a_n n}. \tag{B.20}
$$

A similar result holds for rows and can be esablished by considering the determinant of the transpose of the matrix and using the analogous result for columns just proved, and then apply $\det(A^T) = \det(A)$ on the resulting two determinants.

(e) Multiplication of a column by a constant k

$$
\det \begin{pmatrix} A_{11} & \dots & kA_{1k} & \dots & A_{13} \\ A_{21} & \dots & kA_{2k} & \dots & A_{23} \\ \vdots & \dots & \vdots & \ddots & \vdots \\ A_{n1} & \dots & kA_{nk} & \dots & A_{n} \end{pmatrix} = k \det \begin{pmatrix} A_{11} & \dots & A_{1k} & \dots & A_{13} \\ A_{21} & \dots & A_{2k} & \dots & A_{23} \\ \vdots & \dots & \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nk} & \dots & A_{n} \end{pmatrix} \tag{B.21}
$$

follows from

$$
\epsilon_{a_1 \dots a_k \dots a_n} A_{a_1 1} \cdots (kA_{a_k k}) \cdots A_{a_n n} = k \epsilon_{a_1 \dots a_k \dots a_n} A_{1a_1} A_{2a_2} A_{na_n} \tag{B.22}
$$

Multiplication of a row by a constant k

$$
\epsilon_{a_1 a_2 \dots a_n} (kA_{1a_1}) A_{2a_2} A_{na_n} = k \epsilon_{a_1 a_2 \dots a_n} A_{1a_1} A_{2a_2} A_{na_n} \tag{B.23}
$$

In particular $\det(kA) = k^n \det(A)$.

(f) The determinant of the product of two matrices A and B is equal to the product of the determinant of A and the determinant of B, that is, $\det(AB) = \det(A)\det(B)$.

$$\begin{aligned}
\det|\mathbf{A}| &= \frac{1}{n!}\epsilon^{b_1b_2\ldots b_n}\epsilon^{a_1a_2\ldots a_n}(A_{b_1c_1}B_{c_1a_1})(A_{b_2c_2}B_{c_2a_2})\ldots(A_{b_nc_n}B_{c_3a_3}) \\
&= \frac{1}{n!}\sum_{perms\ of\ c}(\epsilon^{b_1b_2\ldots b_n}A_{b_1c_1}A_{b_2c_2}\ldots A_{b_nc_n})(\epsilon^{a_1a_2\ldots a_n}B_{c_1a_1}B_{c_2a_2}\ldots B_{c_3a_3}) \\
&= \frac{1}{n!}\sum_{perms\ of\ c}\det(A^T)\det(B) \\
&= \det(A)\det(B). \qquad (B.24)
\end{aligned}$$

B.0.3 Inverse Matrix

Demonstarte ideas with 3-dimensionals marices. Recal the determinant can be expanded in the frist terms of co-factors.

We show that the elements of middle row being taken in order from left to right with alternate signs and multiplied by the determinants of 2-diemnsional matrices which remain when the row and column through the element are deleted, except here the first term starts with a minus sign.

$$\begin{aligned}
\det|\mathbf{A}| &= \epsilon^{ijk}A_{1i}A_{2j}A_{3k} \\
&= A_{21}\epsilon^{i1k}A_{1i}A_{3k} + A_{22}\epsilon^{i2k}A_{1i}A_{3k} + A_{23}\epsilon^{i3k}A_{1i}A_{3k} \\
&= -A_{21}\epsilon^{1ik}A_{1i}A_{3k} + A_{22}\epsilon^{i2k}A_{1i}A_{3k} + A_{23}\epsilon^{ik3}A_{1i}A_{3k} \qquad (B.25)
\end{aligned}$$

or

$$\det A = -A_{21}\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} + A_{22}\begin{vmatrix} A_{11} & A_{31} \\ A_{13} & A_{33} \end{vmatrix} - A_{23}\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} \qquad (B.26)$$

The determinant

$$-\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix}$$

is called the cofactor of the matrix element A_{21}.

To find the inverse.

For a square matrix $\mathbf{A}$, form the matrix whose elements are the cofactors of $\mathbf{A}$, i.e.

$$\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{pmatrix} \tag{B.27}$$

Consider the product of the matrix $\mathbf{A}$ with the matrix $\mathbf{C}^T$.

$$\mathbf{AC}^T = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{pmatrix} \tag{B.28}$$

The product of the first column of $\mathbf{C}$ and the first row of $\mathbf{A}$ is

$$A_{11}C_{11} + A_{12}C_{12} + A_{13}C_{13} = \det \mathbf{A}.$$

Generally, the product of any row of $\mathbf{A}$ with its cofactors is equal to $\mathbf{C}$,

$$\det A = \sum_k A_{ik}C_{ik} \quad \text{no summation over } i. \tag{B.29}$$

Other products give zero, for example,

$$\begin{aligned} A_{11}C_{21} + A_{12}C_{22} + A_{13}C_{23} &= A_{11}\begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} - A_{12}\begin{vmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{vmatrix} + A_{13}\begin{vmatrix} A_{11} & A_{12} \\ A_{31} & A_{32} \end{vmatrix} \\ &= \begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{11} & A_{12} & A_{13} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} \\ &= 0 \end{aligned} \tag{B.30}$$

We can see general why the other products are zero; if we replace A_{ik} in the sum in (B.29) by A_{jk}, with $j \neq i$, it gives a new deterninant which can be obtained from the determinant of A by replacing its $i-th$ row with the row $(A_{j1}, A_{j2}, A_{j3},)$. Then as this row will be repeated twice and the resulting derterminant vanishes. In general we have

$$\sum_k A_{ik}C_{jk} = A_{i1}C_{j1} + A_{i2}C_{j2} + A_{i3}C_{j3} = \delta_{ij} \det A \tag{B.31}$$

Or

$$\mathbf{A}\mathbf{C}^T = \mathbb{1} \det \mathbf{A}. \tag{B.32}$$

Same with columns.

$$\begin{aligned} \det|\mathbf{A}| &= \epsilon^{ijk} A_{i1} A_{j2} A_{k3} \\ &= A_{11}\epsilon^{1jk} A_{j2} A_{k3} + A_{21}\epsilon^{2jk} A_{j2} A_{k3} + A_{31}\epsilon^{3jk} A_{j2} A_{k3} \\ &= A_{11}\epsilon^{1jk} A_{j2} A_{k3} - A_{21}\epsilon^{j2k} A_{j2} A_{k3} + A_{31}\epsilon^{jk3} A_{j2} A_{k3} \end{aligned} \tag{B.33}$$

$$\det A = A_{11} \begin{vmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{vmatrix} - A_{21} \begin{vmatrix} A_{12} & A_{13} \\ A_{32} & A_{33} \end{vmatrix} + A_{31} \begin{vmatrix} A_{12} & A_{13} \\ A_{22} & A_{23} \end{vmatrix} \tag{B.34}$$

$$\begin{aligned} \det|\mathbf{A}| &= \epsilon^{ijk} A_{i1} A_{j2} A_{k3} \\ &= \epsilon^{ijk} A_{j2} A_{i1} A_{k3} \\ &= A_{12}\epsilon^{i1k} A_{i1} A_{k3} + A_{22}\epsilon^{i2k} A_{i1} A_{k3} + A_{32}\epsilon^{i3k} A_{i1} A_{k3} \\ &= -A_{12}\epsilon^{1ik} A_{i1} A_{k3} + A_{22}\epsilon^{i2k} A_{i1} A_{k3} - A_{32}\epsilon^{ik3} A_{i1} A_{k3} \end{aligned} \tag{B.35}$$

Consider the product of the matrix $\mathbf{A}$ with the matrix $\mathbf{C}^T$.

$$\mathbf{C}^T\mathbf{A} = \begin{pmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \tag{B.36}$$

$$\mathbf{C}^T\mathbf{A} = \mathbb{1} \det \mathbf{A}. \tag{B.37}$$

Therefore from this and (B.32) we have

$$\mathbf{A}^{-1} = \frac{\mathbf{C}^T}{\det \mathbf{A}}. \tag{B.38}$$

These considerations are easily extends to matrices of any size, and we have that

$$\mathbf{C}^T A = A\mathbf{C}^T = \det \mathbf{A}\mathbb{1} \tag{B.39}$$

holds generally.

This easily extends to higher dimensional square matrices. For any square matrix $\mathbf{A}$, the inverse matrix, $\mathbf{A}^{-1}$, exists if and only if $|\mathbf{A}| \neq 0$ and in this case

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathbf{C}^{\mathbf{T}}. \tag{B.40}$$

Or in component form

$$(\mathbf{A}^{-1})_{ij} = \frac{1}{\det \mathbf{A}}(\mathbf{C})_{ji}. \tag{B.41}$$

Appendix C

Gaussian Elimination

Gaussian Elimination is an algorithm for solving systems of linear equations. It has a huge advantage over most other methods, in that it can be used where the number of equations and the number of unknowns are not equal. The method can be used to calculate the determinant of a square matrix, and calculate the inverse of an invertible square matrix.

