SYMBΩL

Vector Spaces Chapter 7: Self-Adjoint Maps

Companion Guide for PAL Sessions

Jack Tabeart

This Companion Guide exists to assist PAL Leaders (and their students) with understanding the mathematical content of the final chapter in the 'Vector Spaces' module, "Self-Adjoint Maps." It will work through each section of the chapter and explain the concepts involved, as well as providing examples where possible. The Guide aims to stimulate a greater understanding of the individual items in the chapter, such as theorems, proofs or definitions, and will refer to previous sections of the course throughout in order to promote a greater holistic understanding of the chapter in context with the other 'Vector Spaces' chapters.

Last Updated: August 2013

Introduction

The chapter "Self-Adjoint Maps", currently the seventh chapter of 'Vector Spaces', can be a difficult one for students to understand! This may be due to the fact that there are challenging new mathematical concepts in the chapter, which students' prior learning will not have introduced them to before. Whatever the reason, students in the past have been known to have a less than full understanding of this chapter before the exam. For similar reasons, PAL Leaders may equally have felt under-equipped to organise a PAL session about certain sections of Chapter 7.

This Companion Guide should be used in conjunction with the existing lecture notes for Chapter 7. It aims to make leaders and students alike more confident and competent with the material in Chapter 7 of the course. This will mainly be done through breaking the individual sections of the chapter down, and explaining the possibly problematic areas in detail. As the chapter is the final one in the course, it refers to certain previous material within its pages. This Guide will also make explicit what the prior theorems or results are which are being referred to, and aims to describe how they fit into the Chapter 7 material.

The sections as presented in the Guide may not exactly match those used in the lecture notes - as I have aimed to group relevant ideas together, rather than following the headings of the notes. Each section here uses the following structure:

- **Prerequisites** - previous knowledge, which will be required before you start in order to follow what is in the section. Details will be given of where this knowledge comes from.

  I will assume throughout the Guide that students are familiar with all content from Chapter 1 ('Introduction & Preliminaries') of the course - e.g. injectivity/surjectivity/bijectivity, basic set notation, etc.

- **Main Content** - the definitions, theorems etc actually in this part of Chapter 7. Comments may be made on proofs, and relevant examples may be given to clarify understanding.

- **Uses** - some comments on why and how the given section is useful to students' understanding of the module, in terms of links to subsequent material or mathematical applications.

Owing to this self-contained structure, all or part of this Guide may be issued to students during a PAL session. Such a structure will hopefully allow for students to deal with specific gaps in their understanding of Chapter 7, as well as helping them to see how the ideas in this chapter link in to the 'big picture' of the module as a whole.

It should be noted that any comments made should not be considered comprehensive, particularly about the mathematical utility of results. This Guide is written by a student (who is not an expert in the field of linear algebra/functional analysis!), and is only a supplement to - not a replacement for - what the module lecturer provides, in terms of both resources and mathematical knowledge.

Nonetheless, I hope that you find this guide informative and useful when it comes to running your Vector Spaces PAL sessions!

Table Of Contents

1 Riesz Representation Theorem

The (slightly simplified) Riesz Representation Theorem can be stated as follows.

Theorem. *Let V be a finite dimensional real or complex Hilbert space and denote by V^* its dual space. Then the map*

$$\Phi : V \to V^*, \quad v \mapsto \langle v, \cdot \rangle$$

is bijective. It is anti-linear if V is a complex space and it is linear if V is a real space.

In this section, we will study this theorem and its proof in detail, in order to fully understand what it says and means. In order to understand the theorem, though, we will need some previous knowledge.

1.1 Prerequisites

Recall from Chapter 4 of 'Vector Spaces' that the dual space of a K-vector space V is the set of homomorphisms from V to K, i.e.

$$\underbrace{V^*}_{\text{dual space of } V} = \underbrace{\mathrm{Hom}_K(V, K)}_{\text{set of homomorphisms from } V \text{ to } K} = \{K\text{-linear maps } f : V \to K\}.$$

Recall also that $\mathrm{Hom}_K(V, W)$, for general K-vector spaces V, W, is a vector space itself. Hence the dual space V^* of a K-vector space V is a space of K-linear maps from V to K.

You should know that a map from a K-vector space V to its field of scalars K is called a functional. In other words, elements of the dual space V^* are called functionals.

You will also need to be familiar with the concept of finite dimensional inner product spaces (or Hilbert spaces) from Chapter 5, and know about the existence of orthonormal bases in every Hilbert space (Chapter 6).

1.2 Main Content

The version of the Riesz Representation Theorem stated in the Vector Spaces lecture notes (henceforth abbreviated to 'RRT') takes V to be a finite dimensional real or complex Hilbert space - so, by definition, V is a vector space with an associated inner product $\langle \cdot, \cdot \rangle$. As V is a vector space, it has a dual space.

The RRT gives us a (linear or anti-linear, depending on whether V is defined over $\mathbb{R}$ or $\mathbb{C}$) bijection between V and its dual space V^*, given by

$$\Phi : V \to V^*, \quad v \mapsto \langle v, \cdot \rangle$$

Note that the expression $\langle v, \cdot \rangle$ is indeed a functional in V^*: the empty second slot of the inner product will accept any element of V; the result of calculating this inner product is a

(real or complex) scalar, belonging to the same field that the Hilbert space V uses. So, we conclude that $\langle\, v\,,\,\cdot\,\rangle$ maps elements of V to scalars. Furthermore, the second slot of the scalar product is linear by definition. Hence $\langle\, v\,,\,\cdot\,\rangle$ is a K-linear map from V to K, and hence the map Φ is well defined.

A full proof of the RRT is given in the lecture notes. Some comments on the proof are below.

- Having proved (anti-)linearity first, one simply has to show that $\ker\Phi\;=\;\{\mathbf{0}\}$ in order to have proven injectivity. This was a theorem from 'Linear Algebra'. Such a task is straightforward once you realise that you require $\langle\, v\,,\,\cdot\,\rangle$ to be the zero map *regardless* of what is in the second slot of the inner product - thus we require in particular that $\langle v, v\rangle = 0$, and by the definition of the inner product this can clearly only be achieved if the element in the first slot is the zero vector in V.

- The surjectivity part of the proof is probably the hardest bit here. You take a general element $\phi\in V^*$, and want to show that it can be expressed as $\phi = \Phi(v)$, for some $v\in V$ - then you have proven surjectivity of Φ.

- Introduce an orthonormal basis $(e_1,\ldots,e_n)$ in V. This will always exist, by the Gram-Schmidt Process of Chapter 6. By manipulating the functional $\langle\, \sum_j \overline{\phi(e_j)}\, e_j,\,\cdot\,\rangle$ one can show that this is equal to $\phi(\,\cdot\,)$

- The functional $\langle\, \sum_j \overline{\phi(e_j)}\, e_j,\,\cdot\,\rangle$ is indeed in V^*, because the entry in the first slot of the inner product is just a linear combination of basis vectors, which thus is in V - by the earlier reasoning, we see that this means the functional is in our dual space V^*.

- We are essentially trying to prove the following functional equation to be true:

$$\left\langle\, \sum_j \overline{\phi(e_j)}\, e_j\,,\,\cdot\,\right\rangle = \phi(\,\cdot\,) \tag{1}$$

An equivalent equation to this, which may look more familiar, is

$$\left\langle\, \sum_j \overline{\phi(e_j)}\, e_j\,,\,x\,\right\rangle = \phi(\,x\,) \quad \forall\, x\in V. \tag{2}$$

But, because V is a vector space and ϕ is a linear map, we do not need to consider all $x\in V$ - we need only consider the (orthonormal) basis vectors, as all $x\in V$ can be obtained by some linear combination of the basis vectors.

- The rest of the surjectivity proof just consists of evaluating this functional equation on all the basis vectors e_k, and thus showing the following to be true:

$$\left\langle\, \sum_j \overline{\phi(e_j)}\, e_j\,,\,e_k\,\right\rangle = \phi(e_k) \quad \forall\, k = 1,\,\ldots,n\,.$$

Therefore equation (2) - i.e. (1) - is true. Hence $\phi = \Phi(v)$, where $v \in V$ is defined as $v = \sum_j \overline{\phi(e_j)}\, e_j$, and hence you have proven surjectivity and the RRT!

1.3   Uses

The Riesz Representation Theorem is a powerful result - you can make any linear functional on a (finite dimensional) Hilbert space by taking the inner product with respect to a certain, unique, vector in the space. The RRT forms a bijection between a finite dimensional Hilbert space and its dual space. This is nontrivial, because it means that both spaces have the same cardinality - this is an entirely non-obvious fact if you think about it!

In terms of the mathematics in Chapter 7, the RRT is an important result to establish at the beginning of the chapter because it will be used to prove the existence and uniqueness of the adjoint map in the next section of the chapter.

2 The Adjoint Map

The adjoint map is a specific type of map between finite dimensional real or complex Hilbert spaces. In this section we show that it exists and is unique under certain conditions.

2.1 Prerequisites

In order to follow the proof of the theorem in this section, you will need to understand what the Riesz Representation Theorem from earlier in Chapter 7 says and means. You will also need to be confident with working with inner products and functionals, as well as knowing about what it means for a map to be bijective.

2.2 Main Content

Next up in Chapter 7 of Vector Spaces is Theorem 7.2, which says the following.

Theorem. *Suppose that V and W are finite dimensional real or complex Hilbert spaces and suppose $f : V \to W$ is a linear map. Then there exists a unique linear map $f^* : W \to V$ such that for all $v \in V$ and $w \in W$:*

$$\langle\, f^*(w)\,,\, v\,\rangle = \langle\, w\,,\, f(v)\,\rangle\,.$$

A full proof of this theorem is given in the Vector Spaces lecture notes. Some comments on the proof are below.

- In the 'existence and uniqueness' section of the proof, we take some arbitrary $w \in W$ and wish to find $\tilde{v} \in V$ such that $\langle \tilde{v}, v \rangle = \langle w, f(v) \rangle \quad \forall v \in V$. Note first that, although both are defined, the inner product on the left of the equality involves elements from V whereas the right hand one involves elements from W. So it is certainly not an obvious statement to prove! You might think that it will be a difficult thing to prove, but in fact the Riesz Representation Theorem will help us out later on.

  Assume for the moment, though, that it is possible to show that such a $\tilde{v}$ exists in V and is unique. Then we have found exactly one $\tilde{v} \in V$ corresponding to the $w \in W$ (that we arbitrarily chose) which satisfies the required property. Hence, as the reasoning would hold for every w in W, there would be a map $f^* : W \to V, \quad w \mapsto \tilde{v}$ which satisfies the required property.

