

## 2 Inner Product Spaces, part 1

You should be familiar with the *scalar/dot product* in  $\mathbb{R}^2$ . For any vectors  $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ ,  $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ , define

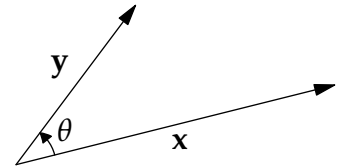
$$\mathbf{x} \cdot \mathbf{y} := x_1 y_1 + x_2 y_2$$

The *norm* or *length* of a vector is  $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ .

The *angle*  $\theta$  between vectors satisfies  $\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta$ .

Vectors are *orthogonal* or *perpendicular* precisely when  $\mathbf{x} \cdot \mathbf{y} = 0$ .

The dot product is what allows us to compute lengths of and angles between vectors in  $\mathbb{R}^2$ . An *inner product* is an algebraic structure that generalizes this idea to other vector spaces.



### 2.1 Real and Complex Inner Product Spaces

Unless explicitly stated otherwise, throughout this chapter  $\mathbb{F}$  is either the real  $\mathbb{R}$  or complex  $\mathbb{C}$  field.

**Definition 2.1.** An *inner product space*  $(V, \langle \cdot, \cdot \rangle)$  is a vector space  $V$  over  $\mathbb{F}$  together with an *inner product*: a function  $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{F}$  satisfying the following properties  $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V, \lambda \in \mathbb{F}$ ,

(a) *Linear*:  $\langle \lambda \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$

(b) *Conjugate-Symmetric*:  $\langle \mathbf{y}, \mathbf{x} \rangle = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$  (complex conjugate!)

(c) *Positive-definite*:  $\mathbf{x} \neq \mathbf{0} \implies \langle \mathbf{x}, \mathbf{x} \rangle > 0$

The *norm* or *length* of  $\mathbf{x} \in V$  is  $\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ . A *unit vector* has  $\|\mathbf{x}\| = 1$ .

Vectors  $\mathbf{x}, \mathbf{y}$  are *perpendicular/orthogonal* if  $\langle \mathbf{x}, \mathbf{y} \rangle = 0$  and *orthonormal* if they are additionally unit vectors.

#### Real inner product spaces

The definition simplifies slightly when  $\mathbb{F} = \mathbb{R}$ .

- Conjugate-symmetry becomes plain *symmetry*:  $\langle \mathbf{y}, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$ .
- Linearity + symmetry yields *bilinearity*: a real inner product is also linear in its second slot

$$\langle \mathbf{x}, \lambda \mathbf{y} + \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$$

A real inner product is often termed a *positive-definite, symmetric, bilinear form*.

The simplest example is the natural generalization of the dot product.

**Definition 2.2.** *Euclidean space* means  $\mathbb{R}^n$  equipped with the *standard inner (dot) product*,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \mathbf{y}^T \mathbf{x} = \sum_{j=1}^n x_j y_j = x_1 y_1 + \cdots + x_n y_n, \quad \|\mathbf{x}\| = \sqrt{\sum_{j=1}^n x_j^2}$$

Unless the inner product is stated explicitly, if we refer to  $\mathbb{R}^n$  as an inner product space, we mean Euclidean space. However, there are many other ways to make  $\mathbb{R}^n$  into an inner product space...

**Example 2.3.** Define an alternative inner product on  $\mathbb{R}^2$  via

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + 3x_2 y_2$$

It is easy to check that this satisfies the required properties:

(a) Linearity: follows from the associative/distributive laws in  $\mathbb{R}$ ,

$$\begin{aligned} \langle \lambda \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle &= (\lambda x_1 + y_1)z_1 + 3(\lambda x_2 + y_2)z_2 = \lambda(x_1 z_1 + 3x_2 z_2) + (y_1 z_1 + 3y_2 z_2) \\ &= \lambda \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle \end{aligned}$$

(b) Symmetry: follows from the commutativity of multiplication in  $\mathbb{R}$ ,

$$\langle \mathbf{y}, \mathbf{x} \rangle = y_1 x_1 + 3y_2 x_2 = x_1 y_1 + 3x_2 y_2 = \langle \mathbf{x}, \mathbf{y} \rangle$$

(c) Positive-definiteness: if  $\mathbf{x} \neq \mathbf{0}$ , then  $\langle \mathbf{x}, \mathbf{x} \rangle = x_1^2 + 3x_2^2 > 0$ .

With respect to  $\langle \cdot, \cdot \rangle$ , the concept of *orthogonality* feels strange: e.g.,  $\left\{ \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{2\sqrt{3}} \begin{pmatrix} 3 \\ -1 \end{pmatrix} \right\}$  is an orthonormal set!

$$\|\mathbf{x}\|^2 = \frac{1}{4}(1^2 + 3 \cdot 1^2) = 1, \quad \|\mathbf{y}\|^2 = \frac{1}{12}(3^2 + 3 \cdot 1^2) = 1, \quad \langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{4\sqrt{3}}(3 - 3) = 0$$

However, with respect to the standard dot product, these vectors are not special:

$$\mathbf{x} \cdot \mathbf{x} = \frac{1}{2}, \quad \mathbf{y} \cdot \mathbf{y} = \frac{5}{6}, \quad \mathbf{x} \cdot \mathbf{y} = \frac{1}{2\sqrt{3}}$$

We have the same *vector space*  $\mathbb{R}^2$ , but different *inner product spaces*:  $(\mathbb{R}^2, \langle \cdot, \cdot \rangle) \neq (\mathbb{R}^2, \cdot)$ .

The above is an example of a *weighted inner product*: choose *weights*  $a_1, \dots, a_n \in \mathbb{R}^+$  and define

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^n a_j x_j y_j = a_1 x_1 y_1 + \dots + a_n x_n y_n$$

It is a simple exercise to check that this defines an inner product on  $\mathbb{R}^n$ . In particular,  $\mathbb{R}^n$  may be equipped with *infinitely many distinct inner products*!

More generally, a symmetric matrix  $A \in M_n(\mathbb{R})$  is *positive-definite* if  $\mathbf{x}^T A \mathbf{x} > 0$  for all non-zero  $\mathbf{x} \in \mathbb{R}^n$ . It is straightforward to check that

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{y}^T A \mathbf{x}$$

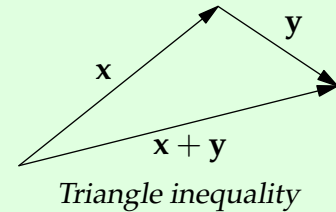
defines an inner product on  $\mathbb{R}^n$ . In fact *all* inner products on  $\mathbb{R}^n$  arise in this fashion! The weighted inner products correspond to  $A$  being diagonal (Euclidean space is  $A = I$ ), but this is not required.

**Example 2.4.** The matrix  $A = \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}$  is positive-definite and thus defines an inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = 3x_1 y_1 + x_1 y_2 + x_2 y_1 + x_2 y_2$$

**Lemma 2.5 (Basic properties).** Let  $V$  be an inner product space, let  $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$  and  $\lambda \in \mathbb{F}$ .

1.  $\langle \mathbf{0}, \mathbf{x} \rangle = 0$
2.  $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$
3.  $\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$
4.  $\langle \mathbf{x}, \mathbf{z} \rangle = \langle \mathbf{y}, \mathbf{z} \rangle$  for all  $\mathbf{z} \implies \mathbf{x} = \mathbf{y}$
5. (Cauchy–Schwarz inequality)  $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$  with equality if and only if  $\mathbf{x}, \mathbf{y}$  are parallel
6. (Triangle Inequality)  $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$  with equality if and only if  $\mathbf{x}, \mathbf{y}$  are parallel and point in the same direction



Be careful with notation:  $|\lambda|$  means the *absolute value/modulus* (of a scalar), while  $\|\mathbf{x}\|$  means the *norm* (of a vector).

In the real case, Cauchy–Schwarz allows us to define *angle* via  $\cos \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$ , since the right-hand side lies in the interval  $[-1, 1]$ . However, except in Euclidean  $\mathbb{R}^2$  and  $\mathbb{R}^3$ , this notion is of limited use; *orthogonality* ( $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ ) and *orthonormality* are usually all we care about.

*Proof.* Parts 1–3 are exercises. For simplicity, we prove 5 and 6 only when  $\mathbb{F} = \mathbb{R}$ .

4. Let  $\mathbf{z} = \mathbf{x} - \mathbf{y}$ , apply the linearity condition and part 2:

$$\langle \mathbf{x}, \mathbf{z} \rangle = \langle \mathbf{y}, \mathbf{z} \rangle \implies 0 = \langle \mathbf{x} - \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle = \|\mathbf{x} - \mathbf{y}\|^2 \implies \mathbf{x} = \mathbf{y}$$

5. If  $\mathbf{y} = \mathbf{0}$ , the result is trivial. WLOG (and by part 3) we may assume  $\|\mathbf{y}\| = 1$ ; if the inequality holds for this, then it holds for all non-zero  $\mathbf{y}$  by parts 1 and 4. Now expand:

$$\begin{aligned} 0 &\leq \|\mathbf{x} - \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{y}\|^2 && \text{(positive-definiteness)} \\ &= \|\mathbf{x}\|^2 + |\langle \mathbf{x}, \mathbf{y} \rangle|^2 \|\mathbf{y}\|^2 - \langle \mathbf{x}, \mathbf{y} \rangle \langle \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{x}, \mathbf{y} \rangle \langle \mathbf{y}, \mathbf{x} \rangle && \text{(bilinearity)} \\ &= \|\mathbf{x}\|^2 - |\langle \mathbf{x}, \mathbf{y} \rangle|^2 && \text{(symmetry)} \end{aligned}$$

Taking square-roots establishes the inequality.