There are thre kinds of row operation:

1) Interchanging two rows
2) Multiplying a row by a non-zero number
3) Adding a multiple of one row to another row.

C.1 Linear equations

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{C.1}$$

1) Form the $m \times (n+1)$ augmented matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} & \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} & b_m \end{pmatrix} \tag{C.2}$$

2) Set $j = 1$

3) For each $i > j$, replace row r_i with

$$r_i - \frac{a_{i1}}{a_{11}} r_1$$

or

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ 0 & a_{22} - \frac{a_{21}}{a_{11}} a_{12} & a_{23} - \frac{a_{21}}{a_{11}} a_{13} & \dots & a_{2n} - \frac{a_{21}}{a_{11}} a_{1n} & b_2 - \frac{a_{21}}{a_{11}} b_1 \\ 0 & a_{32} - \frac{a_{31}}{a_{11}} a_{12} & a_{33} - \frac{a_{31}}{a_{11}} a_{13} & \dots & a_{3n} - \frac{a_{31}}{a_{11}} a_{1n} & b_3 - \frac{a_{31}}{a_{11}} b_1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2} - \frac{a_{m1}}{a_{11}} a_{12} & a_{m3} - \frac{a_{m1}}{a_{11}} a_{13} & \dots & a_{mn} - \frac{a_{m1}}{a_{11}} a_{1n} & b_m - \frac{a_{m1}}{a_{11}} b_1. \end{pmatrix} \tag{C.3}$$

Which we can then write as

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ 0 & \tilde{a}_{22} & \tilde{a}_{23} & \dots & \tilde{a}_{2n} & \tilde{b}_2 \\ 0 & \tilde{a}_{32} & \tilde{a}_{33} & \dots & \tilde{a}_{3n} & \tilde{b}_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \tilde{a}_{m2} & \tilde{a}_{m3} & \dots & \tilde{a}_{mn} & \tilde{b}_m \end{pmatrix} \tag{C.4}$$

4) For each $i > 2$, replace row r_i with

$$r_i - \frac{\tilde{a}_{i2}}{\tilde{a}_{22}} r_j.$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ 0 & \tilde{a}_{22} & \tilde{a}_{23} & \dots & \tilde{a}_{2n} & \tilde{b}_2 \\ 0 & 0 & \tilde{a}_{33} - \frac{\tilde{a}_{32}}{\tilde{a}_{22}} \tilde{a}_{23} & \dots & \tilde{a}_{3n} - \frac{\tilde{a}_{32}}{\tilde{a}_{22}} \tilde{a}_{2n} & \tilde{b}_3 - \frac{\tilde{a}_{32}}{\tilde{a}_{22}} \tilde{b}_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & \tilde{a}_{m3} - \frac{\tilde{a}_{m2}}{\tilde{a}_{22}} \tilde{a}_{23} & \dots & \tilde{a}_{mn} - \frac{\tilde{a}_{m2}}{\tilde{a}_{22}} \tilde{a}_{2n} & \tilde{b}_m - \frac{\tilde{a}_{m2}}{\tilde{a}_{22}} \tilde{b}_2 \end{pmatrix} \tag{C.5}$$

5) For each $i > j$, replace row r_i with

$$r_i - \frac{a_{ij}}{a_{jj}} r_j.$$

6) Solve by back-substitution.

The process hits a hitch if, at any stage, one of the diagonal elements a_{jj} is zero. When this happens, this can usually be dealt with by swapping row r_j with one of the rows below.

C.2 Computing Rank and Basis

Gaussian elimination can be applied to any $m \times n$ matrix A. In this way a matrix can be put in the transformed into a matrix that has a row echelon form like

$$\begin{pmatrix} a & * & * & * & * & * & * & * & * & * \\ 0 & 0 & b & * & * & * & * & * & * & * \\ 0 & 0 & 0 & c & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & e \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{C.6}$$

Example:

Controlled Gaussian elimination.

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 2x_4 &= 1 \\ x_1 - x_2 + 2x_3 &= -1 \\ x_1 + 2x_2 + 3x_3 + 2x_4 &= 1 \\ 2x_1 - 2x_2 + 4x_3 &= -2 \end{aligned}$$

The augmented matrix is:

$$\begin{pmatrix} 1 & 2 & 3 & 2 & \vdots & 1 \\ 1 & -1 & 2 & 0 & \vdots & -1 \\ 1 & 2 & 3 & 2 & \vdots & 1 \\ 2 & -2 & 4 & 0 & \vdots & -2 \end{pmatrix} \tag{C.7}$$

By doing $r_2 \mapsto r_2 - (1/2)r_4$ and $r_3 \mapsto r_3 - r_1$ we obtain

$$
\begin{pmatrix}
1 & 2 & 3 & 2 & \vdots & 1 \\
0 & 0 & 0 & 0 & \vdots & 0 \\
0 & 0 & 0 & 0 & \vdots & 0 \\
2 & -2 & 4 & 0 & \vdots & -2
\end{pmatrix}
\tag{C.8}
$$

we eliminate x_1 in row r_1; $r_1 \mapsto r_1 - (1/2)r_4$

$$
\begin{pmatrix}
0 & 3 & 1 & 2 & \vdots & 2 \\
0 & 0 & 0 & 0 & \vdots & 0 \\
0 & 0 & 0 & 0 & \vdots & 0 \\
2 & -2 & 4 & 0 & \vdots & -2
\end{pmatrix}
\tag{C.9}
$$

$$
\begin{aligned}
3x_2 + x_3 + 2x_4 &= 2 \\
2x_1 - 2x_2 + 4x_3 &= -2
\end{aligned}
$$

Write $x_2 = 2\kappa_1$ and $x_3 = -2\kappa_2$ where κ_1 and κ_2 are arbitrary, then

$$
\begin{aligned}
6\kappa_1 - 2\kappa_2 + 2x_4 &= 2 \\
2x_1 - 4\kappa_1 - 8\kappa_2 &= -2
\end{aligned}
$$

or

$$
\begin{aligned}
x_1 &= -1 + 2\kappa_1 + 4\kappa_2 \\
x_2 &= 2\kappa_1 \\
x_3 &= -2\kappa_2 \\
x_4 &= -1 - 3\kappa_1 + 4\kappa_2.
\end{aligned}
$$

□

C.3 Finding the determinant

The determinant of a matrix in echelon form is simply found by taking the product of diagonal entries. The following examples illustrates this,

$$
\begin{aligned}
\det A &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix} \\
&= a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} 0 & a_{23} \\ 0 & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} 0 & a_{22} \\ 0 & 0 \end{pmatrix} \\
&= a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{pmatrix} \\
&= a_{11}a_{22}a_{33}
\end{aligned} \tag{C.10}
$$

where we have used that when a column of the matrix are all zeros then the determinant is zero. For the 4×4 case we have

$$
\begin{aligned}
\det A &= \det \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{pmatrix} \\
&= a_{11} \det \begin{pmatrix} a_{22} & a_{23} & a_{24} \\ 0 & a_{33} & a_{34} \\ 0 & 0 & a_{44} \end{pmatrix} - a_{12} \begin{pmatrix} 0 & a_{23} & a_{24} \\ 0 & a_{33} & a_{34} \\ 0 & 0 & a_{44} \end{pmatrix} \\
&\quad + a_{13} \det \begin{pmatrix} 0 & a_{22} & a_{24} \\ 0 & 0 & a_{34} \\ 0 & 0 & a_{44} \end{pmatrix} - a_{14} \det \begin{pmatrix} 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \\ 0 & 0 & 0 \end{pmatrix} \\
&= a_{11} \det \begin{pmatrix} a_{22} & a_{23} & a_{24} \\ 0 & a_{33} & a_{34} \\ 0 & 0 & a_{44} \end{pmatrix} \\
&= a_{11}a_{22}a_{33}a_{44}
\end{aligned}
$$

where we used (C.10). It is easy to show that

$$
\begin{aligned}
\det A &= \det \begin{pmatrix} a_{11} & a_{12} & \dots & \dots & \dots & a_{1n} \\ 0 & a_{22} & \dots & \dots & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ddots & \dots & \dots & a_{n-1n-1} & a_{n-1n} \\ 0 & 0 & \dots & \dots & 0 & a_{nn} \end{pmatrix} \\
&= a_{11}a_{22}a_{33}\dots a_{nn}
\end{aligned} \tag{C.11}
$$

When calulating the determinant of the original matrix from the determinant of the row echelon reduced matrix you need to take into account how following operations (proved in the appendix A) change the determinant,

1) Interchanging two rows multiplies the determinant by -1
2) Multiplying a row by a non-zero number multiplies the determinant by that number
3) Adding a multiple of one row to another row does not change the determinant.