- Back to proving that $\tilde{v}$ exists. Rewrite $\langle \tilde{v}, v \rangle = \langle w, f(v) \rangle \quad \forall v \in V$ into the equivalent functional form $\langle\, \tilde{v}\,,\, \cdot\,\rangle = \langle\, w\,,\, f(\cdot)\,\rangle$ (can you see why these are equivalent?). But now note that the left hand side of this equality is in the correct form for us to use the Riesz Representation Theorem! The RRT gives us a bijection between $\langle\, \tilde{v}\,,\, \cdot\,\rangle \in V^*$ and $\tilde{v} \in V$.

  So, there exists unique $\tilde{v} \in V$ such that $\langle\, \tilde{v}\,,\, \cdot\,\rangle = \langle\, w\,,\, f(\cdot)\,\rangle$, i.e.

$$\exists!\, \tilde{v} \in V \text{ s.t. } \langle\, \tilde{v}\,,\, v\,\rangle = \langle\, w\,,\, f(v)\,\rangle \quad \forall\, v \in V.$$

Thus $\exists f^* : W \to V , \quad w \mapsto \tilde{v}$ satisfying this property, by the above reasoning. This map f^* is unique because there is only this one way of mapping w to $\tilde{v}$ so as to satisfy the property.

- In the 'linearity' section of the proof, the RRT is used once again, to go from

$$\langle\, f^*(w_1 + w_2)\,,\, v\,\rangle = \langle\, f^*(w_1) + f^*(w_2)\,,\, v\,\rangle$$

  to

$$f^*(w_1 + w_2) = f^*(w_1) + f^*(w_2) \qquad (w_1, w_2 \in W \text{ and } v \in V)\,,$$

  and also from

$$\langle\, f^*(\lambda \cdot w)\,,\, v\,\rangle = \langle\, \lambda \cdot f^*(w)\,,\, v\,\rangle$$

  to

$$f^*(\lambda \cdot w) = \lambda \cdot f^*(w) \qquad (w \in W\,,\, v \in V \text{ and } \lambda \in K).$$

  This can be seen in a similar way to before: rewrite your inner product equations into functional form; the RRT gives us a bijection into V. The injectivity part of that bijection allows us to conclude that, for both equations, the entries in the first slots of the inner products (on both sides of each equation) must be equal. Hence the map f^* has been shown to be linear.

So, we have proven Theorem 7.2 by following the proof given in the lecture notes. Definition 7.3 states that this unique linear map $f^* : W \to V$ satisfying $\langle\, f^*(w)\,,\, v\,\rangle = \langle\, w\,,\, f(v)\,\rangle$ is called the adjoint of the original linear map $f : V \to W$.

Thus, every linear map $f : V \to W$ $(V, W$ are real or complex Hilbert spaces) has a corresponding adjoint map.

2.3 Uses

Now that we know about the existence and uniqueness of adjoint maps, we may be able to define additional properties on a linear map between finite dimensional Hilbert spaces, depending on its adjoint. This will be covered in the next section.

3 Self-Adjoint and Unitary

It turns out that it is equally valid to talk about matrices having an adjoint, as well as maps. In fact, this should not be so surprising - remember that $\mathbb{R}^n$ and $\mathbb{C}^n$ are both vector spaces, and further that they are Hilbert spaces.

All endomorphisms on $\mathbb{C}^n$ or $\mathbb{R}^n$ are simply matrices. In general, you can uniquely associate any endomorphism f on a n-dimensional real Hilbert space with a matrix $A_f : \mathbb{R}^n \to \mathbb{R}^n$ (so that A_f is the matrix of the endomorphism f), and similarly for complex numbers. Hence, such matrices will by definition have an adjoint.

Being 'self-adjoint' and being 'unitary' are two different properties that a map or matrix may have, if certain conditions are met. These conditions regard which inner product is being used in a given finite dimensional Hilbert space (sometimes any inner product will do, and sometimes a 'standard' inner product must be used), and whether the given map/matrix's adjoint satisfies certain equalities.

In this section we define these conditions, and show equivalent conditions which will allow us to check more easily if a given map or matrix is either self-adjoint or unitary. Some examples will be given at the end of the section.

3.1 Prerequisites

Recall from Chapter 4 of 'Vector Spaces' that the set of endomorphisms on a K-vector space V is the set of homomorphisms from V to V, i.e.

$$\underbrace{\mathrm{End}_K(V)}_{\text{set of endomorphisms on } V} = \underbrace{\mathrm{Hom}_K(V,V)}_{\text{set of homomorphisms from } V \text{ to } V} = \{K\text{-linear maps } f : V \to V\}.$$

You will need to know what an adjoint map is, and in particular which property it must satisfy - this comes from Definition 7.3 in the notes. You will also need to be familiar with the notions of inverse maps / matrices; also, you will need to know what an orthonormal basis is (from Definitions 6.1, 6.3 in Chapter 6).

You will need to be comfortable performing basic manipulations on maps and matrices (including composing, inverting, transposing and conjugating) and also feel comfortable manipulating inner products.

3.2 Main Content

Maps

Although it is necessary to know that a map's adjoint exists, it is generally not necessary to actually calculate the adjoint in order to determine whether a given map is self-adjoint or unitary, as we shall now show.

In what follows, let V be a finite dimensional real or complex Hilbert space (i.e. its field of scalars K is either $\mathbb{R}$ or $\mathbb{C}$). Any inner product $\langle \cdot , \cdot \rangle : V \times V \to K$ may be used. Note

that in the following equalities, all inner products are inner products involving elements of V, as we are dealing with an endomorphism f on V.

Self-Adjoint: A map $f \in \text{End}(V)$ is called *self-adjoint* if $f^* = f$.

The clue is in the name - a self-adjoint map has its adjoint equal to itself! Thus:

Let $v, w \in V$. Then $\langle\, f(v)\,,\, w\,\rangle = \langle\, f^*(v)\,,\, w\,\rangle = \langle\, v\,,\, f(w)\,\rangle$. The first equality comes from being self-adjoint; the second one comes from the definition of the adjoint.

Often, the middle stage is omitted, so we see that

$$\langle\, f(v)\,,\, w\,\rangle = \langle\, v\,,\, f(w)\,\rangle$$

is an equivalent condition for a map to be self-adjoint.

Unitary: A map $f \in \text{End}(V)$ is called *unitary* if it is invertible and $f^* = f^{-1}$. Thus:

Let $v, w \in V$. Then $\langle\, v\,,\, w\,\rangle = \langle\, f^{-1}\,(\,f(v)\,)\,,\, w\,\rangle = \langle\, f^*\,(\,f(v)\,)\,,\, w\,\rangle = \langle\, f(v)\,,\, f(w)\,\rangle$. The first equality comes from the definition of invertible map; the second one comes from being unitary; the third one comes from the definition of the adjoint.

Often, the intermediary stages are omitted, so we see that

$$\langle\, v\,,\, w\,\rangle = \langle\, f(v)\,,\, f(w)\,\rangle$$

is an equivalent condition for a map to be self-adjoint.

Matrices

Before beginning, let us quickly derive how to calculate the adjoint of real and complex matrices. In what follows, let $A \in \text{Mat}(n\,,\,\mathbb{R})$ (A is a $n \times n$ real matrix), let $B \in \text{Mat}(n\,,\,\mathbb{C})$ (B is a $n \times n$ complex matrix) and $\mathbb{1}_n$ be the $n \times n$ identity matrix.

We must first assume that the standard inner product, $\langle\,\cdot\,,\,\cdot\,\rangle_{\text{st}}$, is being used on either $\mathbb{R}^n$ or $\mathbb{C}^n$, as appropriate. This is because we will need the particular properties of this standard inner product to derive our equalities for calculating the adjoint.

Now, the adjoint of A, which we will denote as A^*, must by definition satisfy

$$\langle\, A^*v\,,\, w\,\rangle_{\text{st}} = \langle\, v\,,\, Aw\,\rangle_{\text{st}} \text{ , where } v, w \in \mathbb{R}^n \text{ are arbitrary.}$$

Let us break this down using the properties of the standard scalar product:

$\langle\, A^{*}v\,,\,w\,\rangle_{\mathrm{st}} = (A^{*}v)^{\mathrm{T}}\,w = v^{\mathrm{T}}\,(A^{*})^{\mathrm{T}}\,w$, and $\langle\, v\,,\,Aw\,\rangle_{\mathrm{st}} = v^{\mathrm{T}}\,A\,w$. So we see that our real matrix A^{*} must satisfy $(A^{*})^{\mathrm{T}} = A$ for the two products to be equal. A reformulation of this is clearly $A^{*} = A^{\mathrm{T}}$.

By following very similar steps (recall that the standard inner product on $\mathbb{C}^{n}$ involves complex conjugating the entry in the first slot), it can likewise be shown that the adjoint B^{*} of a $n \times n$ complex matrix B must satisfy

$$\langle\, B^{*}v\,,\,w\,\rangle_{\mathrm{st}} = \langle\, v\,,\,Bw\,\rangle_{\mathrm{st}} \quad (v, w \in \mathbb{C}^{n} \text{ are arbitrary})$$

which upon reformulation is equivalent to $B^{*} = \overline{B}^{\mathrm{T}}$.

Thus we have derived how to calculate (real or complex) adjoint matrices! Unlike the general case for maps, you will need to actually calculate the adjoint matrix every time you want to check if a particular matrix is self-adjoint or unitary.

Self-Adjoint: Just like with maps, a $n \times n$ matrix M is called *self-adjoint* if it satisfies $M^{*} = M$.

Note that this means in the case of a complex matrix B, we have

$$\overline{B}^{\mathrm{T}} = B$$

as an equivalent condition for a matrix to be self-adjoint.

In the case of a real matrix A (where conjugation has no effect) we simply have

$$A^{\mathrm{T}} = A$$

as an equivalent condition for a matrix to be self-adjoint.

'Unitary': Just like with maps, we must check to see if the property $M^{*} = M^{-1}$ holds (i.e. for matrices, $M\,M^{*} = M^{*}\,M = \mathbb{1}_{n}$).

If a complex matrix B satisfies this property, it is called *unitary*. If a real matrix A satisfies this property, it is called *orthogonal*. Although different words are used to describe the matrices (for historical reasons), the property denoted is essentially the same for both real and complex matrices.

Note, finally, that a $n \times n$ square matrix M (which may be real or complex) is orthogonal/unitary if its columns form an orthonormal basis in $\mathbb{R}^{n}$ or $\mathbb{C}^{n}$, as appropriate. This will be a useful way in which we can construct a given matrix to be orthogonal or unitary.