By part 2, equality holds if and only if  $\mathbf{x} = \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{y}$ , which is precisely when  $\mathbf{x}, \mathbf{y}$  are linearly dependent.

6. We establish the squared result.

$$\begin{aligned} (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 - \|\mathbf{x} + \mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \|\mathbf{y}\| + \|\mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \langle \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{y}, \mathbf{x} \rangle - \|\mathbf{y}\|^2 \\ &= 2(\|\mathbf{x}\| \|\mathbf{y}\| - \langle \mathbf{x}, \mathbf{y} \rangle) \\ &\geq 2(\|\mathbf{x}\| \|\mathbf{y}\| - |\langle \mathbf{x}, \mathbf{y} \rangle|) \\ &\geq 0 && \text{(Cauchy–Schwarz)} \end{aligned}$$

Equality requires both equality in Cauchy–Schwarz ( $\mathbf{x}, \mathbf{y}$  parallel) and that  $\langle \mathbf{x}, \mathbf{y} \rangle \geq 0$ ; since  $\mathbf{x}, \mathbf{y}$  are already parallel, this means that one is a non-negative multiple of the other. ■

## Complex Inner Product Spaces

Definition 2.1 is already set up nicely when  $\mathbb{C} = \mathbb{F}$ . One subtle difference comes from how we expand linear combinations in the second slot.

**Lemma 2.6.** An inner product is a positive-definite, conjugate-symmetric, **sesquilinear**<sup>1</sup> form: it is conjugate-linear (anti-linear) in the second slot,

$$\langle \mathbf{z}, \lambda \mathbf{x} + \mathbf{y} \rangle = \bar{\lambda} \langle \mathbf{z}, \mathbf{x} \rangle + \langle \mathbf{z}, \mathbf{y} \rangle$$

The proof is very easy if you remember your complex conjugates; try it!

**Warning!** If you dabble in the dark arts of Physics, be aware that their convention<sup>2</sup> is for an inner product to be conjugate-linear in the *first* entry and linear in the second!

**Definition 2.7.** The standard (Hermitian) inner product and norm on  $\mathbb{C}^n$  are

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x} = \sum_{j=1}^n x_j \bar{y}_j = x_1 \bar{y}_1 + \cdots + x_n \bar{y}_n, \quad \|\mathbf{x}\| = \sqrt{\sum_{j=1}^n |x_j|^2}$$

where  $\mathbf{y}^* = \bar{\mathbf{y}}^T$  is the conjugate-transpose of  $\mathbf{y}$  and  $|x_j|$  is the modulus.

Weighted inner products may be defined as on page 2:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^n a_j x_j \bar{y}_j = a_1 x_1 \bar{y}_1 + \cdots + a_n x_n \bar{y}_n$$

Note that the weights  $a_j$  must still be *positive real numbers*. We may similarly define inner products in terms of positive-definite matrices  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{A} \mathbf{x}$ .

**Definition 2.8.** A matrix  $A \in M_n(\mathbb{C})$  is *Hermitian (self-adjoint)* if  $A^* = A$ . It is moreover *positive-definite* if  $\mathbf{x}^* \mathbf{A} \mathbf{x} > 0$  for all non-zero  $\mathbf{x} \in \mathbb{C}^n$ .

The self-adjoint condition reduces to symmetry ( $A^T = A$ ) if  $A$  is a real matrix.

**Example 2.9.** It can be seen (Exercise 6) that  $A = \begin{pmatrix} 3 & -i \\ i & 3 \end{pmatrix}$  is positive-definite, whence

$$\langle \mathbf{x}, \mathbf{y} \rangle = \begin{pmatrix} \bar{y}_1 & \bar{y}_2 \end{pmatrix} \begin{pmatrix} 3 & -i \\ i & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 3x_1 \bar{y}_1 + ix_1 \bar{y}_2 - ix_2 \bar{y}_1 + 3x_2 \bar{y}_2$$

defines an inner product on  $\mathbb{C}^2$ .

Almost all results in this chapter will be written for general inner product spaces, thus covering the real and complex cases simultaneously. If you don't feel confident with complex numbers, simply let  $\mathbb{F} = \mathbb{R}$  and delete all complex conjugates at first read! Very occasionally a different proof will be required depending on the field. For simplicity, examples will more often use real inner products.

<sup>1</sup>The prefix *sesqui-* means *one-and-a-half*; for instance a *sesquicentenary* is a 150 year anniversary.

<sup>2</sup>The common Physics notation relates to ours via  $\langle \mathbf{x} | \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ .

## Further Examples

As before, the field  $\mathbb{F}$  must be either  $\mathbb{R}$  or  $\mathbb{C}$ .

**Definition 2.10 (Frobenius inner product).** If  $A, B \in M_{m \times n}(\mathbb{F})$ , define

$$\langle A, B \rangle = \operatorname{tr}(B^* A)$$

where  $\operatorname{tr}$  is the *trace* of an  $n \times n$  matrix; this makes  $M_{m \times n}(\mathbb{F})$  into an inner product space.

This isn't really a new example: if we map  $A \rightarrow \mathbb{F}^{m \times n}$  by stacking the columns of  $A$ , then the Frobenius inner product is the standard inner product in disguise.

**Example 2.11.** In  $M_{3 \times 2}(\mathbb{C})$ ,

$$\left\langle \begin{pmatrix} 1 & i \\ 2-i & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 7 \\ 1 & 2i \\ 3-2i & 4 \end{pmatrix} \right\rangle = \operatorname{tr} \begin{pmatrix} 0 & 1 & 3+2i \\ 7 & -2i & 4 \end{pmatrix} \begin{pmatrix} 1 & i \\ 2-i & 0 \\ 0 & -1 \end{pmatrix} = \operatorname{tr} \begin{pmatrix} 2-i & -3-2i \\ 9-4i & -4+7i \end{pmatrix} \\ = 2-i-4+7i = -2-8i$$

$$\left\| \begin{pmatrix} 1 & i \\ 2-i & 0 \\ 0 & -1 \end{pmatrix} \right\|^2 = \operatorname{tr} \begin{pmatrix} 1 & 2+i & 0 \\ -i & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & i \\ 2-i & 0 \\ 0 & -1 \end{pmatrix} = \operatorname{tr} \begin{pmatrix} 6 & i \\ -i & 2 \end{pmatrix} = 8$$

**Definition 2.12 ( $L^2$  inner product).** Given a *real* interval  $[a, b]$ , the function

$$\langle f, g \rangle := \int_a^b f(t) \overline{g(t)} dt$$

defines an inner product on the space  $C[a, b]$  of continuous functions  $f : [a, b] \rightarrow \mathbb{F}$ .

With careful restriction, this works even for infinite intervals and a larger class of functions.<sup>3</sup> Verifying the required properties is straightforward if you know a little analysis; for instance continuity allows us to conclude

$$\|f\|^2 = \int_a^b |f(x)|^2 dx = 0 \iff f(x) \equiv 0$$

This is our first example of an *infinite-dimensional* inner product space.

**Example 2.13.** Let  $f(x) = x$  and  $g(x) = x^2$ ; these lie in the inner product space  $C[-1, 1]$  with respect to the  $L^2$  inner product.

$$\langle f, g \rangle = \int_{-1}^1 x^3 dx = 0, \quad \|f\|^2 = \int_{-1}^1 x^2 dx = \frac{2}{3}, \quad \|g\|^2 = \int_{-1}^1 x^4 dx = \frac{2}{5}$$

With some simple scaling, we see that  $\left\{ \frac{1}{\|f\|} f, \frac{1}{\|g\|} g \right\} = \left\{ \sqrt{\frac{3}{2}} x, \sqrt{\frac{5}{2}} x^2 \right\}$  forms an orthonormal set.

<sup>3</sup>For us, functions will always be continuous (often polynomials) on closed bounded intervals. The *square-integrable functions* and  $L^2$ -spaces for which the inner product is named are a more complicated business and beyond this course.

**Definition 2.14 ( $\ell^2$  inner product).** A sequence  $(x_n)$  is *square-summable* if  $\sum_{n=1}^{\infty} |x_n|^2 < \infty$ . These sequences form a vector space on which we can define an inner product<sup>4</sup>

$$\langle (x_n), (y_n) \rangle = \sum_{n=1}^{\infty} x_n \overline{y_n}$$

In essence we've taken the standard inner product on  $\mathbb{F}^n$  and let  $n \rightarrow \infty$ ! This example, and its  $L^2$  cousin, are the prototypical *Hilbert spaces*, which have great application to differential equations, signal processing, etc. Since a rigorous discussion requires a significant amount of analysis (convergence of series, completeness, integrability), these objects are generally beyond the course.

Our final example of an inner product is a useful, and hopefully obvious, hack to which we shall repeatedly appeal in examples.

**Lemma 2.15.** Let  $V$  be a vector space over  $\mathbb{R}$  or  $\mathbb{C}$ . If  $\beta$  is a basis of  $V$ , then there exists exactly one inner product on  $V$  for which  $\beta$  is an orthonormal set.