Example:

Find

$$\det \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & \alpha \\ 1 & 0 & 2 \end{pmatrix}$$

$r_2 \mapsto r_2 - r_1$ and $r_3 \mapsto r_3 - r_1$

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & \alpha - 1 \\ 0 & -1 & 1 \end{pmatrix}$$

$r_3 \mapsto r_3 - \frac{1}{2}r_2$

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & \alpha - 1 \\ 0 & 0 & 1 - \frac{1}{2}(\alpha - 1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & \alpha - 1 \\ 0 & 0 & -\dfrac{1}{2}(\alpha - 3) \end{pmatrix}.$$

As we have only added a multiple of one row to another row, we have

$$\det \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & \alpha \\ 1 & 0 & 2 \end{pmatrix} = \det \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & \alpha - 1 \\ 0 & 0 & -\dfrac{1}{2}(\alpha - 3) \end{pmatrix} = \alpha - 3.$$

□

C.4 Finding the inverse matrix

In this section we describe a method for finding the inverse matrix of an invertible $n \times n$ matrix. The key is the connection between elementary row operations and matrix multiplication.

C.4.1 Elementary row operations as matrix multiplication

First we demostrate how all elementary row operatios can be achieved by matrix multiplication from the left in the simple case of 3×3 matrices (it is easily generalised to matrices of arbitrary rank).

First, multiplication from the left by

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

exchanges rows 1 and 2:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{C.12}$$

Multiplication from the left by

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

exchanges rows 1 and 3 and multiplication from the left by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

exchanges rows 2 and 3

Multiplication from the left by

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

multiplies

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \lambda a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{C.13}$$

Similarly

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad and \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

multiply the second and third row by λ respectively.

Multiplication from the left by

$$\begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

adds a multiple of row 1 to row 2:

$$\begin{pmatrix} 1 & 0 & 0 \\ \alpha & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} + \alpha a_{11} & a_{22} + \alpha a_{12} & a_{23} + \alpha a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{C.14}$$

Similarly

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{pmatrix} \quad \text{adds a multiple of row 1 to row 3,}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha & 1 \end{pmatrix} \quad \text{adds a multiple of row 2 to row 3,}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix} \quad \text{adds a multiple of row 3 to row 2,}$$

$$\begin{pmatrix} 1 & 0 & \alpha \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{adds a multiple of row 3 to row 1,}$$

$$\begin{pmatrix} 1 & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{adds a multiple of row 2 to row 1.}$$

Completeting all elementary row operations for 3×3 matrices.

C.4.2 Inverting the matrix

We start with the augmented matrix formed from the matrix A and the identity matrix I,

$$[A|I] = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & \vdots & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & \vdots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & \vdots & 0 & 0 & \dots & 1 \end{pmatrix} \tag{C.15}$$

Each time we act on A by a matrix (which effects an elementary row operation), we simultaneously act with it on I of the augmented matrix. The series of matrices which puts A into the form of the identity matrix will be the inverse matrix, which is read off from their cummulative action on the I of the augmented matrix.

So we use gaussian elimination to reduce $[A|I]$. If A reduces to I then $[A|I]$ reduces $[I|A^{-1}]$.

Therefore, to find the inverse matrix we perform a series of elementary row operations on A and I which put A into the form of the identity matrix, we can then just read off the inverse matrix from their effect on I of the augmented matrix.

Example:

Find the inverse matrix of

$$N = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 & \sqrt{2} \end{pmatrix}.$$

Form the augmented matrix

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 & \vdots & 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \vdots & 0 & 1 & 0 \\ 0 & 0 & \sqrt{2} & \vdots & 0 & 0 & 1 \end{pmatrix}$$

$r_3 \mapsto (1/2)r_3$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 & \vdots & 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & \vdots & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \vdots & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

$r_2 \mapsto r_2 + r_3$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 & \vdots & 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & 0 & \vdots & 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \vdots & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

$r_2 \mapsto (2/3)r_2$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{\sqrt{2}}{\sqrt{3}} & 0 & \vdots & 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & 0 & \vdots & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \vdots & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

$r_1 \mapsto r_1 + r_2$

$$\begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 & \vdots & 1 & \frac{2}{3} & \frac{1}{3} \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & 0 & \vdots & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{\sqrt{2}} & \vdots & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Finally, $r_1 \mapsto \sqrt{3}r_1$, $r_2 \mapsto (\sqrt{3}/\sqrt{2})$, and $r_3 \mapsto \sqrt{2}r_3$, so that

$$\begin{pmatrix} 1 & 0 & 0 & \vdots & \sqrt{3} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & 1 & 0 & \vdots & 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & 1 & \vdots & 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

and so

$$N^{-1} = \begin{pmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

□

Appendix D

Spectral Theorem for Hermitian and Unitary Operators

In section 5.1.2 we proved that when two eigenvectors have different eigenvalues that the eigenvector are orthogonal. Here we prove that even in the case when there are degenerate eigenvalues that the set of eigenvectors are all mutually orthogonal.

First we explain why every operator, $\hat{A}$, has at least one eigenvector, i.e, there exists an $\mathbf{x}$ such that

$$\hat{A}(\mathbf{x}) = o^{(m)}\mathbf{x} \tag{D.1}$$

Write $\mathbf{x} = x_i\mathbf{v_i}$, then

$$\begin{aligned}
\hat{A}(\mathbf{x}) &= \hat{A}(\sum_j x_j\mathbf{v_j}) \\
&= \sum_j x_j\hat{A}(\mathbf{v_j}) \\
&= \sum_{i,j} x_j A_{ij}\mathbf{v_i} \\
&= o^{(m)}\mathbf{x} \\
&= o^{(m)}\sum_i x_i\mathbf{v_i}.
\end{aligned} \tag{D.2}$$

From which we see that the operator $\hat{A}$ having an eigenvector $\mathbf{x}$ is equivalent to the matrix A having an eigenvector $\vec{x}$. We now explain why every matrix has at least one eigenvector. Let A be an $n \times n$ matrix. The matrix $A - \lambda\mathbb{1}$ is not invertible if and only if $\det(A - \lambda\mathbb{1}) = 0$. The fundamental theorem of algebra guarantees that this polynomial in λ has at least one root. In fact, it then easily follows that there are at most n distinct roots. For each distinct root λ the equation $(A - \lambda\mathbb{1})\vec{x} = 0$, and thus $A\vec{x} = \lambda\vec{x}$, has at least one solution $\vec{x}$. To understand this, think of each row of the matrix $M \equiv A - \lambda\mathbb{1}$ as a vector then we can write

$$M\vec{x} = \begin{pmatrix} \sum_{j=1}^{n} m_{1j}x_j \\ \sum_{j=1}^{n} m_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} m_{nj}x_j \end{pmatrix} \tag{D.3}$$

As $\det M = 0$, the vectors $\vec{m}_i$ (where $1 \leq i \leq n$) are linearly dependent and so span a vector subspace of dimension less than n. Any $\vec{x}$ orthogonal to this subspace will satisfy $M\vec{x} = 0$.

D.1 Hermitian Operators

Let $\hat{H}$ be a Hertitian linear operator on a finite complex inner product space V of dimension n. We define $\hat{H}$-invariant subspaces of V as linear subspaces W such that $\hat{H}W \subseteq W$. We denote the restriction of the operator $\hat{H}$ to the subspace W as $\hat{H}_W$. Then $\hat{H}_W : W \to W$ is a linear operator on W, and $\hat{H}_W$ is also Hermitian. If $\dim W \geq 1$, $\hat{H}_W$ has at least one eigenvector $w \in W$ because any linear operator at all acting on a (non-empty) finite dimensional complex vector space has at least one eigenvector.

The proof goes as follows: Let W be an invariant subspace with respect to $\hat{H}$. Then $W^\perp$ is also an invariant subspace with respect to $\hat{H}$. The reason for this is that if $w \in W$ and $x \in W^\perp$, then

$$(\mathbf{w}, \hat{H}(\mathbf{x})) = (\hat{H}^\dagger(\mathbf{w}), \mathbf{x}) = (\hat{H}(\mathbf{w}), \mathbf{x}) = 0, \tag{D.4}$$

as $\hat{H}(\mathbf{w}) \in W$ and $\mathbf{x} \in W^\perp$. Thus $H(\mathbf{x}) \in W^\perp$.