3.2.1 Examples

Some examples of matrices which are self-adjoint, not self-adjoint, unitary and not unitary are given below. Owing to the comment made at the start of section 3 of this Guide, it is always possible to convert between an endomorphism of the type we are dealing with in this module and its associated matrix - this is why only examples involving matrices are given.

Self-Adjoint: The matrix $S \in \mathrm{Mat}(3, \mathbb{C})$ given by

$$S = \begin{pmatrix} 7 & 1 + 4i & \frac{5}{2} \\ 1 - 4i & 4\pi & \sqrt{2} - \sqrt{3}i \\ \frac{5}{2} & \sqrt{2} + \sqrt{3}i & 0 \end{pmatrix}$$

is a self-adjoint matrix. This is because when we calculate $S^* = \overline{S}^{\mathrm{T}}$ (do this for yourself!), we see that it is equal to S itself.

However, the matrix $R \in \mathrm{Mat}(3, \mathbb{C})$ given by

$$R = \begin{pmatrix} 5 & 6 & 7 + 8i \\ 6 & 4 + 3i & 2 \\ 7 - 8i & 2 & 1 \end{pmatrix}$$

is not a self-adjoint matrix. Again, we need to calculate R^*. When we do this, we see that the problematic element is the $4 + 3i$ in the r_{22} position of the matrix. Upon conjugation, it becomes $4 - 3i \neq 4 + 3i$. So, although all other elements of R satisfy the required property for self-adjointness, this one diagonal element does not.

In fact, more generally, it is true that all elements on the leading diagonal of our $n \times n$ self-adjoint matrix must be real (i.e. have zero imaginary part), so that conjugating these diagonal elements will not affect them. So if we see a $n \times n$ matrix M with at least one non-real diagonal element, we know immediately that it cannot be self-adjoint.

'Unitary': The matrix $U \in \mathrm{Mat}(2, \mathbb{C})$ given by

$$U = \frac{1}{\sqrt{7}} \begin{pmatrix} 1 + 2i & 1 + i \\ 1 - i & -1 + 2i \end{pmatrix}$$

is a unitary matrix. Clearly, its adjoint is as follows:

$$U^* = \frac{1}{\sqrt{7}} \begin{pmatrix} 1 - 2i & 1 + i \\ 1 - i & -1 - 2i \end{pmatrix}.$$

Now upon calculation of $U\,U^*$ we see that

$$\begin{aligned} U\,U^* &= \frac{1}{7} \begin{pmatrix} 1 - 4i^2 + 1 - i^2 & (1 + i)(1 + 2i - 1 - 2i) \\ (1 - i)(1 - 2i - 1 + 2i) & 1 - i^2 + 1 - 4i^2 \end{pmatrix} \\ &= \frac{1}{7} \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix} \\ &= \mathbb{1}_2 \end{aligned}$$

after simplification. By a very similar calculation, it may be shown that $U^*\,U = \mathbb{1}_2$ as well. Hence, by definition, U is unitary.

However, the matrix $T \in \mathrm{Mat}(2, \mathbb{R})$ given by

$$T = \begin{pmatrix} 1/2 & 1/2 \\ 3/2 & 1/2 \end{pmatrix}$$

is not an orthogonal matrix. Note that its adjoint is given by

$$T^* = \begin{pmatrix} 1/2 & 3/2 \\ 1/2 & 1/2 \end{pmatrix} .$$

Calculate $T T^*$ and $T^* T$ for yourself, and thus verify the following to be true:

$$T T^* = \begin{pmatrix} 1/2 & 1 \\ 1 & 5/2 \end{pmatrix} \quad \text{and} \quad T^* T = \begin{pmatrix} 5/2 & 1 \\ 1 & 1/2 \end{pmatrix} .$$

Clearly, the two matrix products do not even equal each other; they certainly do not equal $\mathbb{1}_2$, the 2×2 identity matrix! Thus, we see that T does not satisfy the definition of orthogonality. We conclude that T is not orthogonal.

3.3   Uses

There are a variety of interesting and powerful results associated with self-adjoint maps, some of which will be covered in the next section of this Guide.

Self-adjoint and unitary maps, and self-adjoint and unitary/orthogonal matrices, are fundamental tools in being able to diagonalise maps and matrices. This will be covered when we get to the Spectral Theorem at the end of the chapter (and the end of the Guide!).

Now that we know what it means to be self-adjoint and unitary/orthogonal, we can begin to use such properties in a mathematically helpful manner.

4   Properties of Self-Adjoint Maps

There are many properties of self-adjoint maps, but in particular the four results in this section are of interest to us. Their proofs are all quite short, if you understand the language and concepts involved, and so it is very possible that you could be asked to prove any of them in an exam.

In this section, we will work through the four statements - and proofs - so as to understand them all before the final section of the Guide.

4.1   Prerequisites

For this section, you will need to know what endomorphisms are. You will need to know what the adjoint map is and which condition it satisfies, as well as knowing what it means to be self-adjoint (this is earlier material in Chapter 7).

In addition, you will need to be comfortable with eigenvalues, eigenvectors and eigenspaces (Chapter 4), and be confident dealing with inner products (Chapter 5). You will need to know what orthogonality means and what the orthogonal complement of a set is (Chapter 6).

4.2   Main Content

The following four items are presented in the next section of the 'Vector Spaces' notes:

i) ('Real Eigenvalues')

   Theorem. *If V is a finite dimensional complex Hilbert space then all eigenvalues of a self-adjoint map are real.*

ii) ('Orthogonal Distinct Eigenvectors')

   Theorem. *Let V be a finite dimensional real or complex Hilbert space and $f \in \mathrm{End}(V)$. Then if f is self-adjoint any two eigenvectors of f for <u>different</u> eigenvalues are orthogonal. In other words for $\mu \neq \lambda$ we have $V_\lambda \subset (V_\mu)^\perp$.*

iii) ('Invariant Subsets')

   Theorem. *Let V be a finite dimensional real or complex Hilbert space and let $f \in \mathrm{End}(V)$ be a self-adjoint map. Suppose that W is an invariant subspace under f i.e. $f(W) \subset W$. Then also $W^\perp$ is invariant for f.*

iv) ('At Least One Eigenvalue')

   Lemma. *Let $V \neq \{\mathbf{0}\}$ be a finite dimensional real or complex Hilbert space and let $f \in \mathrm{End}(V)$ be a self-adjoint map. Then f has at least one eigenvalue.*

Full proofs of these results are given in the Vector Spaces lecture notes. Any of them may be examinable. Some comments on the proofs are below.

i) This proof is very short! The first equality uses linearity in the second slot of the inner product, and the eigenvalue equation $f(v) = \lambda \cdot v$; the second equality uses the definition of self-adjointness (the equivalent condition); the third equality again uses the eigenvalue equation; the fourth equality uses anti-linearity in the first slot of the inner product.

You eventually end up with $\lambda \times \langle\, v\,,\, v\,\rangle = \overline{\lambda} \times \langle\, v\,,\, v\,\rangle$, and because your eigenvector $v \neq \mathbf{0}$ by definition, the inner product on both sides is just some scalar greater than zero. Hence divide by this scalar to get $\lambda = \overline{\lambda}$, which means that the imaginary part of λ must equal zero, i.e. the eigenvalue λ is real. This holds for all eigenvalues, so all eigenvalues are real!

ii) Because V is a finite dimensional Hilbert space and f is self-adjoint, we can immediately use result **i)** in our work here. This means all eigenvalues are real, so throughout the proof there is no need to deal with conjugates.

The proof is very similar to the first proof, in that the eigenvalue equation and the equivalent definition of self-adjointness are used throughout the string of equalities. You arrive at $(\mu - \lambda) \times \langle\, v_1\,,\, v_2\,\rangle = 0$ after some rearrangement. One of your initial assumptions was that the eigenvalues are different, so $\mu - \lambda \neq 0$.

Thus, divide by $(\mu - \lambda)$ on both sides, to arrive at $\langle\, v_1\,,\, v_2\,\rangle = 0$ - but this is nothing more than saying that the eigenvectors are orthogonal!

iii) To succeed in this proof, you'll need to know what the orthogonal complement means:

Let x be an element of a subspace $W \subset V$, where V is a finite dimensional Hilbert space. Let y be an element in the orthogonal complement $W^{\perp} \subset V$. Then $\langle x\,, y \rangle = 0$.

You are given what it means for a (sub)space to be invariant under a map, and thus you know that for all $w \in W$, $f(w) \in W$. So all you need to do is take some $v \in W^{\perp}$ and show that $\langle\, f(v)\,,\, w\,\rangle = 0 \quad \forall\, w \in W$, because then it is true that $f(v) \in W^{\perp}$ i.e. $W^{\perp}$ is invariant under f.

There's only one equality required, and it's a straightforward use of the equivalent definition of self-adjointness. Apply it, state that $f(w) \in W$ and $v \in W^{\perp}$ (both of which are true) - thus the inner product equals zero, and hence the orthogonal complement is invariant under f!

iv) If V is a finite dimensional complex Hilbert space, the proof is trivial because the Fundamental Theorem of Algebra (a.k.a. the fact that $\mathbb{C}$ is algebraically closed) guarantees you at least one root of the characteristic polynomial!

If V is a finite dimensional real Hilbert space, then construct the matrix A_f of the map f, via

$$a_{jk} = \langle\, e_j\,,\, f(e_k)\,\rangle \quad (\, j, k = 1, \ldots, n;\ (e_1, \ldots, e_n) \text{ form an orthonormal basis in } V).$$

Aside

This is the usual way in which we construct a matrix of a self-adjoint map. It makes sense to do this, because of the following (this material adapted from pp. 25-26 in Chapter 3 of the notes):

Let V be a finite dimensional (real/complex) Hilbert space, and let $f : V \to V$ be an endomorphism on V. There will be an orthonormal basis in V, say $(e_1, \ldots, e_n)$, by the Gram-Schmidt Process.

For all e_k in the orthonormal basis, the vector $f(e_k)$ can be expressed uniquely as a linear combination of the basis vectors $e_1, \ldots, e_n$. There is a matrix of the map f, $A_f = (a_{jk})$, with (real/complex) entries, such that

$$f(e_k) = \sum_{i=1}^{n} a_{ik}\, e_i \,.$$

Hence, by applying whichever inner product we have on our Hilbert space V, we get

$$\langle\, e_j\,,\, f(e_k)\,\rangle = \langle\, e_j,\, \sum_{i=1}^{n} a_{ik}\, e_i \,\rangle$$

$$= \sum_{i=1}^{n} a_{ik} \langle\, e_j\,,\, e_i\,\rangle$$

$$= \sum_{i=1}^{n} a_{ik}\, \delta_{ji}$$

where δ_{ji} is the 'Kronecker Delta' (it is 0 when $j \neq i$, and it is 1 when $j = i$).