**Exercises 2.1** 1. Evaluate the inner product of the given vectors.

(a)  $\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$  where  $\langle \mathbf{x}, \mathbf{y} \rangle = 2x_1y_1 + 3x_2y_2 + x_3y_3$

(b)  $\mathbf{x} = \begin{pmatrix} 1 \\ 2i \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 5i \\ 4 \end{pmatrix}$  where  $\langle \mathbf{x}, \mathbf{y} \rangle$  is the standard Hermitian inner product on  $\mathbb{C}^2$

(c)  $\mathbf{x} = \begin{pmatrix} 1 \\ 2i \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 5i \\ 4 \end{pmatrix}$  where  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \begin{pmatrix} 2 & -i \\ i & 2 \end{pmatrix} \mathbf{x}$

(d)  $f(x) = x - 1, g(x) = x + 1$  where  $\langle f, g \rangle$  is the  $L^2$  inner product on  $C[0, 2]$

2. Suppose  $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle$  for all  $\mathbf{x}$ . Prove that  $\mathbf{y} = \mathbf{z}$ .

3. For each  $\mathbf{z} \in V$ , The linearity condition says that the map  $T_{\mathbf{z}} : V \rightarrow \mathbb{F}$  defined by

$$T_{\mathbf{z}}(\mathbf{x}) = \langle \mathbf{x}, \mathbf{z} \rangle$$

is linear ( $T_{\mathbf{z}}$  is an element of the *dual space*  $V^*$ ). What, if anything, can you say about the function  $U_{\mathbf{z}} : \mathbf{x} \mapsto \langle \mathbf{z}, \mathbf{x} \rangle$ ?

4. Define  $\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{j=1}^n x_j y_j$  on  $\mathbb{C}^n$ . Is this an inner product? Which of the properties (a), (b), (c) from Definition 2.1 does it satisfy?

5. (a) Verify that the matrix in Example 2.4 is positive-definite and that  $\langle \mathbf{x}, \mathbf{y} \rangle = 3x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2$  therefore defines an inner product on  $\mathbb{R}^2$ .  
(Hint: Try to write  $\|\mathbf{x}\|^2$  as a sum of squares...)

(b) Let  $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ . With respect to the inner product in part (a), find a non-zero unit vector  $\mathbf{y}$  which is *orthogonal* to  $\mathbf{x}$ .

6. By multiplying out  $|x_1 - ix_2|^2$ , show that the matrix  $\begin{pmatrix} 3 & -i \\ i & 3 \end{pmatrix}$  is positive-definite.

(Hint: recall that  $|a|^2 = a\bar{a}$  for complex numbers!!!)

<sup>4</sup>Neither of these facts are obvious; in particular, we'd need to see that the sum of two square-summable sequences is also square-summable.

7. Show that every eigenvalue of a positive definite matrix is positive.

8. Prove parts 1, 2 and 3 of Lemma 2.5.

9. Let  $V$  be an inner product space, prove Pythagoras' Theorem:

$$\text{If } \langle \mathbf{x}, \mathbf{y} \rangle = 0, \text{ then } \|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

10. Use basic algebra to prove the Cauchy–Schwarz inequality for vectors  $\mathbf{x} = \begin{pmatrix} a \\ b \end{pmatrix}$  and  $\mathbf{y} = \begin{pmatrix} c \\ d \end{pmatrix}$  in  $\mathbb{R}^2$  with the standard (dot) product.

11. Prove the Cauchy–Schwarz and triangle inequalities for a *complex* inner product space. What has to change compared to the proof of Lemma 2.5?

12. Prove the polarization identities:

(a) In any real inner product space:  $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{4} \|\mathbf{x} + \mathbf{y}\|^2 - \frac{1}{4} \|\mathbf{x} - \mathbf{y}\|^2$

(b) In any complex inner product space:  $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{4} \sum_{k=1}^4 i^k \|\mathbf{x} + i^k \mathbf{y}\|^2$

*If you know the length of every vector, then you know the inner product!*

13. Prove that  $\int_0^2 \frac{\sqrt{x}}{x+1} dx \leq \frac{2}{\sqrt{3}}$

*(Hint: use Cauchy–Schwarz)*

14. Let  $m \in \mathbb{Z}$  and consider the complex-valued function  $f_m(x) = \frac{1}{\sqrt{2\pi}} e^{imx}$ . If  $\langle \cdot, \cdot \rangle$  is the  $L^2$  inner product on  $C[-\pi, \pi]$ , prove that  $\{f_m : m \in \mathbb{Z}\}$  is an orthonormal set.

*This example is central to the study of Fourier series.*

*(Hint: If complex functions are scary, use Euler's formula  $e^{imx} = \cos mx + i \sin mx$  and work with the real-valued functions  $\cos mx$  and  $\sin mx$ . The difficulty is that you then need integration by parts...)*

15. Let  $\langle \cdot, \cdot \rangle$  be an inner product on  $\mathbb{F}^n$  (recall that  $\mathbb{F} = \mathbb{R}$  or  $\mathbb{C}$ ). Define the matrix  $A \in M_n(\mathbb{F})$  by  $A_{jk} = \langle \mathbf{e}_k, \mathbf{e}_j \rangle$  where  $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  is the standard basis. Verify that  $A$  is the *matrix of the inner product*:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{F}^n, \langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* A \mathbf{x}$$

In particular,

- $A$  is a *Hermitian/self-adjoint* matrix:  $A^* = A$  (if  $\mathbb{F} = \mathbb{R}$  this is simply *symmetric*);
- $A$  is *positive-definite*: for all  $\mathbf{x} \in \mathbb{F}^n$ ,  $\mathbf{x}^* A \mathbf{x} > 0$ ;

*More generally, if  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  is a basis then  $A_{jk} = \langle \mathbf{v}_k, \mathbf{v}_j \rangle$  defines the matrix of the inner product with respect to  $\beta$ :*

$$\langle \mathbf{x}, \mathbf{y} \rangle = [\mathbf{y}]_{\beta}^* A [\mathbf{x}]_{\beta}$$

## 2.2 Orthogonal Sets and the Gram–Schmidt Process

We start with a simple definition, relating subspaces to orthogonality.

**Definition 2.16.** Let  $U$  be a subspace of an inner product space  $V$ . The *orthogonal complement*  $U^\perp$  is the set

$$U^\perp = \{\mathbf{x} \in V : \forall \mathbf{u} \in U, \langle \mathbf{x}, \mathbf{u} \rangle = 0\}$$

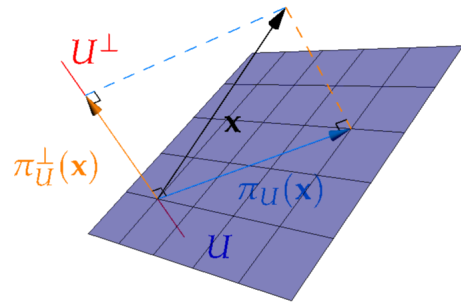
It is easy to check that  $U^\perp$  is itself a subspace of  $V$  and that  $U \cap U^\perp = \{\mathbf{0}\}$ . It can moreover be seen that  $U \subseteq (U^\perp)^\perp$ , though equality need not hold in infinite dimensions (see Exercise 7).

**Example 2.17.**  $U = \text{Span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix} \right\} \leq \mathbb{R}^3$  has orthogonal complement  $U^\perp = \text{Span} \left\{ \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} \right\}$ .

As in the example, we often have a direct sum decomposition  $V = U \oplus U^\perp$ : otherwise said,

$$\forall \mathbf{x} \in V, \exists \text{ unique } \mathbf{u} \in U, \mathbf{w} \in U^\perp \text{ such that } \mathbf{x} = \mathbf{u} + \mathbf{w}$$

In such a case, we'll call  $\mathbf{u} = \pi_U(\mathbf{x})$  and  $\mathbf{w} = \pi_{U^\perp}(\mathbf{x})$  the *orthogonal projections* of  $\mathbf{x}$  onto  $U$  and  $U^\perp$  respectively. As our first result shows, these are easy to compute whenever  $U$  has a finite orthogonal basis.



**Theorem 2.18.** Let  $V$  be an inner product space and let  $U = \text{Span } \beta$  where  $\beta = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  is an orthogonal set of non-zero vectors;

$$\langle \mathbf{u}_j, \mathbf{u}_k \rangle = \begin{cases} 0 & \text{if } j \neq k \\ \|\mathbf{u}_j\|^2 \neq 0 & \text{if } j = k \end{cases}$$

Then:

1.  $\beta$  is a basis of  $U$  and each  $\mathbf{x} \in U$  has unique representation

$$\mathbf{x} = \sum_{j=1}^n \frac{\langle \mathbf{x}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j \quad (*)$$

This simplifies to  $\mathbf{x} = \sum \langle \mathbf{x}, \mathbf{u}_j \rangle \mathbf{u}_j$  if  $\beta$  is an orthonormal set.