Write $V = V_1$. Take one eigenvector $\mathbf{v_1}$. Then $\mathbb{C}\mathbf{v_1}$ is $\hat{H}$-invariant. Hence, $V_2 := (\mathbb{C}\mathbf{v_1})^\perp$ is also $\hat{H}$-invariant. Now apply the same argument to V_2: the restriction of $\hat{H}$ to V_2 has an eigenvector and the perpendicular complement, denote it V_3, to $\mathbb{C}\mathbf{v_2}$ in V_2 is $\hat{H}$-invariant. Continuimg in this way, one obtains a sequence of mutually orthogonal eigenvectors and a decreasing sequence of invariant subspaces, $V = V_1 \supset V_2 \supset V_3 \supset \cdots$ such that V_k has dimension $n - k + 1$. The process will only stop when we get to V_n which has dimension 1.

D.2 Unitary Operators

Let W be an invariant subspace with respect to $\hat{U}$. Then $W^\perp$ is also an invariant subspace with respect to $\hat{U}$. The reason for this is that if $w \in W$ and $x \in W^\perp$, then

$$(\hat{U}(\mathbf{w}), \hat{U}(\mathbf{x})) = (\hat{U}^\dagger \hat{U}(\mathbf{w}), \mathbf{x}) = (\mathbf{w}, \mathbf{x}) = 0, \tag{D.5}$$

as $\mathbf{w} \in W$ and $\mathbf{x} \in W^\perp$. Thus $\hat{U}(\mathbf{x}) \in W^\perp$.

Following the same line of reasoning used in the Hermitician case, one gets a sequence of mutually orthogonal eigenvectors and a decreasing sequence of invariant subspaces, $V = V_1 \supset V_2 \supset V_3 \supset \cdots$ such that V_k has dimension $n - k + 1$. The process will only stop when we get to V_n which has dimension 1.

The eigenvalues of a unitary operator are of the form $e^{i\varphi_m}$. To see this, say $\hat{U}$ is a unitary operator and $\mathbf{v^{(m)}}$ is an eigenvector then

$$(\mathbf{v^{(m)}}, \mathbf{v^{(m)}}) = (\hat{U}(\mathbf{v^{(m)}}), \hat{U}(\mathbf{v^{(m)}})) = (o^{(m)}\mathbf{v^{(m)}}, o^{(m)}\mathbf{v^{(m)}}) = o^{(m)*} o^{(m)} (\mathbf{v^{(m)}}, \mathbf{v^{(m)}}) \tag{D.6}$$

implies that the eigenvalue satisfies: $|o^{(m)}|^2 = 1$.

Appendix E

Proof of the Cayley-Hamilton Theorem

E.1 Proof for Diagonal Matrices

Let $\chi_A(\lambda) = \det(\lambda\mathbb{1} - A)$ be the characteristic polynomial of A. Then $\chi_A(A) = 0$.

First, prove the result for diagonal matrices. Let A be the diagonal matrix

$$A = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}. \tag{E.1}$$

Then $\chi_A(\lambda - \lambda_i) = \prod_{i=1}^{n}(\lambda - \lambda_i)$. We see that

$$\begin{aligned} \chi_A(A) &= \prod_{i=1}^{n}(A - \lambda_i\mathbb{1}) \\ &= \begin{pmatrix} 0 & & & \\ & \lambda_2 - \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_n - \lambda_1 \end{pmatrix} \begin{pmatrix} \lambda_1 - \lambda_2 & & & \\ & 0 & & \\ & & \ddots & \\ & & & \lambda_n - \lambda_2 \end{pmatrix} \cdots \begin{pmatrix} \lambda_1 - \lambda_n & & & \\ & \lambda_2 - \lambda_n & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 \cdot (\lambda_1 - \lambda_2) \cdots (\lambda_1 - \lambda_n) & & & \\ & (\lambda_2 - \lambda_1) \cdot 0 \cdots (\lambda_2 - \lambda_n) & & \\ & & \ddots & \\ & & & (\lambda_n - \lambda_1) \cdot (\lambda_n - \lambda_2) \cdots 0 \end{pmatrix} \\ &= 0. \end{aligned} \tag{E.2}$$

E.2 Proof for Diagonalizable Matrices

We use this to prove that $\chi_A(A) = 0$ for all diagonalizable matrices A. Write $S^{-1}AS = D$ for some diagonal matrix D. Then

$$\begin{aligned} \det(\lambda\mathbb{1} - A) &= \det(S^{-1})\det(\lambda\mathbb{1} - A)\det(S) \\ &= \det(S^{-1}(\lambda_i\mathbb{1} - A)\det(S)) \\ &= \det(\lambda\mathbb{1} - D) \end{aligned} \tag{E.3}$$

Hence $\chi_A(\lambda) = \chi_D(\lambda)$. Now

$$\chi_A(\lambda) = \chi_D(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0. \tag{E.4}$$

Then

$$\begin{aligned} \chi_A(A) &= A^n + a_{n-1}A^{n-1} + \cdots + a_0\mathbb{1} \\ &= SS^{-1}(A^n + a_{n-1}A^{n-1} + \cdots + a_0\mathbb{1})SS^{-1} \\ &= S\left((S^{-1}AS)^n + a_{n-1}S^{-1}AS)^{n-1} + \cdots + a_0\mathbb{1}\right)S^{-1} \\ &= S\chi_D(D)S^{-1} \\ &= 0. \end{aligned} \tag{E.5}$$

Hence, $\chi_A(A) = 0$.

E.2.1 Conditions for diagonalizability

(a) We prove that an $(n \times n)$ matrix A is diagonalizable if and only if $\mathbb{C}^n$ admits a basis which consists of eigenvectors of A. This will in particular prove that the Cayley-Hamilton theorem holds for hermitian and unitary matrices.

(b) Suppose all the eigenvalues of an $(n \times n)$ matrix A are distinct, prove that A is diagonalizable.

Proof:

(a) The "only if" part: Suppose A is diagonalizable. Then $S^{-1}AS = D$ for some diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \tag{E.6}$$

and some invertible matrix S. Let $\mathbf{v_i} \in \mathbb{C}^n$, $i = 1, \ldots, n$, be the columns of S, i.e.,

$$S = \left(\mathbf{v_1} \;\middle|\; \mathbf{v_2} \;\middle|\; \ldots \;\middle|\; \mathbf{v_n} \right) \tag{E.7}$$

Since S is invertible, $\{\mathbf{v_i}\}_{i=1}^n$ forms a basis. Moreover,

$$\left(A\mathbf{v_1} \;|\; A\mathbf{v_2} \;|\; \cdots \;|\; A\mathbf{v_n} \right) = AS = DS = \left(\lambda_1 \mathbf{v_1} \;|\; \lambda_2 \mathbf{v_2} \;|\; \ldots \;|\; \lambda_n \mathbf{v_n} \right) \tag{E.8}$$

Hence, we have $A\mathbf{v_i} = \lambda_i \mathbf{v_i}$. Thus the basis $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ consists of eigenvectors.

The "if part": Given a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{C}^n$ consisting of eiegvectors of A, define the matrix

$$S = \left(\mathbf{v_1} \;|\; \mathbf{v_2} \;|\; \cdots \;|\; \mathbf{v_n} \right) \tag{E.9}$$

then

$$\begin{aligned} S^{-1}AS &= \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{pmatrix} \left(\lambda_1 v_1 \;\middle|\; \lambda_2 v_2 \;\middle|\; \cdots \;\middle|\; \lambda_n v_n \right) \\ &= \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \end{aligned} \tag{E.10}$$

and so $S^{-1}AS$ is diagonal.

(b) Let $\{\lambda_i\}_{i=1}^n$ be the distinct eigenvalues and let $\{\mathbf{v_i}\}_{i=1}^n$ be the corresponding eigenvectors. We prove that $\{\mathbf{v_i}\}_{i=1}^n$ are linearly independent by induction.

The base case $\mathbf{v_1} \neq 0$. $\{\mathbf{v_1}\}$is linearly independent. We now perform the inductive step. Suppose $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ are linearly independent. Let

$$\sum_{i=1}^{k} c_i \mathbf{v_i} = 0. \tag{E.11}$$

Applying A, we get

$$\sum_{i=1}^{k} c_i \lambda_i \mathbf{v_i} = 0. \tag{E.12}$$

Multiply (E.11) by λ_k and subtract (E.12). We obtain

$$\sum_{i=1}^{k-1} c_i (\lambda_k - \lambda_i) \mathbf{v_i} = 0. \tag{E.13}$$

By the linear independence of $\{\mathbf{v_1}, \ldots, \mathbf{v_{k-1}}\}$, we have $c_i(\lambda_k - \lambda_i) = 0$ for all $i \in \{1, \ldots, k-1\}$. Since $\lambda_k \neq \lambda_i$ we have $c_i = 0$ for all $i \in \{1, \ldots, k-1\}$. This then implies also that $c_k = 0$. Hence, $\{\mathbf{v_1}, \ldots, \mathbf{v_{k-1}}\}$ are linearly independent.