The first equality here is just a substitution; the second utilises the linear second slot of the inner product; the third comes from the fact that our basis is orthonormal. Thus, by the definition of the Kronecker Delta, we see that $\langle\, e_j\,,\, f(e_k)\,\rangle = a_{jk}$ indeed!

Every entry in the matrix A_f can therefore be made in this manner, because we have made no specific assumptions about which value the indices j and k can take - i.e. this reasoning holds $\forall\, j, k = 1, \ldots n$.

Now, note that showing f to have an eigenvalue is equivalent to showing that A_f has an eigenvalue.

It is true that the matrix A_f is also self-adjoint (to be expected) because by the definition of A_f, the definition of inner product and the definition of f being self-adjoint:

$$\overline{a_{kj}} = \overline{\langle\, e_k\,,\, f(e_j)\,\rangle} = \langle\, f(e_j)\,,\, e_k\,\rangle = \langle\, e_j\,,\, f(e_k)\,\rangle = a_{jk} \quad \forall\, j, k = 1, \ldots, n$$

i.e. $A^* \left(= \overline{A}^{\mathrm{T}}\right) = A$.

Now, all the entries of our self-adjoint matrix A_f are real. As $\mathbb{R} \subset \mathbb{C}$, we may consider our matrix to in fact be complex. Thus it is now a self-adjoint map $A_f : \mathbb{C}^n \to \mathbb{C}^n$. By the Fundamental Theorem of Algebra, it has at least one eigenvalue; by A_f being self-adjoint and result **i)**, we see that this eigenvalue is real.

Recall that in fact all entries of A_f are real. So this real eigenvalue is an eigenvalue for the real matrix A_f, which thus means it is an eigenvalue for the original self-adjoint map $f : V \to V$!

4.3   Uses

Note that in our proofs of **ii)** and **iv)** we used our result **i)** to allow us to state that certain eigenvalues were real. In this way, a basic result about self-adjoint maps allowed us to develop our understanding further by proving an even stronger statement about self-adjoint maps.

The results **ii)**, **iii)** and **iv)** will all be of value to us when we prove the Spectral Theorem in the final section of the chapter. Hence, the results about self-adjoint maps which we have just proven will allow us to prove a very powerful result about diagonalising self-adjoint maps and matrices.

5 The Spectral Theorem(s)

The Spectral Theorem is a powerful result from the mathematical branches of linear algebra and functional analysis, and can therefore be considered to be one of the major results in both branches. Crucially, students taking 'Vector Spaces' will have seen a simplified version of the Spectral Theorem, as applied to real square matrices, in their first year course 'Linear Algebra' or similar.

A variety of spectral theorems exist - depending on which kind of maps are used and the dimension of the spaces involved - but all can be considered to provide the same information. Essentially, the Spectral Theorem (of any kind) provides conditions under which it is possible to diagonalise a map or a matrix. This diagonalisation, which is sometimes called a *spectral decomposition*, is a powerful tool for manipulating maps and matrices.

There are three versions of the Spectral Theorem which appear in the 'Vector Spaces' lecture notes, along with its proof. The Theorem and proof tie together all of the elements of Chapter 7 of the course (and also a large number of previous concepts from earlier chapters) in what may at first appear to be a remarkable result. An important remark to make to students is that these three versions of the theorem all tell us the same thing!

In this section, we work through the different versions of the Spectral Theorem, explaining why each version provides the same information to us. We work through the given proof of the theorem from the lecture notes, detailing where applicable how previous results are used in the proof. At the end of the section, some uses and examples of how to use the Spectral Theorem are given, both in the context of Vector Spaces material and in other areas of mathematics.

5.1 Prerequisites

As previously stated, the Spectral Theorem is the culmination of all the Chapter 7 material so far - so for this section you will obviously need to fully understand all the previous parts of the chapter! This includes, but is not limited to:

- Understanding what an adjoint of a map or matrix is;

- Knowing what it means for a map or matrix to be self-adjoint, unitary or orthogonal;

- Understanding what the four results from this Guide's section 4 say and mean, particularly results **ii) - iv)**.

Furthermore, the Spectral Theorem uses results from throughout the course. These results include general notions to do with eigenvalues and eigenspaces (Chapter 4 of the notes); the notion of diagonalisability (Definition 4.26 and Proposition 4.27 from Chapter 4); what a finite dimensional Hilbert space is (Chapter 5); orthogonality, orthonormality, and orthonormal bases in finite dimensional Hilbert spaces (Chapters 5 & 6); orthogonal sums and orthogonal complements (Chapter 6); and finally Theorem 6.9 at the end of Chapter 6, about orthogonal projections.

Once all of these things are clear, we come to the theorem itself.

5.2 Main Content

The three versions of the Spectral Theorem as given in the lecture notes are the following.

Theorem (ST1). *Let V be a finite dimensional real or complex Hilbert space and let $f \in \mathrm{End}(V)$ be a self-adjoint map. Then there exists an orthonormal basis of V consisting of eigenvectors of f. In particular every self-adjoint map is diagonalizable.*

If $(e_1, \ldots, e_n)$ is an orthonormal basis consisting of eigenvectors then we can express f by

$$f = \sum_{i=1}^{n} \lambda_i \langle e_i, \cdot \rangle e_i$$

Theorem (ST2). *Let V be a finite dimensional real or complex Hilbert space and let $f \in \mathrm{End}(V)$ be a self-adjoint map. Denote by P_λ the orthogonal projection onto the λ eigenspace. Then*

$$f = \sum_{\lambda} \lambda P_\lambda$$

where the sum is over all eigenvalues of f.

Theorem (ST3). *Let $A \in \mathrm{Mat}(n, \mathbb{R})$ be a symmetric matrix. Then there is an orthogonal matrix C such that*

$$C^T A C$$

is a diagonal matrix.

Let $A \in \mathrm{Mat}(n, \mathbb{C})$ be a self-adjoint matrix. Then there is an unitary matrix C such that

$$C^* A C$$

is a diagonal matrix.

Perhaps the first thing that should be pointed out is that

<div align="center">

THESE THREE THEOREMS ALL TELL US THE SAME THING!

</div>

I will now detail exactly why these three formulations of the Spectral Theorem can be interchanged depending on what the problem you are trying to solve is.

Why are the first and second versions of the Spectral Theorem the same?

Upon inspection of the conditions for **ST1** and **ST2**, it is immediately clear that both theorems make the same initial assumptions - that V is a finite dimensional Hilbert space (real or complex), and that $f \in \text{End}(V)$ is a self-adjoint map. The consequents look different, though, but they are the same.

Hidden within the statement of both **ST1** and **ST2** is the conclusion that the eigenspaces of f span the Hilbert space V. This must be the case for the equations for f given in each version of the theorem to even make sense. Note first that we have not yet proved this statement - indeed, a proof of "the eigenspaces span the space" constitutes a proof of the Spectral Theorem itself, and we have not yet achieved this (the proof will be worked through in a couple of sections' time).

Next, note that neither theorem explicitly states that "the eigenspaces span the space", yet both rely on it implicitly. **ST1** states its implication, that one can construct an orthonormal basis for V consisting of eigenvectors of f; **ST2** implicitly relies on the eigenspaces spanning V when it represents f exclusively involving orthogonal projections onto its eigenspaces - otherwise, there would be some elements of V which are missed out when redefining f in this new way.

Having concluded implicitly that the eigenspaces of f span V, the two theorems give us what appear to be two different ways to represent f. **ST1** uses functional form, and the sum in its equation is indexed over the number of elements (eigenvectors) in your overall orthonormal basis for V. **ST2** simply expresses f in terms of a specific linear combination of projections onto its eigenspaces - and hence the sum involved is indexed over the number of eigenvalues, not eigenvectors. We now show that these two expressions are the same, though they may look different. Hence we will have shown that **ST1** and **ST2** are exactly the same! An argument to this effect is what follows:

Consider **ST1** first. Now, each of the n eigenvectors in the orthonormal basis for V comes from one of the eigenspaces of f, and thus has an associated eigenvalue; in the equation we are given, λ_i means "the eigenvalue associated with the ith eigenvector of the orthonormal basis for V." Note that several of these n eigenvectors may have the same eigenvalue, depending on how our eigenspaces are comprised. We don't really know anything about how our eigenspaces are comprised (the Gram-Schmidt Process only guarantees us an orthonormal basis in each eigenspace - it doesn't give any further conditions on how many elements are in each basis). But, so long as we are aware that there may be fewer eigenvalues than eigenvectors in the orthonormal basis for V, we can proceed without concern.

As the right hand side of our equation in **ST1** is a linear combination of linear maps, we may rearrange the linear combination into a different order, without consequence, if we so desire. Let us rearrange the linear combination so that it is now ordered by eigen*values* and not by eigenvectors in the orthonormal basis. Factorise the expression (using the distributive law) so that each of our eigenvalues is outside a bracket containing any expressions associated with it - by the above, there may well be less than n brackets now, but that's OK.

Consider **ST2** now. Recall that if $W \subset V$ is a finite dimensional subspace (with $(e_1, \ldots, e_k)$

being an orthonormal basis in W), then

$$P_W : V \to V, \quad v \mapsto P_W(v) = \sum_{i=1}^{k} \langle\, e_i \,,\, v \,\rangle\, e_i$$

is the orthogonal projection onto W. Note that an equivalent functional form of this orthogonal projection is

$$P_W = \sum_{i=1}^{k} \langle\, e_i \,,\, \cdot \,\rangle\, e_i$$

Let us now consider the given sum in **ST2**, taken over all orthogonal projections onto eigenspaces of V. This will be a sum of a sequence of terms, each consisting of an eigenvalue multiplied by a bracket containing a sum of terms of the form $\langle\, e_j \,,\, \cdot \,\rangle\, e_j$, where the number of summands will vary depending on how many elements form an orthonormal basis in the particular eigenspace you are currently considering.

Compare the two expressions you have produced for f in both **ST1** and **ST2**. You should be able to see - possibly after rearrangement - that the expressions are identical. Hence both formulae for f are exactly the same, so the first and second versions of the Spectral Theorem are the same after all!