2.  $V = U \oplus U^\perp$ . For any  $\mathbf{x} \in V$ , we may write  $\mathbf{x} = \mathbf{u} + \mathbf{w}$  where

$$\mathbf{u} = \sum_{j=1}^n \frac{\langle \mathbf{x}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j \in U \quad \text{and} \quad \mathbf{w} = \mathbf{x} - \mathbf{u} \in U^\perp$$

Observe that (\*) essentially calculates the co-ordinate vector  $[\mathbf{x}]_\beta \in \mathbb{F}^n$ . Recalling how unpleasant such calculations have been in the past, often requiring large matrix inversions, we immediately see the power of inner products and orthogonal bases.



*Proof.* 1. Since  $\beta$  spans  $U$ , a given  $\mathbf{x} \in U$  may be written

$$\mathbf{x} = \sum_{k=1}^n a_k \mathbf{u}_k$$

for some scalars  $a_k$ . The orthogonality of  $\beta$  recovers the required expression for  $a_j$ :

$$\langle \mathbf{x}, \mathbf{u}_j \rangle = \sum_{k=1}^n a_k \langle \mathbf{u}_k, \mathbf{u}_j \rangle = a_j \|\mathbf{u}_j\|^2$$

Finally, let  $\mathbf{x} = \mathbf{0}$  to see that  $\beta$  is linearly independent.

2. Clearly  $\mathbf{u} \in U$ . For each  $\mathbf{u}_k$ , the orthogonality of  $\beta$  tells us that

$$\langle \mathbf{w}, \mathbf{u}_k \rangle = \langle \mathbf{x} - \mathbf{u}, \mathbf{u}_k \rangle = \langle \mathbf{x}, \mathbf{u}_k \rangle - \sum_{j=1}^n \frac{\langle \mathbf{x}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \langle \mathbf{u}_j, \mathbf{u}_k \rangle = \langle \mathbf{x}, \mathbf{u}_k \rangle - \langle \mathbf{x}, \mathbf{u}_k \rangle = 0$$

Since  $\mathbf{w}$  is orthogonal to a basis of  $U$  it is orthogonal to any element of  $U$ ; we conclude that  $\mathbf{w} \in U^\perp$ . Finally,  $U \cap U^\perp = \{\mathbf{0}\}$  forces the uniqueness of the decomposition  $\mathbf{x} = \mathbf{u} + \mathbf{w}$ , whence  $V = U \oplus U^\perp$ . ■

**Examples 2.19.** 1. Consider the standard orthonormal basis  $\beta = \{\mathbf{e}_1, \mathbf{e}_2\}$  of  $\mathbb{R}^2$ . For any  $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ , we easily check that

$$\sum_{j=1}^2 \langle \mathbf{x}, \mathbf{e}_j \rangle \mathbf{e}_j = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 = \mathbf{x}$$

2. In  $\mathbb{R}^3$ ,  $\beta = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \\ -5 \end{pmatrix} \right\}$  is an orthogonal set and thus a basis. We compute the co-ordinates of  $\mathbf{x} = \begin{pmatrix} 7 \\ 4 \\ 2 \end{pmatrix}$  with respect to  $\beta$ :

$$\begin{aligned} \mathbf{x} &= \sum_{j=1}^3 \frac{\langle \mathbf{x}, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j = \frac{7+8+6}{1+4+9} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \frac{14-4+0}{4+1+0} \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + \frac{21+24-10}{9+36+25} \begin{pmatrix} 3 \\ 6 \\ -5 \end{pmatrix} \\ &= \frac{3}{2} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 3 \\ 6 \\ -5 \end{pmatrix} \implies [\mathbf{x}]_\beta = \begin{pmatrix} 3/2 \\ 2 \\ 1/2 \end{pmatrix} \end{aligned}$$

Compare this with the painfully slow augmented matrix method for finding co-ordinates!

3. Revisiting Example 2.17, let  $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ . Since  $\beta = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} \right\}$  is an orthogonal basis of  $U$ , we observe that

$$\pi_U(\mathbf{x}) = \frac{1}{1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{2}{10} \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 5 \\ -1 \\ 3 \end{pmatrix}, \quad \pi_{U^\perp}(\mathbf{x}) = \mathbf{x} - \pi_U(\mathbf{x}) = \frac{2}{5} \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$$

are the orthogonal projections corresponding to  $\mathbb{R}^3 = U \oplus U^\perp$ .

## The Gram–Schmidt Process

Theorem 2.18 tells us how to compute the orthogonal projections corresponding to  $V = U \oplus U^\perp$ , provided  $U$  has a finite, orthogonal basis. Given how useful such bases are, our next goal is to see that such exist for any finite-dimensional subspace. Helpfully there exists a constructive algorithm.

**Theorem 2.20 (Gram–Schmidt).** Suppose  $S = \{\mathbf{s}_1, \dots, \mathbf{s}_n\}$  is a linearly independent subset of an inner product space  $V$ . Construct a sequence of vectors  $\mathbf{u}_1, \dots, \mathbf{u}_n$  inductively:

- Choose  $\mathbf{u}_1 = a_1 \mathbf{s}_1$  where  $a_1 \neq 0$
- For each  $k \geq 2$ , choose  $\mathbf{u}_k = a_k \left( \mathbf{s}_k - \sum_{j=1}^{k-1} \frac{\langle \mathbf{s}_k, \mathbf{u}_j \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j \right)$  where  $a_k \neq 0$  (†)

Then  $\beta := \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  is an orthogonal basis of  $\text{Span } S$ .

The purpose of the scalars  $a_k$  is to give you some freedom; choose them to avoid unpleasant fractions! If you want a set of *orthonormal* vectors, it is easier to scale everything after the algorithm is complete. Indeed, by taking  $S$  to be a basis of  $V$  and normalizing the resulting  $\beta$ , we conclude:

**Corollary 2.21.** Every finite-dimensional inner product space has an orthonormal basis.

**Example 2.22.**  $S = \{\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$  is a linearly independent subset of  $\mathbb{R}^3$ .

1. Choose  $\mathbf{u}_1 = \mathbf{s}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \implies \|\mathbf{u}_1\|^2 = 1$
2.  $\mathbf{s}_2 - \frac{\langle \mathbf{s}_2, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} - \frac{2}{1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix}$  : choose  $\mathbf{u}_2 = \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix} \implies \|\mathbf{u}_2\|^2 = 10$
3.  $\mathbf{s}_3 - \frac{\langle \mathbf{s}_3, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 - \frac{\langle \mathbf{s}_3, \mathbf{u}_2 \rangle}{\|\mathbf{u}_2\|^2} \mathbf{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{2}{10} \begin{pmatrix} 0 \\ -1 \\ 3 \end{pmatrix} = \frac{2}{5} \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$  : choose  $\mathbf{u}_3 = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$

The orthogonality of  $\beta = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  is clear. It is now trivial to observe that  $\left\{ \mathbf{u}_1, \frac{1}{\sqrt{10}} \mathbf{u}_2, \frac{1}{\sqrt{10}} \mathbf{u}_3 \right\}$  is an orthonormal basis of  $\mathbb{R}^3$ .

*Proof of Theorem 2.20.* For each  $k \leq n$ , define  $S_k = \{\mathbf{s}_1, \dots, \mathbf{s}_k\}$  and  $\beta_k = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ . We prove by induction that each  $\beta_k$  is an orthogonal set of non-zero vectors and that  $\text{Span } \beta_k = \text{Span } S_k$ . The Theorem is then the terminal case  $k = n$ .

(Base case  $k = 1$ ) Certainly  $\beta_1 = \{a_1 \mathbf{s}_1\}$  is an orthogonal set and  $\text{Span } \beta_1 = \text{Span } S_1$ .

(Induction step) Fix  $k \geq 2$ , assume  $\beta_{k-1}$  is an orthogonal non-zero set and that  $\text{Span } \beta_{k-1} = \text{Span } S_{k-1}$ . By Theorem 2.18,  $\mathbf{u}_k \in (\text{Span } \beta_{k-1})^\perp$ . We also see that  $\mathbf{u}_k \neq \mathbf{0}$ , for if not,

$$(\dagger) \implies \mathbf{s}_k \in \text{Span } \beta_{k-1} = \text{Span } S_{k-1}$$

and  $S$  would be linearly dependent. It follows that  $\beta_k$  is an orthogonal set of non-zero vectors.

Moreover,  $\mathbf{s}_k \in \text{Span } \beta_k \implies \text{Span } S_k \leq \text{Span } \beta_k$ . Since these spaces have the same (finite) dimension  $k$ , we conclude that  $\text{Span } \beta_k = \text{Span } S_k$ .

By induction,  $\beta$  is an orthogonal, non-zero spanning set for  $\text{Span } S$ ; by Theorem 2.18, it is a basis. ■

**Example 2.23.** This time we work in the space of polynomials  $P(\mathbb{R})$  equipped with the  $L^2$  inner product  $\langle f, g \rangle = \int_0^1 f(x)g(x) dx$  on the interval  $[0, 1]$ . Let  $S = \{1, x, x^2\}$  and apply the algorithm:

1. Choose  $f_1(x) = 1 \implies \|f_1\|^2 = \int_0^1 1 dx = 1$

2.  $x - \frac{\langle x, f_1 \rangle}{\|f_1\|^2} f_1 = x - \int_0^1 x dx = x - \frac{1}{2}$

We choose  $f_2(x) = 2x - 1$ , with  $\|f_2\|^2 = \int_0^1 (2x - 1)^2 dx = \frac{1}{3}$

3.  $x^2 - \frac{\langle x^2, f_1 \rangle}{\|f_1\|^2} f_1 - \frac{\langle x^2, f_2 \rangle}{\|f_2\|^2} f_2 = x^2 - \int_0^1 x^2 dx - \frac{\int_0^1 x^2(2x-1) dx}{1/3} (2x - 1) = x^2 - x + \frac{1}{6}$

We choose  $f_3(x) = 6x^2 - 6x + 1$  with  $\|f_3\|^2 = \int_0^1 (6x^2 - 6x + 1)^2 dx = \frac{1}{5}$

It follows that  $\text{Span } S$  has an orthonormal basis

$$\beta = \left\{ 1, \sqrt{3}(2x - 1), \sqrt{5}(6x^2 - 6x + 1) \right\}$$

This example can be extended to arbitrary degree since the countable set  $\{1, x, x^2, \dots\}$  is basis of  $P(\mathbb{R})$ . Indeed this shows that  $(P(\mathbb{R}), \langle \cdot, \cdot \rangle)$  has an orthonormal basis.