It follows that $\{\mathbf{v_1}, \ldots, \mathbf{v_n}\}$ are linearly independent. It therefore forms a basis of $\mathbb{C}^n$. Hence, A is diagonalizable. This is a sufficient but not necessary criterion for diagonalizability.

E.3 Proof of Cayley-Hamilton Theorem for all Matrices

From appendix B we have for an arbitrary matrix A that

$$\operatorname{adj}(A) \cdot A = \det(A) \mathbb{1}_d = A \cdot \operatorname{adj}(A). \tag{E.14}$$

Consider the adguagte

$$B = \operatorname{adj}(x\mathbb{1}_d - M).$$

According to (E.14)

$$(x\mathbb{1}_d - M)B = \det(x\mathbb{1}_d - M)\mathbb{1}_d = P(x)\mathbb{1}_d. \tag{E.15}$$

Since B is a matrix with polynomials as entries, we can write

$$B = \sum_{i=0}^{d-1} x^i B_i \tag{E.16}$$

where B_i are d constant matrices.

Now we expand

$$\begin{aligned} P(x)\mathbb{1}_d &= (x\mathbb{1}_d - M)B \\ &= (x\mathbb{1}_d - M)\sum_{i=0}^{d-1} x^i B_i \\ &= \sum_{i=0}^{d-1} x^{i+1} B - \sum_{i=0}^{d-1} x^i M B_i \\ &= x^d B_{d-1} + \sum_{i=1}^{d-1} x^i (B_{i-1} - MB_i) - MB_0. \end{aligned} \tag{E.17}$$

Writing

$$P(x)\mathbb{1}_d = x^d \mathbb{1}_d + x^{d-1} c_{d-1}\mathbb{1}_d + \cdots + x c_1 \mathbb{1}_d + c_0 \mathbb{1}_d.$$

and comparing this to (E.17) we find

$$B_{d-1} = \mathbb{1}_d, \quad B_{i-1} - MB_i = c_i \mathbb{1}_d \quad \text{for } 1 \leq i \leq d-1, \quad -MB_0 = c_0 \mathbb{1}_d. \tag{E.18}$$

Substituting these expressions into

$$P(M) = M^d \mathbb{1}_d + M^{d-1} c_{d-1}\mathbb{1}_d + \cdots + M c_1 \mathbb{1}_d + c_0 \mathbb{1}_d$$

we obtain

$$
\begin{aligned}
P(M) &= M^d \mathbb{1}_d + M^{d-1} c_{d-1} \mathbb{1}_d + \cdots + M c_1 \mathbb{1}_d + c_0 \mathbb{1}_d \\
&= M^d B_{d-1} + \sum_{i=1}^{d-1} (M^i B_{i-1} - M^{i+1} B_i) - M B_0 \\
&= M^d B_{d-1} + (M B_0 - M^2 B_1 \\
&\qquad + M^2 B_1 - M^3 B_2 \\
&\qquad + M^3 B_2 - M^4 B_3 \\
&\qquad \vdots \\
&\qquad + M^{d-2} B_{d-3} - M^{d-1} B_{d-2} \\
&\qquad + M^{d-1} B_{d-2} - M^d B_{d-1}) - M B_0 \\
&= 0. \qquad \text{(E.19)}
\end{aligned}
$$

Therefore,

$$
P(M) = M^d + c_{d-1} M^{d-1} + \cdots + c_1 M + c_0 \mathbb{1}_d = 0. \qquad \text{(E.20)}
$$

□

Appendix F

Dividing polynomials, Bezout's identity, and partial fractions

F.1 Dividing polynomials

We divide the polynomial $p(x) = x^3 - 4x^2 + 2x - 3$ by $q(x) = x + 2$. We take the obvious steps:

(1a) $x^3 - 4x^2 + 2x - 3 = (x^3 + 2x^2) - 6x^2 + 2x - 3$
(1b) $x^3 - 4x^2 + 2x - 3 = x^2(x + 2) - 6x^2 + 2x - 3$
(2a) $-6x^2 + 2x - 3 = (-6x^2 - 12x) + 14x - 3$
(2b) $-6x^2 + 2x - 3 = (-6x)(x + 2) + 14x - 3$
(3a) $14x - 3 = 14x + 28 - 31$
(3b) $14x - 3 = 14(x + 2) - 31$

So that

$$\frac{x^3 - 4x^2 + 2x - 3}{x + 2} = x^2 - 6x + 14 - \frac{31}{x + 2}.$$

Euclid's algorithm for polynomials

Divide $p(x)$ by $m(x)$

$$p(x) = m(x)q(x) + r(x) \qquad (0 \leq \deg r < \deg m)$$

$\square$

F.2 Bezout's identity

(Bezout's identity). Let f, g be polynomials. Then there exists a greatest common divisor which is unique up to multiplication by constants. Moreover, for the greatest common divisor h, there exists polynomials p and q such that

$$pf + qg = h.$$

Proof:

Consider the set $S = \{pf + qg : p, q \text{ polynomials}\}$. Let $h = p_h f + q_h g$ be a non-zero polynomial in S with minimal degree. We claim that h divides any polynomial in S.

Suppose not. Then there exists some p_d and q_d such that

$$p_d f + q_d g = sh + r$$

for some polynomials s and r, $r \neq 0$ and $\deg(r) < \deg(h)$. But then

$$r = (p_d - p_h s)f + (q_d - p_d s)g$$

is a non-zero element in S with smaller degree than h, which is a contradiction.

We now claim that h is a greatest common divisor of f and g (note that h is a divisor of f because $f = 1 \cdot f + 0 \cdot g \in S$. Also, h is a divisor of g because $g = 0 \cdot f + 1 \cdot g \in S$). Suppose c is a common divisor of f and g, so that $f = cp_c$ and $g = cq_c$ for some polynomials p_c and q_c. Then

$$h = p_h f + q_h g = p_h p_c c + q_h q_c c = (p_h p_c + q_h q_c)c.$$

Hence, c divides h. In other words, any other common divisor of f and g must divide h. This shows that h is a greasest common divisor. The uniqueness of the greatest common divisor up to multiplicative constant follows from the fact that $\deg(c) = \deg(h)$ only if $\deg(p_h p_c + q_h q_c) = 0$. That is, if $p_h p_c + q_h q_c$ is a constant.

□

If two polynomials f and g have no roots in common then the greatest common divisor is a constant number. The polynomials f and g are said to be relatively prime.

F.3 Partial fractions

If f and g are polynomials such that $\deg f < \deg g$, and

$$g = g_1 g_2$$

where g_1 and g_2 are relatively prime polynomials, then there exist polynomials f_1 and f_2 such that

$$\frac{f}{g} = \frac{f_1}{g_1} + \frac{f_2}{g_2} \tag{F.1}$$

and

$$\deg f_1 < \deg g_1 \quad \text{and} \quad \deg f_2 < \deg g_2.$$

We prove this using Bezout's identity. According to Bezout's identity there exist polynomials c and d such that

$$cg_1 + dg_2 = 1$$

(by hypothesis, 1 is the greatest common divisor of g_1 and g_2).

Let $df = g_1 q + f_1$ with $\deg f_1 < \deg g_1$ be the Euclidean division of df by g_1. Setting $f_2 = cf + qg_2$, one gets

$$\begin{aligned} \frac{f}{g} &= \frac{f(cg_1 + dg_2)}{g} \\ &= \frac{df}{g_1} + \frac{cf}{g_2} \\ &= \frac{f_1 + g_1 q}{g_1} + \frac{f_2 - g_2 q}{g_2} \\ &= \frac{f_1}{g_1} + \frac{f_2}{g_2}. \end{aligned} \tag{F.2}$$

It remains to show $\deg f_2 < \deg g_2$. Multiplying both sides of (F.1) by $g_1 g_2$ gives $f = f_2 g_1 + f_1 g_2$, and thus

$$\begin{aligned}
\deg f_2 &= \deg(f - f_1 g_2) - \deg g_1 \\
&\leq \max(\deg f, \deg(f_1 g_2)) - \deg g_1 \\
&< \max(\deg g, \deg(g_1 g_2)) - \deg g_1 \\
&= \deg g_2.
\end{aligned}$$

This is easily generalised by iteration: If f and g are polynomials such that $\deg f < \deg g$, and

$$g = g_1 \cdots g_n \tag{F.3}$$

where every pair of polynomials from the list $g_1, \ldots, g_n$ are relatively prime, then there exist polynomials $f_1, \ldots f_n$ such that

$$\frac{f}{g} = \frac{f_1}{g_1} + \cdots + \frac{f_n}{g_n} \tag{F.4}$$

and

$$\deg f_j < \deg g_j \quad \text{for } j = 1, \ldots, n.$$

Appendix G

Jordan Normal Form

G.1 The main theorem

Let A be a $n \times n$ matrix. There exists an invertible matrix S such that

$$S^{-1}AS = J \tag{G.1}$$

where J is a block matrix,

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_m \end{pmatrix} \tag{G.2}$$

and each block J_i is a square matrix of the form

$$J_j = \lambda_j \mathbb{1} + N = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix} \tag{G.3}$$

where λ is an eigenvalue of A.