Aside

Students may have difficulties following this rather general argument, given that it involves manipulating quite abstract expressions if no specific dimensions, eigenvalues or eigenvectors are given. You may find it helpful to use a specific example when covering this section. Present it to students so as to show them that the two expressions will equate in this case - then point out that it will always hold in a more arbitrary case, due to a simple rearrangement of the linear combination. Hopefully students will then be able or willing to generalise from this specific example!

The following is one specific example which you could present to students:

Example

Let V be a 6-dimensional (real or complex) Hilbert space, and let $f \in \text{End}(V)$ be a self-adjoint map.

Assume that f has eigenvalues 4, $\sqrt{2}$ and π. Assume further (WLOG) that $\{e_4, e_1, e_5\}$ forms an orthonormal basis in the 4-eigenspace, $\{e_3, e_6\}$ forms an orthnormal basis in the $\sqrt{2}$-eigenspace and $\{e_2\}$ is an orthonormal basis for the π-eigenspace.

If we express f by

$$f = \sum_{i=1}^{n} \lambda_i \langle\, e_i \,,\, \cdot \,\rangle\, e_i$$

then we get

$$f = 4 \langle e_1, \cdot \rangle e_1 + \pi \langle e_2, \cdot \rangle e_2 + \sqrt{2} \langle e_3, \cdot \rangle e_3 + 4 \langle e_4, \cdot \rangle e_4 + 4 \langle e_5, \cdot \rangle e_5 + \sqrt{2} \langle e_6, \cdot \rangle e_6$$
$$= \ldots$$
$$= 4 \left(\langle e_1, \cdot \rangle e_1 + \langle e_4, \cdot \rangle e_4 + \langle e_5, \cdot \rangle e_5 \right) + \sqrt{2} \left(\langle e_3, \cdot \rangle e_3 + \langle e_6, \cdot \rangle e_6 \right) + \pi \left(\langle e_2, \cdot \rangle e_2 \right)$$

after some rearrangement.

If we express f by

$$f = \sum_{\lambda} \lambda P_{\lambda}$$

then we get

$$f = 4 \left(\langle e_4, \cdot \rangle e_4 + \langle e_1, \cdot \rangle e_1 + \langle e_5, \cdot \rangle e_5 \right) + \sqrt{2} \left(\langle e_3, \cdot \rangle e_3 + \langle e_6, \cdot \rangle e_6 \right) + \pi \left(\langle e_2, \cdot \rangle e_2 \right)$$
$$= \ldots$$
$$= 4 \left(\langle e_1, \cdot \rangle e_1 + \langle e_4, \cdot \rangle e_4 + \langle e_5, \cdot \rangle e_5 \right) + \sqrt{2} \left(\langle e_3, \cdot \rangle e_3 + \langle e_6, \cdot \rangle e_6 \right) + \pi \left(\langle e_2, \cdot \rangle e_2 \right)$$

after some rearrangement.

Clearly, the two end expressions for f are identical! Hence our two expressions for f in terms of eigenvalues and eigenvectors equate - i.e.

$$\sum_{i=1}^{n} \lambda_i \langle e_i, \cdot \rangle e_i = \sum_{\lambda} \lambda P_{\lambda}$$

Why is the third version of the Spectral Theorem the same as the others?

This is considerably easier to see - **ST3** involves square matrices of finite order n, which can be considered to be endomorphisms on $\mathbb{R}^n$ or $\mathbb{C}^n$, as appropriate.

ST3 assumes initially that A is either a real, square, symmetric matrix of order n (i.e. a self-adjoint endomorphism on the finite dimensional real Hilbert space $\mathbb{R}^n$) or a complex, square, self-adjoint matrix of order n (i.e. a self-adjoint endomorphism on the finite dimensional complex Hilbert space $\mathbb{C}^n$). Hence the assumptions in **ST3** are exactly the same as in **ST1** or **ST2**.

What occurs in **ST3** can be considered to be a particular subset of cases from the first or second version of the Spectral Theorem. As such, the immediate (implicit) next conclusion is that the eigenspaces of A span $\mathbb{R}^n$ or $\mathbb{C}^n$, as appropriate. Hence an orthonormal basis of eigenvectors of A exists, and it spans $\mathbb{R}^n$ or $\mathbb{C}^n$.

Next, recall from either page 10 of this Guide or pages 58-59 in the lecture notes that a $n \times n$ square matrix M is orthogonal/unitary if its columns form an orthonormal basis in $\mathbb{R}^n$ or

$\mathbb{C}^n$. Thus, **ST3** allows us to conclude that we can create an orthogonal or unitary matrix by taking the orthonormal basis eigenvectors (which we know exist) and combining them in order into a matrix C. This is achieved in the simplest way possible: the first (column) vector becomes the first column of C, etc.

So far, we have the existence of an orthogonal/unitary matrix C in both cases in **ST3**. The fact that this matrix can be used to diagonalise A (by calculating the given matrix products $C^{\mathrm{T}}AC$ or C^*AC) should not be a new idea to you - we saw it in first year, in 'Linear Algebra', or its equivalent. Admittedly, we saw only the real case and not the complex case that appears here, but this result can just be considered to be a natural extension of that result from 'Linear Algebra' to some more general cases.

Hence we have shown that **ST3** is exactly the same as **ST1** and **ST2** when it comes down to what is being assumed and what is being concluded! Thus, indeed, all three versions of the Spectral Theorem given in the lecture notes can be considered to be interchangeable.

General comments on the Spectral Theorem and its proof

We have now concluded that all three versions of the Spectral Theorem can in a sense be considered 'the same.' In exactly which way is this?

Recall from Definition 4.26 and Proposition 4.27 (from Chapter 4 in the notes) what it means to be called *diagonalisable*. In **ST1** and **ST2** we have shown the existence of an orthonormal basis of eigenvectors - this means that we have shown our map f to be diagonalisable. What follows in the first two versions of the Spectral Theorem is a statement of two 'different' (but actually the same!) ways in which we can actually diagonalise f. So we show that f is diagonalisable and explicitly state how to diagonalise it.

In the case (**ST3**) that we are dealing with matrices, we have shown the existence of an orthonormal basis of eigenvectors, then we put these eigenvectors into the columns of a matrix C. By definition, that matrix is unitary/orthogonal. As such, it satisfies $C^* = C^{-1}$ and hence can be used to diagonalise A according to the statements that follow in **ST3**. So we show that A is diagonalisable and explicitly state how to diagonalise it.

Therefore, we see that

> *the reason the Spectral Theorem is of use to us is because it gives us a set of conditions under which it is possible to diagonalise certain maps or matrices, as well as an explicit method for this diagonalisation.*

Such a representation of a map or matrix in diagonal form is sometimes called a *spectral decomposition* - the word 'spectral' relates to the spectrum of a map or matrix. In finite dimensional Hilbert spaces, like the ones considered in Chapter 7 of this course, the spectrum of a map is just the set of its eigenvalues; in an infinite dimensional space, this is not necessarily true.

Now, remember that in each of the Spectral Theorems written in the lecture notes, the proof relied on proving the assertion "the eigenspaces span the space", because then we could conclude the existence of an orthonormal basis in our domain consisting of eigenvectors of our map/matrix. A proof of this, given the appropriate initial assumptions, constitutes a proof of the Spectral Theorem, because once this statement has been proven then all the other parts of the theorem follow.

A proof of "the eigenspaces span the space" is given in the 'Vector Spaces' lecture notes for the case where f is a self-adjoint endomorphism (the matrix case is a specific subset of cases of this more general case). Some comments on the proof are below.

- In symbol form, "the eigenspaces span the space" can be written as:

  Let $\{\lambda_1, \ldots, \lambda_k\}$ be the set of eigenvalues of f and let the spaces $V_{\lambda_1}, \ldots, V_{\lambda_k}$ be the corresponding eigenspaces. Then $V = V_1 \perp V_2 \perp \ldots \perp V_k$.

- We know to begin with (from an application of theorem **ii)** from section 4 of this Guide - 'Orthogonal Distinct Eigenvectors') that any two of the eigenspaces are orthogonal to each other, and thus that all the eigenspaces are orthogonal to each other. Thus, it only remains to show that $V_1 \perp V_2 \perp \ldots \perp V_k = V$.

- Denote the orthogonal sum of the eigenspaces by $W \subset V$. It can be easily shown - by taking a linear combination of orthonormal basis eigenvectors, applying the linear map f and using the eigenvalue equation - that W is invariant under f. We can immediately apply theorem **iii)** from section 4 ('Invariant Subsets') to conclude that $W^\perp \subset V$ is invariant under f.

- $W^\perp$ is a finite dimensional subspace of the finite dimensional Hilbert space V and, by definition of W, it cannot contain any eigenvectors. The restriction endomorphism $f|_{W^\perp}$ is a self-adjoint map.

- A proof by contradiction is used to conclude that $W^\perp = \{\mathbf{0}\}$:

  Assume $W^\perp \neq \{\mathbf{0}\}$. Then the map $f|_{W^\perp}$ would have to have at least one eigenvalue (by lemma **iv)** from section 4, 'At Least One Eigenvalue'). This would mean that $W^\perp$ would contain an eigenvector - but this is impossible. CONTRADICTION! Hence we conclude that $W^\perp = \{\mathbf{0}\}$ after all.

- Recall Theorem 6.9 from the notes - for any finite dimensional subspace $W \subset V$, it is true that $V = W \perp W^\perp$.

  We have just concluded that $W^\perp = \{\mathbf{0}\}$, which by Theorem 6.9 means that $W = V$, i.e.
  $$V_1 \perp V_2 \perp \ldots \perp V_k = V.$$

  Thus, the eigenspaces span the space, and hence the Spectral Theorem has been proved!

5.3   Use of the Spectral Theorem

We know that the Spectral Theorem is valuable to us because it allows us to diagonalise certain maps/matrices, under specific conditions. But why is being able to diagonalise such maps/matrices so useful to us?

Perhaps the main reason in the 'Vector Spaces' course is that this diagonal representation - in terms of eigenvalues and eigenvectors - is in some sense very simple to manipulate. We can calculate many functions of a self-adjoint (real or complex) endomorphism (and hence a self-adjoint, real or complex, square matrix), by appropriately applying that function to the diagonalised form. This is surprisingly helpful, and increases the number of mathematical problems we can solve.

Some theory on functions of matrices and functions of self-adjoint endomorphisms, including a precise statement of how to evaluate functions of both, is now presented.