Gram–Schmidt also shows that our earlier discussion of orthogonal projections is generic.

**Corollary 2.24.** *If  $U$  is a finite-dimensional subspace of an inner product space  $V$ , then  $V = U \oplus U^\perp$  and the orthogonal projections may be computed as in Theorem 2.18.*

If  $U$  is an infinite-dimensional subspace of  $V$ , then we need not have  $V = U \oplus U^\perp$  and the orthogonal projections might not be well-defined (see, for example, Exercises 7 and 8). Instead, if  $\beta$  is an orthonormal basis of  $U$ , it is common to describe the coefficients  $\langle \mathbf{x}, \mathbf{u} \rangle$  for each  $\mathbf{u} \in \beta$  as the *Fourier coefficients*, and the infinite sum

$$\sum_{\mathbf{u} \in \beta} \langle \mathbf{x}, \mathbf{u} \rangle \mathbf{u}$$

as the *Fourier series* of  $\mathbf{x}$ , provided the sum converges.

**Exercises 2.2** 1. Apply Gram–Schmidt to obtain an orthogonal basis  $\beta$  for  $\text{Span } S$ . Then obtain the co-ordinate representation (Fourier coefficients) of the given vector with respect to  $\beta$ .

(a)  $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{R}^3, \quad \mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

(b)  $S = \left\{ \begin{pmatrix} 3 & 5 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 9 \\ 5 & -1 \end{pmatrix}, \begin{pmatrix} 7 & -17 \\ 2 & -6 \end{pmatrix} \right\} \subseteq M_2(\mathbb{R}), \quad X = \begin{pmatrix} -1 & 27 \\ -4 & 8 \end{pmatrix}$  (use the Frobenius product)

(c)  $S = \left\{ \begin{pmatrix} 1 \\ i \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ i \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{C}^3$  with  $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

(d)  $S = \{1, x, x^2\}$  with  $\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx$  and  $f(x) = x^2$ .

*Important!* You'll likely need *much* more practice than this to get comfortable with Gram–Schmidt; make up your own problems!

2. Let  $S = \{\mathbf{s}_1, \mathbf{s}_2\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \right\}$  and  $U = \text{Span } S \leq \mathbb{R}^3$ . Find  $\pi_U(\mathbf{x})$  if  $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \\ -2 \end{pmatrix}$ .  
(Hint: First apply Gram–Schmidt)
3. Find the orthogonal complement to  $U = \text{Span}\{x^2\} \leq P_2(\mathbb{R})$  with respect to the inner product  $\langle f, g \rangle = \int_0^1 f(t)g(t) dt$
4. Let  $T \in \mathcal{L}(V, W)$  where  $V, W$  are inner product spaces with orthonormal bases  $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  and  $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$  respectively. Prove that the matrix  $A = [T]_{\beta}^{\gamma} \in M_{m \times n}(\mathbb{F})$  of  $T$  with respect to these bases has  $jk^{\text{th}}$  entry
- $$A_{jk} = \langle T(\mathbf{v}_k), \mathbf{w}_j \rangle$$
5. Suppose that  $\beta$  is an orthonormal basis of an  $n$ -dimensional inner product space  $V$ . Prove that,
- $$\forall \mathbf{x}, \mathbf{y} \in V, \langle \mathbf{x}, \mathbf{y} \rangle = [\mathbf{y}]_{\beta}^* [\mathbf{x}]_{\beta}$$
- Otherwise said, the co-ordinate isomorphism  $\phi_{\beta} : V \rightarrow \mathbb{F}^n$  defined by  $\phi_{\beta}(\mathbf{x}) = [\mathbf{x}]_{\beta}$  is an isomorphism of inner product spaces where we use the standard inner product on  $\mathbb{F}^n$
6. Let  $U$  be a subspace of an inner product space  $V$ . Prove the following:
- $U^{\perp}$  is a subspace of  $V$ .
  - $U \cap U^{\perp} = \{\mathbf{0}\}$
  - $U \subseteq (U^{\perp})^{\perp}$
  - If  $V = U \oplus U^{\perp}$ , then  $U = (U^{\perp})^{\perp}$  (this is always the case when  $\dim U < \infty$ )
7. Let  $\ell^2$  be the set of square-summable sequences of real numbers (Definition 2.14). Consider the sequences  $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots$ , where  $\mathbf{u}_j$  is the zero sequence except for a single 1 in the  $j^{\text{th}}$  entry. For instance,
- $$\mathbf{u}_4 = (0, 0, 0, 1, 0, 0, 0, \dots)$$
- Let  $U = \text{Span}\{\mathbf{u}_j : j \in \mathbb{N}\}$ . Prove that  $U^{\perp}$  contains only the zero sequence.
  - Show that the sequence  $\mathbf{y} = (\frac{1}{n})$  lies in  $\ell^2$ , but does not lie in  $U$ .  
 $U$  is therefore a proper subset of  $(U^{\perp})^{\perp} = \ell^2$  and  $\ell^2 \neq U \oplus U^{\perp}$ .
8. Recall Exercise 2.1.14 where we saw that the set  $\beta = \{\frac{1}{\sqrt{2\pi}}e^{imx} : m \in \mathbb{Z}\}$  is orthonormal with respect to  $\langle f, g \rangle = \int_{-\pi}^{\pi} f(t)\overline{g(t)} dt$ .
- Show that the Fourier series of  $f(x) = x$  is
- $$\mathcal{F}(x) := \sum_{m=-\infty}^{\infty} \left\langle x, \frac{1}{\sqrt{2\pi}}e^{imx} \right\rangle \frac{1}{\sqrt{2\pi}}e^{imx} = \sum_{m=1}^{\infty} \frac{2(-1)^{m+1}}{m} \sin mx$$
- Briefly explain why the Fourier series is *not* an element of  $\text{Span } \beta$ .
  - Sketch a few of the Fourier approximations (sum up to  $m = 5$  or  $7 \dots$ ) and observe, when extended to  $\mathbb{R}$ , how they approximate a discontinuous periodic function.
9. (Hard) Much of Theorem 2.18 remains true, with suitable modifications, even if  $\beta$  is an *infinite* set. Restate and prove as much as you can, and identify the false part(s).

## 2.3 The Adjoint of a Linear Operator

Recall how the standard inner product on  $\mathbb{F}^n$  may be written in terms of the conjugate-transpose

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x} = \overline{\mathbf{y}}^T \mathbf{x}$$

We start by inserting a matrix into this expression and interpreting in two different ways. Suppose  $A \in M_{m \times n}(\mathbb{F})$ ,  $\mathbf{v} \in \mathbb{F}^n$  and  $\mathbf{w} \in \mathbb{F}^m$ , then

$$\underbrace{\langle A^* \mathbf{w}, \mathbf{v} \rangle}_{\text{in } \mathbb{F}^n} = \mathbf{v}^* (A^* \mathbf{w}) = (\mathbf{v}^* A^*) \mathbf{w} = (A \mathbf{v})^* \mathbf{w} = \underbrace{\langle \mathbf{w}, A \mathbf{v} \rangle}_{\text{in } \mathbb{F}^m} \quad (\dagger)$$

**Example 2.25.** As a sanity check, let  $A = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \in M_2(\mathbb{R})$ ,  $\mathbf{w} = \begin{pmatrix} x \\ y \end{pmatrix}$  and  $\mathbf{v} = \begin{pmatrix} p \\ q \end{pmatrix}$ . Then,

$$\begin{aligned} \left\langle A^T \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} x \\ 2x + 3y \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle = xp + (2x + 3y)q \\ \left\langle \begin{pmatrix} x \\ y \end{pmatrix}, A \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} p + 2q \\ 3q \end{pmatrix} \right\rangle = x(p + 2q) + 3yq \end{aligned}$$

Note how the inner products are evaluated on *different spaces*. At the level of linear maps this is a relationship between  $L_A \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$  and  $L_{A^*} \in \mathcal{L}(\mathbb{F}^m, \mathbb{F}^n)$ , one that is easily generalizable.

**Definition 2.26.** Let  $T \in \mathcal{L}(V, W)$  where  $V, W$  are inner product spaces over the same field  $\mathbb{F}$ . The *adjoint* of  $T$  is a function  $T^* : W \rightarrow V$  (read ‘ $T$ -star’) satisfying

$$\forall \mathbf{v} \in V, \mathbf{w} \in W, \quad \langle T^*(\mathbf{w}), \mathbf{v} \rangle = \langle \mathbf{w}, T(\mathbf{v}) \rangle$$

Note that the first inner product is computed within  $V$  and the second within  $W$ .