In the proof it is convenient to work with linear operators on vector spaces rather their matrix representations in some basis. Throughout the rest of the notes we shall therefore assume that V is an $n-$dimensional complex vector space and that A is a linear operator on V.

The kernel (or null space) of A is defined by

$$\ker A = \{x \in V : A\mathbf{x} = 0\} \tag{G.4}$$

and the range

$$\text{range } A = \{A\mathbf{x} :\in V\} \tag{G.5}$$

The kernel and the range are both linear subspaces of V and the dimension theorem says that

$$\dim \ker A + \dim \text{range } A = n \tag{G.6}$$

The subspace $\ker(A - \lambda \mathbb{1})$ of V, that is, the subspace spanned by the eigenvectors belonging to λ, is called the eigenspace corresponding to λ. The number $\dim \ker(A - \lambda \mathbb{1})$ is called the geometric multiplicity of λ. Recall that $\lambda \in \mathbb{C}$ is an eigenvalue if and only if it is a root of the characteristic polynomial

$$p(z) = (-1)^n (z - \lambda_1)^{a_1} (z - \lambda_2)^{a_2} \cdots (z - \lambda_1)^{a_k} \tag{G.7}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of A. The positive integer a_j is called the algebraic multiplicity of the eigenvalue λ_j. The corresponding geometric multiplicity will be denoted g_j.

G.2 Decomposition into invariant subspaces

We begin with some definitions. Let $V_1, \ldots, V_k$ be subspaces of V. We say that V is the direct sum of $V_1, \ldots, V_k$ if each vector $\mathbf{x} \in V$ can be written in a unique way as

$$\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_k, \quad \text{where} \quad \mathbf{x}_j \in V_j,\ j = 1, \ldots, k. \tag{G.8}$$

We use the notation

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k. \tag{G.9}$$

We say that a subspace W of V is invariant under A if

$$\mathbf{x} \in W \quad \text{implies} \quad A\mathbf{x} \in W. \tag{G.10}$$

Example: Suppose that A has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ with corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$. It then follows that the vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ are linearly independent and thus form a basis for V. Let

$$\ker(A - \lambda_k \mathbb{1}) = \{z\mathbf{u}_k : z \in \mathbb{C}\}, \quad k = 1, \ldots, n,$$

be the corresponding eigenspaces. Each eigenspace is invariant under A since

$$(A - \lambda_k \mathbb{1})\mathbf{u} = \mathbf{0} \quad \text{implies} \quad (A - \lambda_k \mathbb{1})A\mathbf{u} = A(A - \lambda_k \mathbb{1})\mathbf{u} = \mathbf{0}.$$

Moreover,

$$V = (A - \lambda_1 \mathbb{1}) \oplus (A - \lambda_2 \mathbb{1}) \oplus \cdots \oplus (A - \lambda_n \mathbb{1})$$

by the definition of a basis.

□

More generally, suppose that A has k distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ and that the geometric multiplicity g_j of each λ_j equals the algebraic multiplicity a_j. Let $\ker(A - \lambda_j \mathbb{1})$, $j = 1, \ldots, k$, be the corresponding eigenspaces. We can then find a basis for each eigenspace consisting of g_j eigenvectors. The union of these bases consists of $g_1 + \cdots + g_k = a_1 + \cdots + a_k = n$ elements and is linearly independent, since eigenvectors belonging to different eigenvalues are linearly independent. We thus obtain a basis for V and it follows that

$$V = (A - \lambda_1 \mathbb{1}) \oplus (A - \lambda_2 \mathbb{1}) \oplus \cdots \oplus (A - \lambda_k \mathbb{1}).$$

In this basis, A has the matrix

$$D = \begin{pmatrix} \lambda_1 \mathbb{1}_1 & & \\ & \ddots & \\ & & \lambda_k \mathbb{1}_k \end{pmatrix}$$

where $\mathbb{1}_j$ is the $g_j \times g_j$ unit matrix. In other words, D is a diagonal matrix with the eigenvalues on the diagonal, each repeated g_j times. One says that A is diagonalised in the new basis.

□

Not all matrices can be diagonalised.

G.3 Minimal polynomial

There exists a non-zero polynomial p such that $p(\mathbf{A}) = \mathbf{0}$.

Proof:

Here it is convenient to identify A with its matrix in some basis. Note that $\mathbb{C}^{n\times n}$ is an n^2-dimensional vector space. It follows that the n^2+1 matrices $\mathbb{1}, A, A^2, \ldots, A^{n^2}$ are linearly dependent. But this means that there exists numbers $\alpha_0, \ldots, \alpha_{n^2}$, not all zero, such that

$$\alpha_{n^2} A^{n^2} + \alpha_{n^2-1} A^{n^2-1} + \cdots + \alpha_1 A + \alpha_0 \mathbb{1} = 0,$$

that is, $p(A) = 0$, where $p(z) = \alpha_{n^2} z^{n^2} + \alpha_{n^2-1} z^{n^2-1} + \cdots + \alpha_1 z + \alpha_0$.

□

Let $p_{min}(z)$ be a monic polynomial (a monic polynomial is a polynomial with leading coefficient 1) of minimal degree such that $p_{min}(A) = 0$. If $p(z)$ is any polynomial such that $p(A) = 0$ it follows that $p(z) = q(z)p_{min}(z)$ for some polynomial q. To see this, use the division alorithm on p and p_{min}:

$$p(z) = q(z)p_{min}(z) + r(z), \quad \text{where } r = 0 \text{ or } \deg(r) < \deg(p_{min}).$$

Thus $r(A) = p(A) - q(A)p_{min}(A) = 0$. But this implies that $r(z) = 0$, since p_{min} has minimal degree. This also shows that the polynomial p_{min} is unique.

□

It is called the minimal poynomial for A.

By the fundamental theorem of algebra, we can write the minimal polynomial as a product

$$p_{min}(z) = (z-\lambda_1)^{m_1}(z-\lambda_2)^{m_2}\cdots(z-\lambda_k)^{m_k} \tag{G.11}$$

where the numbers λ_j and distinct and $m_j \geq 1$. It will be shown later that the roots λ_j coincide with the eigenvalues of A.

Suppose that $p(z) = p_1(z)p_2(z)$ where p_1 and p_2 are relatively prime. If $p(A) = \mathbf{0}$ we have that

$$V = \ker p_1(A) \oplus \ker p_2(A)$$

and each subspace $\ker p_j(A)$ is invariant under A.

Proof:

The invariance follows from $p_j(A)A\mathbf{x} = Ap_j(A)\mathbf{x} = \mathbf{0}$ for $\mathbf{x} \in \ker p_j(A)$. Since p_1 and p_2 are relatively prime it follows from Bezout's identity that there exists polynmials q_1, q_2 such that

$$p_1(z)q_1(z) + p_2(z)q_2(z) = 1.$$

Thus

$$p_1(A)q_1(A) + p_2(A)q_2(A) = I.$$

Applying this identity to the vector $\mathbf{x} \in V$, we obtain

$$\mathbf{x} = \underbrace{p_1(A)q_1(A)\mathbf{x}}_{\mathbf{x}_2} + \underbrace{p_2(A)q_2(A)\mathbf{x}}_{\mathbf{x}_1},$$

where

$$p_2(A)\mathbf{x}_2 = p_2(A)p_1(A)q_1(A)\mathbf{x} = p(A)q_1(A)\mathbf{x} = \mathbf{0},$$

so that $\mathbf{x}_2 \in \ker p_2(A)$. Similarly, $\mathbf{x}_1 \in p_1(A)q_1(A)$. Thus $V = \ker p_1(A) + \ker p_2(A)$.