Functions of Self-Adjoint Matrices

First of all, let us assume we have a suitable self-adjoint finite order square matrix M. Then **the Spectral Theorem says** we can find a unitary matrix C (consisting of orthonormal basis eigenvectors of M) such that $C^* M C = D$, where D is a diagonal matrix comprised of eigenvalues of M. Immediately, then, by the definition of unitary matrices, we can multiply these equal expressions on the left by C and on the right by C^*, in order to arrive at

$$CC^* M CC^* = C D C^*$$

or, upon simplification,

$$M = C D C^*.$$

Hence, we have a new way of expressing M - in terms of matrices containing eigenvectors and eigenvalues. This representation is useful to us because we can use it to easily calculate functions which take matrices as their arguments!

Take for example the function f defined by $f(x) = x^5$. We often think of such functions as taking a real number (say) as its argument, and then returning another real number as its output. However, there is no reason that such a function could not take a square matrix as its argument and then return another square matrix as its output, for example. So we could evaluate $f(M) = M^5$ if we so wished.

But such a calculation will often be difficult to do by hand, particularly if the entries of M are not relatively easy to work with, or if M has a large order (e.g. multiplying a 6×6 matrix by itself 5 times will usually take a lot of work!). Having said this, it *can* be quite simple to perform such calculations - and being able to diagonalise a matrix is exactly one way we can simplify the process.

If we know that $M = C D C^*$, then $f(M) = M^5$ (say) becomes

$$
\begin{aligned}
f(M) = M^5 &= (\,C D C^*\,)^5 \\
&= (\,C D C^*\,)(\,C D C^*\,)(\,C D C^*\,)^3 \\
&= (\,C D\,)(\,C^* C\,)(\,D C^*\,)(\,C D C^*\,)^3 \\
&= (\,C D\,)\,\mathbb{1}\,(\,D C^*\,)(\,C D C^*\,)^3 \qquad (\mathbb{1}\text{ is the appropriate identity matrix}) \\
&= (\,C D^2 C^*\,)(\,C D C^*\,)^3 \\
&= \cdots \\
&= C D^5 C^* \qquad \text{(upon simplification, using the definition of unitary matrices)} \\
&= C\,(\,f(D)\,)\,C^*.
\end{aligned}
$$

In a similar way, any power of M can inductively be calculated (by calculating $C\,(D^n)\,C^*$) - and this is much quicker than calculating it in the usual way for M, because most of the elements in D are zero, and the ones that are not are eigenvalues which are relatively simple to manipulate. In fact, because D is a diagonal matrix, when you do the matrix multiplication any off-diagonal elements will always be zero (try an example for yourself to see why this is the case!) and the diagonal elements will just end up raised to whatever power you are evaluating.

Recall that the '1' of matrix multiplication (i.e. the multiplicative identity element) is the appropriately sized identity matrix $\mathbb{1}$; recall also that matrices can be multiplied by scalars by multiplying each entry of the matrix by the given scalar. Hence it is possible to consider a function such as $g(x) = 4x^2 + 7$, and evaluate it on matrices by calculating $g(M) = 4 M^2 + 7\,\mathbb{1}$. Thus, by combining the above approaches with the general method for evaluating powers of diagonalisable matrices, we arrive at the following:

Theorem. *Let M be a $n \times n$ self-adjoint (real or complex) matrix. Let $M = C D C^*$, where C and D are appropriate matrices for diagonalising M.*

Let $p(x) = a_k x^k + a_{k-1} x^{k-1} + \ldots + a_1 x + a_0$
(where the a_i are all real or complex constants, as appropriate) be any polynomial in x.

Then one can calculate $p(M)$ by calculating $p(M) = C\,(\,p(D)\,)\,C^*$.

If the leading diagonal of the diagonal matrix D contains in order the eigenvalues $\lambda_1, \ldots, \lambda_n$, then the diagonal matrix $p(D)$ has the elements $p(\lambda_1), \ldots, p(\lambda_n)$ in order on its leading diagonal (all other entries are zero).

This theorem can be extended to encompass infinite power series

$$
P(x) = a_0 + a_1 x + \ldots + a_n x^n + \ldots
$$

as well, by adapting it in the obvious required ways.

Although this theorem is good, we can go further! Recall that there are certain series expansions (Maclaurin Series) for functions such as $\cos(x)$, $\sin(x)$ and e^x. Such expansions express functions in terms of infinite power series in x. However, we now know how to calculate infinite power series in a diagonalisable matrix M. Hence, we can meaningfully calculate expressions such as $\cos(M)$ or e^M in an analogous manner:

$$\text{If } M = C\,D\,C^*, \text{ then:} \quad \cos(M) = C\,(\,\cos(D)\,)\,C^*;$$
$$e^M \quad = C\,(\,e^D\,)\,C^* \text{ etc.}$$

The square matrices $\cos(D)$, e^D etc encompassed in the above equations are all diagonal matrices. Hence the elements not on the leading diagonal are always zero. If the leading diagonal of D contains in order the eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\cos(D)$ has the elements $\cos(\lambda_1), \ldots, \cos(\lambda_n)$ in order on its leading diagonal; similarly for e^D etc.

In fact, we can go further than even this! Recall that the eigenvalues of our self-adjoint matrix M are all real, by theorem **i)** from section 4.2 of this Guide (or Theorem 7.7 on page 5 of Chapter 7). Let $F \in C(\mathbb{R})$ be any continuous real-valued function. Then for this function F, we can evaluate $F(M)$ in the following way:

$$\text{If } M = C\,D\,C^*, \text{ then } F(M) = C\,(\,F(D)\,)\,C^*$$
(where $F(D)$ is a diagonal matrix, with diagonal entries $F(\lambda_1)$, $F(\lambda_2)$ etc).

As we have already seen, to evaluate a function on a self-adjoint matrix we in effect only evaluate it on the diagonal elements of the diagonal matrix - i.e. the eigenvalues of the original matrix. Thus, as the eigenvalues are all real, we can in fact consider evaluating any real function on a self-adjoint square matrix.

Hence there are a great deal of possible functions which we could evaluate on any given self-adjoint square matrix! In general, functions of matrices may be used for a wide variety of purposes - one such purpose is to solve systems of differential equations - but the uses of such functions will not be elaborated on further in this Guide.

Functions of Self-Adjoint Endomorphisms

We have covered how to calculate functions of self-adjoint matrices in the previous section. However, self-adjoint matrices are just one subtype of self-adjoint endomorphisms on a finite dimensional (real or complex) Hilbert space. The Spectral Theorem in fact gives us more generality than just diagonalising (and hence evaluating functions of) self-adjoint matrices. Therefore it is worthwhile to consider how we can evaluate functions of endomorphisms in a more general manner!

Let $f \in \mathrm{End}(V)$ be a self-adjoint endomorphism on a finite dimensional (real or complex) Hilbert space. Then **the Spectral Theorem tells us** we can express f in terms of its eigenvalues and orthonormal basis eigenvectors:

$$f = \sum_{i=1}^{n} \lambda_i \langle\, e_i\,,\, \cdot\,\rangle\, e_i \qquad \text{(where } (e_1, \ldots, e_n) \text{ is an orthonormal basis of eigenvectors)}$$

As was the case before, here λ_i means "the eigenvalue associated with the ith eigenvector of the orthonormal basis for V."

We can use quite a bit of the theory discussed regarding diagonalising matrices when we come to talk about diagonalising endomorphisms - this is because the matrix case is a specific set of subcases of the more general situation for endomorphisms. Although partially different in specifics, in broad terms the method can be considered to be the same, as we shall now see.

It can be shown, using similar methods to the ones employed in the previous section, that any polynomial $p(x)$ can be evaluated on f by

$$p(f) = \sum_{i=1}^{n} p(\lambda_i) \langle e_i, \cdot \rangle e_i.$$

Similarly, we can extend this notion to encompass power series and certain common functions like the exponential function. In a similar manner, we can further evaluate any real-valued continuous function $F \in C(\mathbb{R})$ on f, by

$$F(f) = \sum_{i=1}^{n} F(\lambda_i) \langle e_i, \cdot \rangle e_i.$$

So we see that we can express many functions of self-adjoint endomorphisms in an easy to compute way, by using the Spectral Theorem to diagonalise the endomorphism first. Sometimes it will be more advantageous to directly diagonalise an endomorphism f when calculating such a function on it, rather than calculating the associated matrix A_f; sometimes, equally, we will prefer for the purposes of our question to manipulate matrices instead. However, as we can do either (and know how to do both!), you should use whichever approach would appear to suit your specific question best.

5.4   Examples and Applications

In the last section we found out how to use the Spectral Theorem to evaluate certain functions on self-adjoint endomorphisms (or square matrices), via diagonalisation. We can use this knowledge to solve several mathematical problems, and prove a variety of mathematical results. Some such problems appear on Problem Sheet 7 for 'Vector Spaces'; some applications of this knowledge appear at the end of the lecture notes for Chapter 7; plenty of others exist too!

In this final section, I present a few examples of how we can use diagonalisation and the Spectral Theorem, both within the Vector Spaces course and in other areas of mathematics. An application to previously encountered first year material is also given. The material presented here should obviously not be considered exhaustive - indeed, please have a look at other questions online and read further about applications of the Spectral Theorem in order to get a greater sense of its versatility.

Example of Diagonalising a Matrix: Calculating the Fibonacci Numbers

Have you ever wondered what the first 100 (or more) Fibonacci numbers are? Using the Spectral Theorem, we can easily calculate the entire sequence of Fibonacci numbers ...

Let us begin in a more general fashion. Suppose that we want to solve a general second order linear recurrence relation of the form

$$x_{n+2} = a\,x_n + b\,x_{n+1}\,, \tag{3}$$

where a and b are constants. Make the substitutions $u_n := x_n$ and $v_n := x_{n+1}$. Hence it is true that $u_{n+1} = x_{n+1} = v_n$ and, by (3), $v_{n+1} = x_{n+2} = a\,x_n + b\,x_{n+1} = a\,u_n + b\,v_n$.

Thus we have $u_{n+1} = v_n$, $v_{n+1} = a\,u_n + b\,v_n$. We can write this in matrix form as

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix} \begin{pmatrix} u_n \\ v_n \end{pmatrix},$$

or $U_{n+1} = AU_n$, using obvious notation. By induction, it is true that

$$U_{n+1} = AU_n = A\,(\,AU_{n-1}\,) = A^2 U_{n-1} = A^2\,(\,AU_{n-2}\,) = \cdots = A^{n+1}\,U_0\,,$$

i.e.

$$U_n = A^n\,U_0\,.$$

Upon resubstitution for U_n, U_0, A, u_n and v_n, we see that

$$\begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}^n \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \tag{4}$$

is an expression which will allow us to calculate the nth and $(n+1)^{\text{th}}$ terms in our sequence, provided that we are given the first two terms x_0 and x_1.