The adjoint effectively extends the conjugate-transpose to linear maps. We now use the same notation for three objects, so be careful!

- If  $A$  is a real or complex matrix, then  $A^* = \overline{A}^T$  is its *conjugate-transpose*.
- If  $T$  is a linear map, then  $T^*$  is its *adjoint*.
- If  $V$  is a vector space, then  $V^* = \mathcal{L}(V, \mathbb{F})$  is its *dual space*.

Thankfully the two notations line up nicely, as part 3 of our first result shows.

**Theorem 2.27 (Basic Properties).** 1. If an adjoint exists,<sup>5</sup> then it is unique and linear.

2. If  $T$  and  $S$  have adjoints, then

$$(T^*)^* = T, \quad (TS)^* = S^*T^*, \quad (\lambda T + S)^* = \overline{\lambda}T^* + S^*$$

3. Suppose  $V, W$  are finite-dimensional with orthonormal bases  $\beta, \gamma$  respectively. Then the matrix of the adjoint of  $T \in \mathcal{L}(V, W)$  is the conjugate-transpose of the original:  $[T^*]_{\gamma}^{\beta} = ([T]_{\beta})^*$ .

<sup>5</sup>Existence of adjoints is trickier, so we postpone this a little: see Corollary 2.34 and Exercise 12.

*Proof.* 1. (Uniqueness) Suppose  $T^*$  and  $S^*$  are adjoints of  $T$ . Then

$$\langle T^*(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, T(\mathbf{y}) \rangle = \langle S^*(\mathbf{x}), \mathbf{y} \rangle$$

Since this holds for all  $\mathbf{y}$ , Lemma 2.5 part 4 says that  $\forall \mathbf{x}$ ,  $T^*(\mathbf{x}) = S^*(\mathbf{x})$ , whence  $T^* = S^*$ .

(Linearity) Simply translate across, use the linearity of  $T$ , and again appeal to Lemma 2.5:

$$\begin{aligned} \forall \mathbf{z}, \langle T^*(\lambda \mathbf{x} + \mathbf{y}), \mathbf{z} \rangle &= \langle \lambda \mathbf{x} + \mathbf{y}, T(\mathbf{z}) \rangle = \lambda \langle \mathbf{x}, T(\mathbf{z}) \rangle + \langle \mathbf{y}, T(\mathbf{z}) \rangle = \lambda \langle T^*(\mathbf{x}), \mathbf{z} \rangle + \langle T^*(\mathbf{y}), \mathbf{z} \rangle \\ &= \langle \lambda T^*(\mathbf{x}) + T^*(\mathbf{y}), \mathbf{z} \rangle \\ \implies T^*(\lambda \mathbf{x} + \mathbf{y}) &= \lambda T^*(\mathbf{x}) + T^*(\mathbf{y}) \end{aligned}$$

2. These may be proved similarly to part 1 and are left as an exercise.

3. By Exercise 2.2.4, the  $jk^{\text{th}}$  entry of  $[T^*]_{\gamma}^{\beta}$  is

$$\langle T^*(\mathbf{w}_k), \mathbf{v}_j \rangle = \langle \mathbf{w}_k, T(\mathbf{v}_j) \rangle = \overline{\langle T(\mathbf{v}_j), \mathbf{w}_k \rangle} = \overline{A_{kj}}$$

We revisit our motivating set-up (+) in the language of part 3. Suppose:

- $V = \mathbb{F}^n$  and  $W = \mathbb{F}^m$  have standard orthonormal bases  $\beta = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$  and  $\gamma = \{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ .
- $T = L_A \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ .

Since the matrix of  $T$  with respect to the standard bases is simply  $A$  itself, the theorem confirms our earlier observation that the adjoint of  $L_A$  is left multiplication by the conjugate-transpose  $A^*$ :

$$[T^*]_{\gamma}^{\beta} = ([T]_{\beta}^{\gamma})^* = A^* = [L_A^*]_{\gamma}^{\beta} \implies T^* = (L_A)^* = L_{A^*}$$

Here is another straightforward example, this time using the standard inner products on  $\mathbb{C}^2$  and  $\mathbb{C}^3$ .

**Example 2.28.** Let  $T = L_A \in \mathcal{L}(\mathbb{C}^3, \mathbb{C}^2)$  where  $A = \begin{pmatrix} -i & 1 & -3 \\ 2 & 1-i & 4+2i \end{pmatrix}$ .

Plainly  $A = [T]_{\beta}^{\gamma}$  with respect to  $\beta = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$  and  $\gamma = \{\mathbf{e}_1, \mathbf{e}_2\}$ . We conclude that  $T^* = L_{A^*}$ :

$$[T^*]_{\gamma}^{\beta} = A^* = \begin{pmatrix} i & 2 \\ 1 & 1+i \\ -3 & 4-2i \end{pmatrix}$$

As a sanity check, multiply out a few examples of  $\langle A^* \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{w}, A \mathbf{v} \rangle$ ; make sure you're comfortable with the fact that the left inner product is on  $\mathbb{C}^2$  and the right on  $\mathbb{C}^3$ !

The Theorem tells us that every linear map  $T \in \mathcal{L}(V, W)$  between finite-dimensional spaces has an adjoint and moreover how to compute it:

1. Choose orthonormal bases (exist by Corollary 2.21) and find the matrix  $[T]_{\beta}^{\gamma}$ .
2. Take the conjugate-transpose  $([T]_{\beta}^{\gamma})^*$  and translate back to find  $T^* \in \mathcal{L}(W, V)$ .

The prospect of twice applying Gram–Schmidt and translating between linear maps and their matrices is unappealing; calculating this way can quickly become an enormous mess! In practice, it is often better to try a modified approach; see for instance part 2(b) of the next Example.

**Examples 2.29.** Let  $T = \frac{d}{dx} \in \mathcal{L}(P_1(\mathbb{R}))$  be the derivative operator;  $T(a + bx) = b$ . We treat  $P_1(\mathbb{R})$  as an inner product space in two ways.

1. Equip the inner product for which the standard basis  $\epsilon = \{1, x\}$  is orthonormal. Then

$$[T]_{\epsilon} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \implies [T^*]_{\epsilon} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \implies T^*(a + bx) = ax$$

2. Equip the  $L^2$  inner product  $\langle f, g \rangle = \int_0^1 f(x)g(x) dx$ . As we saw in Example 2.23, the basis  $\beta = \{f_1, f_2\} = \{1, 2x - 1\}$  is orthogonal with  $\|f_1\| = 1$  and  $\|f_2\| = \frac{1}{\sqrt{3}}$ . We compute the adjoint of  $T = \frac{d}{dx}$  in two different ways.

- (a) The basis  $\gamma = \{g_1, g_2\} = \left\{f_1, \sqrt{3}f_2\right\} = \left\{1, \sqrt{3}(2x - 1)\right\}$  is orthonormal. Observe that

$$\begin{aligned} T(g_1) = 0, \quad T(g_2) = 2\sqrt{3} &\implies [T]_{\gamma} = \begin{pmatrix} 0 & 2\sqrt{3} \\ 0 & 0 \end{pmatrix} \implies [T^*]_{\gamma} = \begin{pmatrix} 0 & 0 \\ 2\sqrt{3} & 0 \end{pmatrix} \\ \implies T^*(a + bx) = T^*\left(a + \frac{b}{2} + \frac{b}{2\sqrt{3}}\sqrt{3}(2x - 1)\right) &= \left(a + \frac{b}{2}\right) \cdot 2\sqrt{3}g_2 \\ &= 3(2a + b)(2x - 1) \end{aligned}$$

- (b) Use the orthogonal basis  $\beta$  and the projection formula (Theorem 2.18). With  $p(x) = a + bx$ ,

$$\begin{aligned} T^*(p) &= \frac{\langle T^*(p), f_1 \rangle}{\|f_1\|^2} f_1 + \frac{\langle T^*(p), f_2 \rangle}{\|f_2\|^2} f_2 = \langle p, T(1) \rangle + \langle p, T(2x - 1) \rangle \cdot 3(2x - 1) \\ &= \langle p, 0 \rangle + 3 \langle p, 2 \rangle (2x - 1) = 3 \left( \int_0^1 2(a + bx) dx \right) (2x - 1) \\ &= 3(2a + b)(2x - 1) \end{aligned}$$

Note the advantages here: no square roots and no need to change basis at the end!

The calculations for the second example were much nastier, even though we were already in possession of an orthogonal basis. The crucial point is that the two examples produce *different maps*  $T^*$ : the adjoint depends on the inner product!

### Why should we care about adjoints?

Adjoints might seem merely to be an abstraction of something simple (transposes) for its own sake. A convincing explanation of why adjoints are useful takes a lot of work; here is a short version.

Given a linear map  $T \in \mathcal{L}(V)$  on an inner product space, we now have *two desirable types of basis*.

1. Eigenbasis: diagonalizes  $T$ .
2. Orthonormal basis: recall (Theorem 2.18) how these simplify computations.

The capstone result of this course is the famous *spectral theorem* (Theorem 2.37) which says, in short, that *self-adjoint operators* ( $T^* = T$ ) have an *orthonormal eigenbasis*, the holy grail of easy computation! Such operators are important both theoretically and in applications such as quantum mechanics.