Next we prove the uniqueness of $\mathbf{x}_1$ and $\mathbf{x}_2$. If

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 = \mathbf{x}_1' + \mathbf{x}_2', \quad \mathbf{x}_j, \mathbf{x}_j' \in \ker p_j(A),\ j = 1, 2,$$

we obtain that

$$\mathbf{u} = \mathbf{x}_1 - \mathbf{x}_1' = \mathbf{x}_2' - \mathbf{x}_2 \in \ker p_1(A) \cap \ker p_2(A),$$

so that

$$\mathbf{u} = p_1(A)q_1(A)\mathbf{u} + p_2(A)q_2(A)\mathbf{u} = \mathbf{0}.$$

□

With $\lambda_1, \ldots, \lambda_k$ and $m_1, \ldots, m_k$ as in () we have

$$V = \ker(A - \lambda_1 I)^{m_1} \oplus \cdots \oplus \ker(A - \lambda_k I)^{m_k},$$

where each $\ker(A - \lambda_k I)^{m_j}$ is invariant under A. The numbers $\lambda_1, \ldots, \lambda_k$ are eigenvalues of A.

Proof:

We begin by noting that the polynomials $(z - \lambda_j)^{m_j}$, $j = 1, \ldots, k$, are relatively prime. Repeated of the previous result shows that

$$V = \ker(A - \lambda_1 I)^{m_1} \oplus \cdots \oplus \ker(A - \lambda_k I)^{m_k},$$

with each $\ker(A - \lambda_j I)^{m_j}$ invariant.

Consider the restriction of linear operator A to the subspace $\ker(A - \lambda_j I)^{m_j}$. Obvious, $\ker(A - \lambda_j I)^{m_j} \neq \{\mathbf{0}\}$, for otherwise p_{min} would not be minimal. Since every linear operator on a (non-trivial) finite dimensional complex vector space has an eigenvalue, it follows that there is some non-zero element $\mathbf{u} \in \ker(A - \lambda_j I)^{m_j}$ with $A\mathbf{u} = \lambda \mathbf{u}$, $\lambda \in \mathbb{C}$. But then

$$\mathbf{0} = (A - \lambda_j I)^{m_j} \mathbf{u} = (\lambda - \lambda_j)^{m_j} \mathbf{u},$$

so $\lambda = \lambda_j$. This proves that the roots λ_j of the minima; polynomial are eigenvalues of A. On the other hand if $\mathbf{u}$ is an eigenvector of A corresponding to the eigenvalue λ, we have

$$\mathbf{0} = p_{min}(A)\mathbf{u} = (A - \lambda_1 I)^{m_1} \cdots (A - \lambda_k I)^{m_k} \mathbf{u} = (\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_k)^{m_k} \mathbf{u},$$

so $\lambda = \lambda_j$ for some j, that is, every eigenvalue is a root of the minimal polynomial.

□

The subspace $\ker(A - \lambda_j I)^{m_j}$ is called the generalised eigenspace corresponding to λ_j and a non-zero vector $\mathbf{x} \in \ker(A - \lambda_j I)^{m_j}$ is called a generalised eigenvector. The number m_j is the smallest exponent m such that $(A - \lambda_j I)^m$ vanishes on $\ker(A - \lambda_j I)^{m_j}$. Writing $\mathbf{x} \in V$ as $\mathbf{x} = \mathbf{x}_1 + \tilde{\mathbf{x}}$ according to the decomposition

$$V = (A - \lambda_1 I)^{m_1} \oplus \ker \tilde{p}(A),$$

where $\tilde{p}(z) = (z - \lambda_2)^{m_2} \cdots (z - \lambda_k)^{m_k}$, we would then obtain that

$$(A - \lambda_1 I)^{m_1 - 1} \tilde{p}(A)\mathbf{x} = \tilde{p}(A)(A - \lambda_1 I)^{m_1 - 1} \mathbf{x}_1 + (A - \lambda_1 I)^{m_1 - 1} \tilde{p}(A)\tilde{\mathbf{x}} = \mathbf{0},$$

contradicting the definition of the minimal polynomial.

□

If we can select a basis $\{\mathbf{u}_{j,1}, \ldots, \mathbf{u}_{j,n_j}\}$ for each generalized eigenspace, then the union

$$\{\mathbf{u}_{1,1}, \ldots, \mathbf{u}_{1,n_1}, \mathbf{u}_{2,1}, \ldots, \mathbf{u}_{2,n_2}, \ldots, \mathbf{u}_{k,n_k}, \mathbf{u}_{k,1}, \ldots, \mathbf{u}_{k,n_k}\}$$

will be a basis for V. Since each generalized eigenspace is invariant under the linear operator A, the matrix for A in this basis will have the block form

$$\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{pmatrix},$$

where each A_j is a $n_j \times n_j$ square matrix, as

$$\begin{aligned} & A\{\mathbf{u}_{1,1}, \ldots, \mathbf{u}_{1,n_1}, \ldots, \mathbf{u}_{1,1}, \ldots, \mathbf{u}_{k,n_k}\} \\ &= \{\mathbf{u}_{1,1}, \ldots, \mathbf{u}_{1,n_1}, \ldots, \mathbf{u}_{1,1}, \ldots, \mathbf{u}_{k,n_k}\} \begin{pmatrix} & \vdots & & & & \\ \cdots & A_1 & \cdots & & & \\ & \vdots & & & & \\ & & & \ddots & & \\ & & & & \vdots & \\ & & & \cdots & A_k & \cdots \\ & & & & \vdots & \end{pmatrix}, \end{aligned}$$

What remains in order to prove the main theorem is to show that we can select a basis for each generalised eigenspace so that each block A_j takes the form (G.3).

G.4 Proof of the main theorem

Denote the restriction of A to a generalised eigenspace $G(\lambda_j, A) = \ker(A - \lambda_j I)^{m_j}$ by $A_{|G(\lambda_j,A)}$. We have that $A_{|G(\lambda_j,A)}$ has only one eigenvalue. Set $N = A_{|G(\lambda_j,A)} - \lambda_j I$ and then $_j$ is the smallest integer such that $N^{m_j} = 0$. A linear operator N with the property that $N^m = \mathbf{0}$ for some m is called nilpotent.

There is some vector $\mathbf{u} \in \ker(A - \lambda_j I)^{m_j}$ such that $N^{m_j - 1}\mathbf{u} \neq \mathbf{0}$. It follows that the vectors $\mathbf{u}, N\mathbf{u}, \ldots, N^{m_j-1}\mathbf{u}$ are linearly independent. Indeed, suppose that

$$\alpha_1 \mathbf{u} + \alpha_2 N\mathbf{u} + \cdots + \alpha_n N^{m_j-1}\mathbf{u} = \mathbf{0}.$$

Applying N^{m_j-1} to both sides of this equation we obtain that $\alpha_1 N^{m_j-1}\mathbf{u} = \mathbf{0}$ implies $\alpha_1 = 0$. Then applying N^{m_j-2} to

$$\alpha_2 N\mathbf{u} + \cdots + \alpha_n N^{m_j-1}\mathbf{u} = \mathbf{0}$$

implies that $\alpha_2 = 0$. Proceeding this way we obtain $\alpha_j = 0$ for each j. Thus $\{N^{m_j-1}\mathbf{u}, \ldots, N\mathbf{u}, \mathbf{u}\}$ is a basis for $\ker(A - \lambda_j I)^{m_j}$. The matrix for N in this basis is

$$\begin{pmatrix} 0 & 1 & & & 0 \\ 0 & 0 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & 0 & & & 0 \end{pmatrix}$$

as

$$\begin{aligned} N\{N^{m_j-1}\mathbf{u}, N^{m_j-2}\mathbf{u}, \ldots, N\mathbf{u}, \mathbf{u}\} &= \{\mathbf{0}, N^{n-1}\mathbf{u}, \ldots, N^2\mathbf{u}, N\mathbf{u}\} \\ &= \{N^{m_j-1}\mathbf{u}, N^{m_j-2}\mathbf{u}, \ldots, N\mathbf{u}, \mathbf{u}\} \begin{pmatrix} 0 & 1 & & & 0 \\ 0 & 0 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & 0 & & & 0 \end{pmatrix}. \end{aligned}$$

In general, a set of non-zero vectors $\mathbf{u}, \ldots, N^{l-1}\mathbf{u}$ with $N^l\mathbf{u} = \mathbf{0}$ is called a Jordan chain. We prove the theorem in general by showing that there is a basis for V consisting of Jordan chains.

Suppose N is nilpotent. Then there exists vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$ and non-negative integers $m_1, \ldots, m_k$ such that

(i) $N^{m_1}\mathbf{u}_1, \ldots, N\mathbf{u}_1, \mathbf{u}_1, \ldots, N^{m_k}\mathbf{u}_k, \ldots, N\mathbf{u}_k, \mathbf{u}_k$ is a basis for V;

(ii) $N^{m_1+1}\mathbf{u}_1 = \cdots = N^{m_k+1}\mathbf{u}_k = \mathbf{0}$.