Note that in general this expression is difficult to calculate, as it involves raising a matrix to a power. However, if that matrix A is diagonalisable, then the problem becomes far simpler! The Spectral Theorem can be used to determine if a given matrix is diagonalisable, and hence can be used to determine the general solution to certain second order linear recurrence relations! An example of this, in the case where our sequence is the Fibonacci Sequence now follows.

Take now the specific second order linear recurrence relation

$$x_{n+2} = x_n + x_{n+1}$$

with $x_0 = 0$ and $x_1 = 1$ (this is the Fibonacci Sequence). We wish to find the general solution of this recurrence relation - i.e. find a way of calculating the nth term in the sequence.

Note that we have $a = 1$, $b = 1$ in our recurrence relation. Thus, by equation (4), it is true that

$$\begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}^n \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

in our particular case. Write $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ for purposes of reference. Then we know that

$$\begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = A^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{5}$$

gives us the nth and $(n+1)^{\text{th}}$ terms in the Fibonacci Sequence.

Note that A is self-adjoint (it is easily seen that $A^* = A$); note further that A is a 2×2 real matrix. Hence, **by the Spectral Theorem**, it is possible to diagonalise A - that is, find an orthogonal matrix C such that $C^{\mathrm{T}} A C$ equals a diagonal matrix D. Because this is possible, we can then express A as $A = C D C^{\mathrm{T}}$, and hence evaluate (5) easily, via

$$\begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = C D^n C^{\mathrm{T}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

To actually diagonalise A, we must calculate the eigenvalues and eigenspaces of A first. We find bases of eigenvectors in each eigenspace, normalise the eigenvectors, and then combine these vectors into the orthogonal matrix C.

The eigenvalues of A can in our case be calculated as the roots of the characteristic polynomial for A, $\chi_A(\lambda) := \det(A - \lambda \mathbb{1})$:

$$\chi_A(\lambda) = \det(A - \lambda \mathbb{1})$$
$$= \det \begin{pmatrix} -\lambda & 1 \\ 1 & 1 - \lambda \end{pmatrix}$$
$$= \lambda^2 - \lambda - 1.$$

So, the eigenvalues of A are the solutions of $\lambda^2 - \lambda - 1 = 0$, which can be obtained using the quadratic formula:

$$\lambda = \frac{1 \pm \sqrt{5}}{2}$$

i.e.

$$\lambda_1 = \frac{1 + \sqrt{5}}{2} \text{ and } \lambda_2 = \frac{1 - \sqrt{5}}{2}.$$

Note that $\lambda_1 \lambda_2 = -1$, or equivalently that $-\frac{1}{\lambda_1} = \lambda_2$ and $-\frac{1}{\lambda_2} = \lambda_1$. Also note that

$\lambda_1 + \lambda_2 = 1$, or equivalently that $1 - \lambda_1 = \lambda_2$ and $1 - \lambda_2 = \lambda_1$.

Often, our two eigenvalues are given special names: λ_1 is more commonly known as ϕ or "the golden ratio"; $\lambda_2 = 1 - \phi$ can be expressed as $-\Phi$, where $\Phi = \frac{1}{\phi}$ is sometimes known as "'the golden ratio conjugate'."

Next, we calculate the eigenspaces:

$V_{\lambda_1} = \ker(A - \lambda_1 \mathbb{1})$, i.e. the solution of the augmented matrix system

$$\left(\begin{array}{cc|c} -\lambda_1 & 1 & 0 \\ 1 & \lambda_2 & 0 \end{array} \right).$$

Gaussian Elimination can be used to rewrite the above augmented matrix system as

$$\left(\begin{array}{cc|c} 1 & \lambda_2 & 0 \\ 0 & 0 & 0 \end{array} \right).$$

Hence we can say that $V_{\lambda_1} = \mathrm{span}_{\mathbb{R}} \left\{ \left(\begin{array}{c} \lambda_2 \\ -1 \end{array} \right) \right\}$.

$V_{\lambda_2} = \ker(A - \lambda_2 \mathbb{1})$, i.e. the solution of the augmented matrix system

$$\left(\begin{array}{cc|c} -\lambda_2 & 1 & 0 \\ 1 & \lambda_1 & 0 \end{array} \right).$$

Gaussian Elimination can be used to rewrite the above augmented matrix system as

$$\left(\begin{array}{cc|c} 1 & \lambda_1 & 0 \\ 0 & 0 & 0 \end{array} \right).$$

Hence we can say that $V_{\lambda_2} = \mathrm{span}_{\mathbb{R}} \left\{ \left(\begin{array}{c} \lambda_1 \\ -1 \end{array} \right) \right\}$.

Next, normalise these eigenvectors:

$$N_1 := \left\| \left(\begin{array}{c} \lambda_2 \\ -1 \end{array} \right) \right\| = \sqrt{\frac{10}{4} - \frac{\sqrt{5}}{2}} = \sqrt{\frac{5 - \sqrt{5}}{2}} \, ;$$

$$N_2 := \left\| \left(\begin{array}{c} \lambda_1 \\ -1 \end{array} \right) \right\| = \sqrt{\frac{10}{4} + \frac{\sqrt{5}}{2}} = \sqrt{\frac{5 + \sqrt{5}}{2}} \, .$$

Having normalised our eigenvectors, we can combine the resultant orthonormal basis eigenvectors into an orthogonal matrix, as discussed previously in the Guide:

$$C = \begin{pmatrix} \frac{\lambda_2}{N_1} & \frac{\lambda_1}{N_2} \\[2ex] -\frac{1}{N_1} & -\frac{1}{N_2} \end{pmatrix}$$

It is possible to check that C is indeed orthogonal, should you wish to - although it involves a lot of manipulation of surds, it can be done! But C is by definition orthogonal **as a result of the Spectral Theorem**. This orthogonal matrix C can be used to diagonalise A, as follows:

$$C^{\mathrm{T}} A C = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

If we set $\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = D$, then by the definition of orthogonality we thus have

$$A = C D C^{\mathrm{T}}.$$

We can use this as a substitution in equation (5) in order to calculate any term in the Fibonacci Sequence we like! Upon substitution we arrive at

$$\begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = C \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^n C^{\mathrm{T}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{6}$$

(with C as defined above in terms of λ_1, λ_2, N_1 and N_2) as an explicit formula for the nth and $(n+1)^{\mathrm{th}}$ Fibonacci numbers.

A computer algebra package such as Maple can be coded to calculate these terms, and thus produce a list of the Fibonacci numbers for reasonably large values of n. For example, Maple can calculate the first 100 Fibonacci numbers incredibly quickly, and can evaluate any term up to the $100,000^{\mathrm{th}}$ in the Fibonacci Sequence after a maximum of a few seconds' calculation. Beyond that, it can take a lot longer to evaluate (or Maple may crash whilst trying to calculate using huge numbers) - but Maple can still compute up to the $100,000^{\mathrm{th}}$ term far faster than we possibly could by hand!

Here we have seen that the Spectral Theorem has allowed us to completely solve Fibonacci's famous recurrence relation! The Fibonacci numbers (and ϕ, the 'golden ratio') occur in a diverse range of mathematical and non-mathematical areas, including number theory, biology, music and architecture.

Example of Diagonalising an Endomorphism: 'Square Rooting' a Matrix

Let us assume that $A \in \mathrm{Mat}(n, K)$ is a self-adjoint (real or complex) matrix with non-negative eigenvalues. Then the Spectral Theorem can be used to 'square root' this matrix.

That is, we can find a self-adjoint matrix $B \in \mathrm{Mat}(n, K)$ such that $B^2 = A$. An argument to this extent follows.

By assumption, A is self-adjoint and of finite order. Let us consider A as a self-adjoint endomorphism on a finite dimensional Hilbert space (i.e. K^n with the standard inner product). Hence, **by the Spectral Theorem**, we can represent A as

$$A = \sum_i \lambda_i \langle e_i \,, \, \cdot \,\rangle e_i$$

where we have also already assumed the eigenvalues $\lambda_i \geq 0$. Note that by theorem **i)** from section 4.2 of this Guide (or Theorem 7.7 on page 5 of Chapter 7), all eigenvalues are real.

Take, for example, $B \in \mathrm{Mat}(n, K)$ to be the self-adjoint matrix (i.e. endomorphism) which can be represented as

$$B = \sum_i \sqrt{\lambda_i} \langle e_i \,, \, \cdot \,\rangle e_i \,, \tag{7}$$

where $\sqrt{\lambda_i}$ means 'the square root of the eigenvalue associated with the ith eigenvector in the expression for A.' This (non-negative) square root always exists, because we assumed all the eigenvalues to be non-negative. Clearly, of course, the orthonormal basis of eigenvectors $\{e_i\}_i$ is the same in both expressions.

Now, consider $B^2 = B \circ B$ (as we are treating B as an endomorphism). Let us substitute in our expression for B and explicitly calculate B^2 in this manner:

$$
\begin{aligned}
B^2 = B \circ B &= \sum_i \sqrt{\lambda_i} \left\langle e_i \quad, \quad \sum_k \sqrt{\lambda_k} \langle e_k \,, \, \cdot \,\rangle e_k \right\rangle e_i \\
&= \sum_i \sum_k \sqrt{\lambda_i} \sqrt{\lambda_k} \langle e_k \,, \, \cdot \,\rangle \langle e_i \,, \, e_k \,\rangle e_i \quad \text{(using linear second slot of the inner product)} \\
&= \sum_i \sum_k \sqrt{\lambda_i \lambda_k} \langle e_k \,, \, \cdot \,\rangle \langle e_i \,, \, e_k \,\rangle e_i \\
&= \sum_i \sum_k \sqrt{\lambda_i \lambda_k} \langle e_k \,, \, \cdot \,\rangle \delta_{ik} \, e_i \quad \text{(by definition of orthonormal basis)} \\
&= \sum_i \sqrt{\lambda_i^2} \langle e_i \,, \, \cdot \,\rangle e_i \quad \text{(by definition of } \delta_{ik}) \\
&= \sum_i \lambda_i \langle e_i \,, \, \cdot \,\rangle e_i \\
&= A \,.
\end{aligned}
$$

So, we have found one particular $B \in \mathrm{Mat}(n, K)$, given by equation (7) above, which satisfies $B^2 = A$. Thus, we have found a 'square root' of the original self-adjoint matrix A!