## The Fundamental Subspaces Theorem

To every linear map are associated its range and nullspace. These interact nicely with the adjoint...

**Theorem 2.30.** If  $T \in \mathcal{L}(V, W)$  has adjoint  $T^*$ , then,

1.  $\mathcal{R}(T^*)^\perp = \mathcal{N}(T)$
2. If  $V$  is finite dimensional, then  $\mathcal{R}(T^*) = \mathcal{N}(T)^\perp$

The corresponding results hold if we swap  $V \leftrightarrow W$  and  $T \leftrightarrow T^*$ .

The proof is left to Exercise 6. You've likely observed this with transposes of small matrices.

**Example 2.31.** Let  $A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 3 & -2 \end{pmatrix}$ . Viewed as a linear map between Euclidean spaces,  $T = L_A$  has adjoint  $T^* = L_{A^T}$ . It is easy to compute the relevant subspaces:

$$\mathcal{R}(A) = \mathbb{R}^2, \quad \mathcal{N}(A^T) = \{\mathbf{0}\}, \quad \mathcal{R}(A^T) = \text{Span} \left\{ \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix} \right\}, \quad \mathcal{N}(A) = \text{Span} \left( \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} \right)$$

## The Riesz Representation Theorem

This powerful result demonstrates a natural relation between an inner product space and its dual-space  $V^* = \mathcal{L}(V, \mathbb{F})$ .

**Theorem 2.32.** If  $V$  is finite-dimensional and  $g : V \rightarrow \mathbb{F}$  is linear, then there exists a unique  $\mathbf{y} \in V$  such that  $g(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle$ .

**Example 2.33.**  $g(p) := \int_0^1 p(x) dx$  is a linear map  $g : P_2(\mathbb{R}) \rightarrow \mathbb{R}$ . Equip  $P_2(\mathbb{R})$  with the inner product for which the standard basis  $\{1, x, x^2\}$  is orthonormal. Then

$$g(a + bx + cx^2) = a + \frac{1}{2}b + \frac{1}{3}c = \langle a + bx + cx^2, 1 + \frac{1}{2}x + \frac{1}{3}x^2 \rangle$$

We conclude that  $g(p) = \langle p, q \rangle$ , where  $q(x) = 1 + \frac{1}{2}x + \frac{1}{3}x^2$ .

The idea of the proof is very simple: if  $g(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle$  then the nullspace of  $g$  must equal  $\text{Span}\{\mathbf{y}\}^\perp \dots$

*Proof.* If  $g$  is the zero map, take  $\mathbf{y} = \mathbf{0}$ . Otherwise,  $\text{rank } g = 1$  and

$$V = \mathcal{N}(g) \oplus \mathcal{N}(g)^\perp \quad \text{where} \quad \dim \mathcal{N}(g)^\perp = 1 \quad (\text{rank-nullity theorem and Exercise 2.2.6})$$

Let  $\mathbf{u} \in \mathcal{N}(g)^\perp$  be either of the two unit vectors and define, independently of  $\mathbf{u}$ ,

$$\mathbf{y} := \overline{g(\mathbf{u})} \mathbf{u} \in V$$

Following the decomposition, write  $\mathbf{x} = \mathbf{n} + \alpha \mathbf{u}$  where  $\mathbf{n} \in \mathcal{N}(g)$  and observe that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{n} + \alpha \mathbf{u}, \overline{g(\mathbf{u})} \mathbf{u} \rangle = g(\mathbf{u})\alpha = 0 + g(\alpha \mathbf{u}) = g(\mathbf{n} + \alpha \mathbf{u}) = g(\mathbf{x})$$

The uniqueness of  $\mathbf{y}$  follows from the cancellation property (Lemma 2.5, part 4). ■



Due to the tight correspondence, the map is often decorated as  $g_y$ . Riesz's theorem indeed says that  $\mathbf{y} \mapsto g_y$  is an isomorphism  $V \cong V^*$ . While there are infinitely many isomorphisms between these spaces, the inner product structure identifies a *canonical* or preferred choice.

**Corollary 2.34.** *Every linear map on a finite-dimensional inner product space has an adjoint.*

Note how only the domain is required to be finite-dimensional! Riesz's Theorem and the Corollary also apply to continuous linear operators on (infinite-dimensional) Hilbert spaces, though the proof is a little trickier.

*Proof.* Let  $T \in \mathcal{L}(V, W)$  where  $\dim V < \infty$ , and suppose  $\mathbf{z} \in W$  is given. Simply *define*  $T^*(\mathbf{z}) := \mathbf{y}$  where  $\mathbf{y} \in V$  is the unique vector in Riesz's Theorem arising from the linear map

$$g : V \rightarrow \mathbb{F}, \quad g(\mathbf{x}) = \langle T(\mathbf{x}), \mathbf{z} \rangle$$

and check the required property:

$$\langle T^*(\mathbf{z}), \mathbf{x} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle = \overline{g(\mathbf{x})} = \langle \mathbf{z}, T(\mathbf{x}) \rangle \quad \blacksquare$$

**Exercises 2.3** 1. For each inner product space  $V$  and linear operator  $T \in \mathcal{L}(V)$ , evaluate  $T^*$  on the given vector.

(a)  $V = \mathbb{R}^2$  with the standard inner product,  $T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x+y \\ x-3y \end{pmatrix}$  and  $\mathbf{x} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$

(b)  $V = \mathbb{C}^2$  with the standard inner product,  $T \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} 2z+iw \\ (1-i)z \end{pmatrix}$  and  $\mathbf{x} = \begin{pmatrix} 3-i \\ 1+2i \end{pmatrix}$

(c)  $V = P_1(\mathbb{R})$  with  $\langle f, g \rangle = \int_0^1 f(t)g(t) dt$ ,  $T(f) = f' + 3f$  and  $f(t) = 4 - 2t$

2. Suppose  $A = \begin{pmatrix} 1 & 1 \\ 4 & 3 \end{pmatrix}$  and consider the linear map  $T = L_A \in \mathcal{L}(\mathbb{R}^2)$  where  $\mathbb{R}^2$  is equipped with the weighted inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = 4x_1y_1 + x_2y_2$$

(a) Find the matrix of  $T$  with respect to the orthonormal basis  $\beta = \{\mathbf{v}_1, \mathbf{v}_2\} = \{\frac{1}{2}\mathbf{e}_1, \mathbf{e}_2\}$ .

(b) Find the adjoint  $T^*$  and its matrix with respect to the standard basis  $\epsilon = \{\mathbf{e}_1, \mathbf{e}_2\}$ .

(Hint: the answer isn't  $A^T$ !)

3. Extending Examples 2.29, find the adjoint of  $T = \frac{d}{dx} \in \mathcal{L}(P_2(\mathbb{R}))$  with respect to:

(a) The inner product where the standard basis  $\epsilon = \{1, x, x^2\}$  is orthonormal.

(b) (Hard!) The  $L^2$  inner product  $\int_0^1 f(x)g(x) dx$ .

4. Let  $T(f) = f''$  be a linear transformation of  $P_2(\mathbb{R})$  and let  $\epsilon = \{1, x, x^2\}$  be the standard basis. Find  $T^*(a + bx + cx^2)$ :

(a) With respect to the inner product where  $\epsilon$  is orthonormal;

(b) With respect to the  $L^2$  inner product  $\langle f, g \rangle = \int_{-1}^1 f(t)g(t) dt$ .

(Hint:  $\{1, x, 3x^2 - 1\}$  is orthogonal)

5. Prove part 2 of Theorem 2.27.
6. Prove the Fundamental Subspaces Theorem 2.30.
7. For each inner product space  $V$  and linear transformation  $g : V \rightarrow \mathbb{F}$ , find a vector  $\mathbf{y} \in V$  such that  $g(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle$  for all  $\mathbf{x} \in V$ .
- $V = \mathbb{R}^3$  with the standard inner product, and  $g\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = x - 2y + 4z$
  - $V = \mathbb{C}^2$  with the standard inner product, and  $g\left(\begin{pmatrix} z \\ w \end{pmatrix}\right) = iz - 2w$
  - $V = P_2(\mathbb{R})$  with the  $L^2$  inner product  $\langle f, h \rangle = \int_0^1 f(x)h(x) dx$ , and  $g(f) = f'(1)$
8. (a) In the proof of Theorem 2.32, explain why  $\mathbf{y}$  depends only on  $g$  (not  $\mathbf{u}$ ).  
 (b) In the proof of Corollary 2.34, check that  $g(\mathbf{x}) := \langle T(\mathbf{x}), \mathbf{z} \rangle$  is linear.
9. Let  $\mathbf{y}, \mathbf{z} \in V$  be fixed vectors and define  $T \in \mathcal{L}(V)$  by  $T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle \mathbf{z}$ . Show that  $T^*$  exists and find an explicit expression.
10. Suppose  $A \in M_{m \times n}(\mathbb{F})$ . Prove that  $A^*A$  is diagonal if and only if the columns of  $A$  are orthogonal. What additionally would it mean if  $A^*A = I$ ?
11. Suppose  $T \in \mathcal{L}(V)$  where  $V$  is a finite-dimensional inner product space.
- Prove that the eigenvalues of  $T^*$  are the complex conjugates of those of  $T$ .  
 (Hint: relate the characteristic polynomial  $p^*(t) = \det(T^* - tI)$  to that of  $T$ )
  - Prove that  $T^*$  is diagonalizable if and only if  $T$  is.
12. (Hard) We present two linear maps which do not have an adjoint!
- Since  $\epsilon = \{1, x, x^2, \dots\}$  is a basis of  $P(\mathbb{R})$ , we may define a linear map  $T \in \mathcal{L}(P(\mathbb{R}))$  via  $T(x^n) = 1$  for all  $n$ ; for instance
 