Proof:

We prove the theorem by induction on the dimension of V. Clearly the theorem holds when the dimension is 1 (in that case, the only nilpotent operator is the 0 operator, take $\mathbf{u}_1$ to be any non-zero vector and $m_1 = 0$).

Suppose now that the theorem holds for all complex vector spaces of dimension less than n, where $n \geq 2$, and assume that $\dim V = n$. Since N is nilpotent it is not one-to-one and therefore dim range $N < n$. By the inductive hypothesis, we can therefore find a basis of Jordan chains

$$N^{m_i-1}\mathbf{u}_i, \ldots, N\mathbf{u}_i, \mathbf{u}_i, \quad i = 1, \ldots, k, \tag{G.12}$$

for range N. For each $\mathbf{u}_i$ we can find a $\mathbf{v}_i \in V$ such that $N\mathbf{v}_i = \mathbf{u}_i$ (since $\mathbf{u}_i \in$ range N). That is, each Jordan chain in the basis for range N can be extended by one element. We claim that the vectors

$$N^{m_i}\mathbf{v}_i, \ldots, N\mathbf{v}_i, \mathbf{v}_i, \quad i = 1, \ldots, k, \tag{G.13}$$

are linearly independent. Indeed, suppose that

$$\sum_{i=1}^{k}\sum_{j=0}^{m_i} \alpha_{i,j} N^j \mathbf{v}_i = \mathbf{0}. \tag{G.14}$$

Applying N to both sides of this, we find that

$$\sum_{i=1}^{k}\sum_{j=0}^{m_i-1} \alpha_{i,j} N^j \mathbf{u}_i = \sum_{i=1}^{k}\sum_{j=0}^{m_i} \alpha_{i,j} N^{j+1} \mathbf{u}_i = \mathbf{0}, \tag{G.15}$$

which, by hypothesis implies that $\alpha_{i,j} = 0$ for $1 \leq i \leq k$, $0 \leq j \leq l_i - 1$. Substituting this into (G.14) gives that

$$\sum_{i=1}^{k} \alpha_{i,m_i} N^{l_i} \mathbf{u}_i = \sum_{i=1}^{k} \alpha_{i,l_i} N^{m_i} \mathbf{u}_i = \mathbf{0}, \tag{G.16}$$

which again implies that $\alpha_{i,m_i} = 0$ for $1 \leq i \leq k$, by our inductive hypothesis.

Extend the vectors in (G.13) to a basis for V by adding vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$:

$$N^{m_1}\mathbf{v}_1, \ldots, N\mathbf{v}_1, \mathbf{v}_1, \ldots, N^{m_k}\mathbf{v}_k, \ldots, N\mathbf{v}_k, \mathbf{v}_k, \mathbf{w}_1, \ldots, \mathbf{w}_p, \tag{G.17}$$

Each $N\mathbf{w}_j$ is in the range of N and hence we can find an element $\mathbf{x}_j$ in the span of (G.13) such that $N\mathbf{w}_j = N\mathbf{x}_j$. Now let

$$\mathbf{v}_{n+j} = \mathbf{w}_j - \mathbf{x}_j$$

Then $N\mathbf{v}_{n+j} = \mathbf{0}$. Furthermore

$$N^{m_1-1}\mathbf{v}_1, \ldots, N\mathbf{v}_1, \mathbf{v}_1, \ldots, N^{m_k-1}\mathbf{v}_k, \ldots, N\mathbf{v}_k, \mathbf{v}_k, \mathbf{v}_{n+1}, \ldots, \mathbf{v}_{n+p} \tag{G.18}$$

spans V because its span contains each $\mathbf{x}_j$ and each $\mathbf{v}_{n+j}$ and hence $\mathbf{w}_j$ (and because (G.17) spans V). The elements $\mathbf{v}_{n+j}$ are chains of length 1.

□

Let A be a $n \times n$ matrix. There exists an invertible matrix S such that

$$S^{-1}AS = J \tag{G.19}$$

where J is a block matrix,

$$J = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_m \end{pmatrix} \tag{G.20}$$

and each block J_i is a square matrix of the form

$$J_j = \lambda_j \mathbb{1} + N = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix} \tag{G.21}$$

The λ_j need not be distinct. Also, J_j may be a 1×1 matrix (λ_j) containing just the eigenvalue of A.

G.5 Uniqueness of the Jordan normal form up to permutation of Jordan blocks

The matrix J is not completely unique, since in that case we can change the order of the Jordan blocks, this is the only thing which is unique. In other words, the number of blocks and their sizes are uniquely determined.

Let β be the total number of blocks and $\beta(k)$ the number of blocks of size $\beta \times \beta$.

$$\begin{aligned}
\dim\ker N &= \beta, \\
\dim\ker N^2 &= \dim\ker N + \beta - \beta(1), \\
&\vdots \\
\dim\ker N^{k+1} &= \dim\ker N^k + \beta - \beta(1) - \cdots - \beta(k).
\end{aligned}$$

□

G.6 Jordan Form for real matrices

G.6.1 Jordan Form for real matrices with compex eigenvalues

Example: Two complex and on real eigenvalue

Consider

$$A\{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} = \{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \begin{pmatrix} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & \\ 0 & 0 & \overline{\lambda}_1 & 1 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \tag{G.22}$$

First define the matrix Q:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{with inverse} \quad Q^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{G.23}$$

This simply permutes the basis $\{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\}$ to the basis $\{\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,1}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\}$.

Employing these matrices in (G.22)

$$
\begin{aligned}
& A\{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \\
= & \{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & \overline{\lambda}_1 & 1 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\
= & \{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\} \begin{pmatrix} \lambda_1 & 0 & 1 & 0 & 0 \\ 0 & \overline{\lambda}_1 & 0 & 1 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix}
\end{aligned} \tag{G.24}
$$

Define the matrix X:

$$
X = \begin{pmatrix} \frac{1}{2} & -i\frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & i\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -i\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & i\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{with inverse} \quad X^{-1} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ i & -i & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & i & -i & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{G.25}
$$

This takes us from the basis

$$
\{\mathbf{u}_{1,1}, \overline{\mathbf{u}}_{1,1}, \mathbf{u}_{1,2}, \overline{\mathbf{u}}_{1,2}, \mathbf{u}_3\}
$$

to the basis

$$
\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\}
$$

where:

$$
\Re(\mathbf{u}_{1,1}) = \frac{1}{2}(\mathbf{u}_{1,1} + \overline{\mathbf{u}}_{1,1}), \quad \Im(\mathbf{u}_{1,1}) = \frac{1}{2i}(\mathbf{u}_{1,1} - \overline{\mathbf{u}}_{1,1}), \quad \text{etc.} \tag{G.26}
$$

Emplying this in (G.24)

$$
\begin{aligned}
&A\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\} \\
&= \{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\} \times \\
&\times \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ i & -i & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & i & -i & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 1 & 0 & 0 \\ 0 & \overline{\lambda}_1 & 0 & 1 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \overline{\lambda}_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 & -i & 0 & 0 & 0 \\ 1 & i & 0 & 0 & 0 \\ 0 & 0 & 1 & -i & 0 \\ 0 & 0 & 1 & i & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \\
&\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\} \begin{pmatrix} a_1 & b_1 & 1 & 0 & 0 \\ -b_1 & a_1 & 0 & 1 & 0 \\ 0 & 0 & a_1 & b_1 & 0 \\ 0 & 0 & -b_1 & a_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{pmatrix} \\
&\{\Re(\mathbf{u}_{1,1}), \Im(\mathbf{u}_{1,1}), \Re(\mathbf{u}_{1,2}), \Im(\mathbf{u}_{1,2}), \mathbf{u}_3\}B
\end{aligned} \tag{G.27}
$$

where $\lambda_1 = a_1 + ib_1$.

General case

Thus easily generalises. Let

$$
\Lambda = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \quad \mathcal{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathcal{O} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{G.28}
$$

The Jordan block matrix of dimension $2m \times 2m$ is given by the formula

$$
B = \begin{pmatrix} \Lambda & \mathcal{I} & \mathcal{O} & \cdots & \mathcal{O} & \mathcal{O} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \Lambda & \mathcal{I} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \cdots & \mathcal{O} & \Lambda \end{pmatrix} \tag{G.29}
$$

The matrices Q and X are easily generalised. We go from the equation

$$
S^{-1}AS = J
$$

to

$$
(SQX)^{-1}A(SQX) = \mathcal{P}^{-1}A\mathcal{P} = B. \tag{G.30}
$$