Aside

Note that this 'square root' matrix that we have found is not unique. We could just as easily have taken

$$-B = \sum_i -\sqrt{\lambda_i} \langle e_i, \cdot \rangle e_i$$

to be the matrix we used instead. Upon evaluating $(-B)^2$ (and using "$-1 \times -1 = 1$", which was a basic result about fields from Problem Sheet 2), one can see that it too is equal to A. Thus, there exist two square roots of the original self-adjoint matrix A.

Application: Determinants of Binary Quadratic Forms

A binary quadratic form is an expression of the type

$$q(x, y) = a\,x^2 + 2b\,xy + c\,y^2$$

(where a, b and c are constants). From now onwards, let us assume that a, b and c are all real numbers.

This binary quadratic form may be written as a matrix product:

$$a\,x^2 + 2b\,xy + c\,y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

If you cannot see why these expressions are equivalent, then multiply the expression on the right hand side out and see what you get!

Now, the matrix in the centre of the above matrix product,

$$A_q = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

is clearly a self-adjoint matrix (it is easy to see that it is a symmetric real matrix). The matrix A_q represents the binary quadratic form $q(x, y)$ in a similar way to how a matrix of an inner product represents that inner product, or how a matrix of an endomorphism represents that endomorphism. Thus we can call A_q the matrix of the binary quadratic form q; we can swap between a binary quadratic form and its associated matrix as we like.

As A_q is self-adjoint, we can diagonalise it in terms of its eigenvalues, **by applying the Spectral Theorem**. Hence we arrive at

$$A_q = C \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} C^{\mathrm{T}}$$

where λ_1 and λ_2 are the eigenvalues of A_q, and C is an appropriate orthogonal matrix. Note that, as C is orthogonal, we have $C^{\mathrm{T}} = C^{-1}$ and so

$$\det(C^{\mathrm{T}}) = \det(C^{-1}) = {}^{1}/_{\det(C)}.$$

Now, there are certain occasions when we wish to calculate the determinant of the matrix of a binary quadratic form - as it will be mathematically useful to us for some reason. Then, in fact, this is nothing more than the product of its eigenvalues:

$$\det(A_q) = \det(C)\det\left(\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\right)\det(C^{\mathrm{T}})$$
$$= \frac{\det(C)}{\det(C)}\lambda_1\lambda_2$$
$$= \lambda_1\lambda_2.$$

So, if we ever need to calculate the determinant $\Delta_q = \det(A_q)$ of a binary quadratic form q, what we are really doing is computing the product of the eigenvalues of q. Hence, any statement that this determinant is greater than/less than/equal to 0 can be understood as a statement about the sign or value of q's eigenvalues:

If $\Delta_q > 0$ then both eigenvalues have the same sign; if $\Delta_q = 0$ then one of the eigenvalues is 0; if $\Delta_q < 0$ then the two eigenvalues have opposite signs.

In fact, you will already have come across a couple of situations where you have been calculating the determinant of a binary quadratic form - perhaps without knowing it! These were in your first year modules 'Geometry, Vectors & Complex Numbers' and 'Calculus'. So we see that theory from 'Vector Spaces' was (perhaps only implicitly) helping us out in other areas of maths:

1) Recall that in geometry a conic section is a curve obtained from the intersection of a plane with a cone. A general conic section takes the form

$$a\,x^2 + 2b\,xy + c\,y^2 + \ldots = 0 \quad \text{(where } a, b, c \in \mathbb{R} \text{ cannot all be zero).}$$

This is clearly a binary quadratic form plus some lower order terms. The binary quadratic form is of interest to us, because that is what geometrically classifies our conic section as an ellipse, a parabola or a hyperbola.

Take the binary quadratic form from your conic section and write its associated matrix A_q. Now, $\Delta_q = ac - b^2 = -\frac{1}{4}\Delta$, where $\Delta = 4b^2 - 4ac$ is the discriminant of the conic section. We know that Δ is used to determine whether our conic section is an ellipse, a parabola or a hyperbola. But, by the above relation, what we're really saying is that the eigenvalues of the quadratic form in the conic are what determines the type of the conic!

Hence:

$$
\text{The conic section is}
\begin{cases}
\text{an ellipse} & \text{if } \Delta_q > 0 \\
\text{(a circle} & \text{if also } b = 0 \text{ and } a = c) \\
\\
\text{a parabola} & \text{if } \Delta_q = 0 \\
\\
\text{a hyperbola} & \text{if } \Delta_q < 0
\end{cases}
$$

or, equivalently,

$$
\text{the conic section is}
\begin{cases}
\text{an ellipse} & \text{if both the eigenvalues of } q \text{ have the same sign} \\
\text{(a circle} & \text{if there is only one - repeated - eigenvalue, } \lambda = a) \\
\\
\text{a parabola} & \text{if one of the eigenvalues is } 0 \\
\\
\text{a hyperbola} & \text{if the eigenvalues have different signs}
\end{cases}
$$

2) Recall that in real calculus a differentiable function of two variables $f(x, y)$, whose second order partial derivatives exist and are continuous, can be expanded about the point $(0, 0)$ by the following Taylor Series expansion:

$$
f(x, y) = f(0, 0) + \begin{pmatrix} x & y \end{pmatrix} \nabla f + \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} H(x, y) \begin{pmatrix} x \\ y \end{pmatrix} + \dots, \quad (8)
$$

where ∇f is the gradient of f and

$$
H(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}
$$

is the Hessian Matrix, comprised of the second order partial derivatives of f.

The Hessian Matrix can be viewed as the matrix of the binary quadratic form which we see in (8) - the matrix is symmetric because the mixed partial derivatives are equal for the f we are considering (this was known as Clairaut's Theorem in 'Calculus' or 'Analysis'). We can ignore the $1/2$ which appears in the series expansion, because we could always eliminate it by appropriately changing coordinates.

Hence $H(x, y)$ is self-adjoint.

For critical points of the function $f(x, y)$ (points (a, b) such that $\nabla f \,|_{(a,b)} = (0, 0)$), we know how to perform the 'second derivative test' to work out if the critical point is a local minimum, a local maximum or a saddle point of f. This involves calculating

$$
D(x, y) = f_{xx}(x, y)\, f_{yy}(x, y) - f_{yx}(x, y)\, f_{xy}(x, y)
$$

then evaluating $D(a, b)$, and checking whether the result is greater than/equal to/less than 0.

However, note that $D(x, y)$ is nothing more than det($H(x, y)$). So our second derivative test boils down to no more than calculating the determinant of the Hessian Matrix, which we know to be self-adjoint. Hence, **by the Spectral Theorem**, we can diagonalise $H(x, y)$ in terms of its eigenvalues!

We obtain an expression of the form

$$H(x, y) = C \begin{pmatrix} \lambda_1(x, y) & 0 \\ 0 & \lambda_2(x, y) \end{pmatrix} C^{\mathrm{T}}$$

and, by similar working out to in the previous item, we obtain

$$D(a, b) = \det(H(a, b)) = \ \ldots \ = \lambda_1(a, b) \, \lambda_2(a, b) \,.$$

Therefore, the 'second derivative test' is nothing more than calculating the product of the eigenvalues of the Hessian Matrix!

You will probably have encountered the second derivative test before this moment in a form similar to the following:

The critical point (a, b) is $\begin{cases} \text{a local minimum of } f & \text{if } D(a, b) > 0 \text{ and } f_{xx}(a, b) > 0 \\[2ex] \text{a local maximum of } f & \text{if } D(a, b) > 0 \text{ and } f_{xx}(a, b) < 0 \\[2ex] \text{a saddle point of } f & \text{if } D(a, b) < 0 \\[2ex] \text{any of the above (test is inconclusive)} & \text{if } D(a, b) = 0 \end{cases}$

However, we now know that an equivalent version of the second derivative test, in terms of the eigenvalues of the Hessian Matrix, is given by:

(a, b) is $\begin{cases} \text{a local minimum of } f & \text{if both eigenvalues of } H(a, b) \text{ are positive} \\[2ex] \text{a local maximum of } f & \text{if both eigenvalues are negative} \\[2ex] \text{a saddle point of } f & \text{if the two eigenvalues have different signs} \\[2ex] \text{any of the above (test is inconclusive)} & \text{if one of the eigenvalues is 0} \end{cases}$

In fact, this approach (in terms of eigenvalues of the Hessian Matrix) has a generalisation to functions of more than just two variables. You need to know whether all eigenvalues are positive, all are negative, or some positive and some negative, at your particular critical point (a, b) in order to determine which type of critical point it is.

Overall, the examples given throughout section 5.4 show us that knowledge of self-adjoint matrices and the Spectral Theorem can allow us to gain a deeper understanding of the mathematics we have already encountered in our degree! Thus, the content in Chapter 7 of 'Vector Spaces' is both powerful and applicable in a wide variety of ways.

Conclusion

This Companion Guide for the final chapter of 'Vector Spaces' (Chapter 7 - "Self-Adjoint Maps") has worked through each part of the chapter in detail, breaking down the mathematical content into smaller chunks and explaining how each given result follows from previous material throughout the course, whether it be from Chapter 7 itself or earlier chapters. Where appropriate, examples have been given and comments made so as to develop understanding in possibly problematic areas.

I hope that this Companion Guide has enabled you to better understand the chapter as a whole, as well as its place within the Vector Spaces course - and indeed how it might link to other areas of mathematics. If you remain unsure about some aspect of the chapter, please feel free to discuss it with your (other) PAL leaders - they are there to help! - or with the module lecturer. I am also happy to try and answer any questions you may have: email me at `j.tabeart-10@student.lboro.ac.uk`

Jack Tabeart

Acknowledgements and References

I would like to thank the following lecturers at Loughborough University for providing highly useful feedback on an earlier draft of this Guide:

Dr Alexander Strohmaier, Dr Thomas Bartsch and Professor Tony Croft.

In addition to my own understanding & explanations of the mathematical concepts involved in Chapter 7, and my own generation of examples to use within the Guide, I occasionally drew upon other external resources when creating this document.

To make the Guide, I used or adapted material from the following sources:

- Dr Alexander Strohmaier's 'Vector Spaces' lecture notes, 2011-12 version.

- `http://www.math.upenn.edu/~kazdan/504/la.pdf` (accessed: 25/07/13) - the questions here provided inspiration for section 5.4's Fibonacci Sequence example.

- A variety of Wikipedia articles from `http://www.en.wikipedia.org` (accessed between 20/07/13 and 03/08/13) - primarily regarding the mathematical terminology associated with the Spectral Theorem, and the applications in section 5.4.