$$T(4 + 3x + 2x^5) = 9$$
 Let  $\langle \cdot, \cdot \rangle$  be the inner product for which  $\epsilon$  is orthonormal. If  $T^*$  existed, show that
 
$$T^*(1) = \sum_{n=0}^{\infty} x^n$$
 would be an *infinite series*:  $T^*$  therefore does not exist.
  - For a related challenge, recall the space  $\ell^2$  of square-summable real sequences. For any sequence  $(x_n)_{n=1}^{\infty} \in \ell^2$ , define  $T \in \mathcal{L}(\ell^2)$  via
 
$$T(x_n) = \left( \sum_{n=1}^{\infty} \frac{1}{n} x_n, 0, 0, 0, 0, \dots \right)$$
 Find the adjoint  $T^*$ . If  $V \leq \ell^2$  is the subspace whose elements have only finitely many non-zero terms, show that the restriction  $T_V$  does not have an adjoint.

## 2.4 Normal & Self-Adjoint Operators and the Spectral Theorem

We now come to the fundamental question of this chapter: for which linear operators  $T \in \mathcal{L}(V)$  can we find an orthonormal eigenbasis? Many linear maps are, of course, not even diagonalizable, so in general this is far too much to hope for! Let's see what happens if such a basis exists...

If  $\beta$  is an orthonormal basis of eigenvectors of  $T$ , then

$$[T]_{\beta} = \text{diag}(\lambda_1, \dots, \lambda_n) \implies [T^*]_{\beta} = \text{diag}(\overline{\lambda_1}, \dots, \overline{\lambda_n})$$

If  $V$  is a *real* inner product space, then these matrices are identical and so  $T^* = T$ . In the complex case, we instead observe that

$$[TT^*]_{\beta} = \text{diag}(|\lambda_1|^2, \dots, |\lambda_n|^2) = [T^*T]_{\beta} \implies TT^* = T^*T$$

**Definition 2.35.** Suppose  $T$  is a linear operator on an inner product space  $V$  and assume  $T$  has an adjoint. We say that  $T$  is,

*Normal* if  $TT^* = T^*T$ ,

*Self-adjoint* if  $T^* = T$ .

The definitions for square matrices over  $\mathbb{R}$  and  $\mathbb{C}$  are identical, where  $*$  now denotes the conjugate-transpose.

A real self-adjoint matrix  $A \in M_n(\mathbb{R})$  is plainly *symmetric*:  $A^T = A$ . A complex self-adjoint matrix is also called *Hermitian*:  $A^* = A$ .

If  $T$  is self-adjoint then it is certainly normal, but the converse is false as the next example shows.

**Example 2.36.** The (non-symmetric) real matrix  $A = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$  is normal but not self-adjoint:

$$AA^T = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} = A^T A$$

More generally, every non-zero *skew-hermitian* matrix  $A^* = -A$  is normal but *not* self-adjoint:

$$A^* = -A \implies AA^* = -A^2 = A^*A$$

We saw above that linear maps with an orthonormal eigenbasis are either self-adjoint or normal depending whether the inner product space is real or complex. Amazingly, this provides a complete characterisation of such maps!

**Theorem 2.37 (Spectral Theorem, version 1).** Let  $T$  be a linear operator on a finite-dimensional inner product space  $V$ .

*Complex case:*  $V$  has an orthonormal basis of eigenvectors of  $T$  if and only if  $T$  is normal.

*Real case:*  $V$  has an orthonormal basis of eigenvectors of  $T$  if and only if  $T$  is self-adjoint.

The theorem gets its name from the *spectrum* (set of eigenvalues) of  $T$ .

**Examples 2.38.** 1. We diagonalize the self-adjoint linear map  $T = L_A \in \mathcal{L}(\mathbb{R}^2)$  where  $A = \begin{pmatrix} 6 & 3 \\ 3 & -2 \end{pmatrix}$ .

$$\text{Characteristic polynomial } p(t) = (6-t)(-2-t) - 9 = t^2 - 4t - 21 = (t-7)(t+3)$$

$$\text{Eigenvalues } \lambda_1 = 7, \quad \lambda_2 = -3$$

$$\text{Eigenvectors (normalized)} \quad \mathbf{w}_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

The basis  $\beta = \{\mathbf{w}_1, \mathbf{w}_2\}$  is orthonormal, with respect to which  $[T]_\beta = \begin{pmatrix} 7 & 0 \\ 0 & -3 \end{pmatrix}$  is diagonal.

2. The map  $T = L_A \in \mathcal{L}(\mathbb{R}^2)$  where  $A = \begin{pmatrix} 1 & 3 \\ 0 & -2 \end{pmatrix}$  is neither self-adjoint nor normal:

$$AA^* = \begin{pmatrix} 1 & 3 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 3 & -2 \end{pmatrix} = \begin{pmatrix} 10 & -6 \\ -6 & 4 \end{pmatrix} \neq \begin{pmatrix} 1 & 3 \\ 3 & 13 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 0 & -2 \end{pmatrix} = A^*A$$

It is diagonalizable, indeed

$$\gamma = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \end{pmatrix} \right\} \rightsquigarrow [T]_\gamma = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$$

In accordance with the spectral theorem,  $\gamma$  is not orthogonal.

3. Let  $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$  and consider  $T = L_A$  acting on both  $\mathbb{C}^2$  and  $\mathbb{R}^2$ . Since  $T$  is normal but not self-adjoint, we'll see how the field really matters in the spectral theorem.

First the complex case:  $T \in \mathcal{L}(\mathbb{C}^2)$  is normal and thus diagonalizable with respect to an orthonormal basis of eigenvectors. Here are the details.

$$\text{Characteristic polynomial } p(t) = t^2 + 1 = (t-i)(t+i)$$

$$\text{Eigenvalues } \lambda_1 = i, \quad \lambda_2 = -i$$

$$\text{Eigenvectors (normalized)} \quad \mathbf{w}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ 1 \end{pmatrix}$$

Certainly  $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$ ,  $\beta = \{\mathbf{w}_1, \mathbf{w}_2\}$  is orthonormal, and  $[T]_\beta = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$  is diagonal.

Now for the real case:  $T \in \mathcal{L}(\mathbb{R}^2)$  is not self-adjoint and thus should not be diagonalizable with respect to an orthonormal basis of eigenvectors. Indeed this is trivial; the characteristic polynomial has no roots in  $\mathbb{R}$  and so there are no real eigenvalues! It is also clear geometrically:  $T$  is rotation by  $90^\circ$  counter-clockwise around the origin, so it has no eigenvectors.

4. Let  $A = \begin{pmatrix} 3 & i \\ -i & 3 \end{pmatrix}$  and consider the self-adjoint operator  $T = L_A \in \mathcal{L}(\mathbb{C}^2)$ .

$$\text{Characteristic polynomial } p(t) = t^2 - 6t + 9 - 1 = (t-2)(t-4)$$

$$\text{Eigenvalues } \lambda_1 = 2, \quad \lambda_2 = 4$$

$$\text{Eigenvectors (normalized)} \quad \mathbf{w}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \mathbf{w}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

With respect to the orthonormal basis  $\beta = \{\mathbf{w}_1, \mathbf{w}_2\}$ , we have  $[T]_\beta = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$ .

## Proving the Spectral Theorem for Self-Adjoint Operators

**Lemma 2.39 (Basic properties of self-adjoint operators).** Let  $T \in \mathcal{L}(V)$  be self-adjoint.

1. If  $W \leq V$  is  $T$ -invariant then the restriction  $T_W$  is self-adjoint;
2. Every eigenvalue of  $T$  is real;
3. If  $\dim V$  is finite then  $T$  has an eigenvalue.

It is irrelevant whether  $V$  is real or complex. The previous example demonstrates part 2; even when  $V = \mathbb{C}^2$  is a complex inner product space, the eigenvalues of a self-adjoint matrix are real.

*Proof.* 1. Let  $\mathbf{w}_1, \mathbf{w}_2 \in W$ . Then

$$\langle T_W(\mathbf{w}_1), \mathbf{w}_2 \rangle = \langle T(\mathbf{w}_1), \mathbf{w}_2 \rangle = \langle \mathbf{w}_1, T(\mathbf{w}_2) \rangle = \langle \mathbf{w}_1, T_W(\mathbf{w}_2) \rangle$$

2. Suppose  $(\lambda, \mathbf{x})$  is an eigenpair. Then

$$\lambda \|\mathbf{x}\|^2 = \langle T(\mathbf{x}), \mathbf{x} \rangle = \langle \mathbf{x}, T(\mathbf{x}) \rangle = \bar{\lambda} \|\mathbf{x}\|^2 \implies \lambda \in \mathbb{R}$$

3. This is trivial if  $V$  is complex since every characteristic polynomial splits over  $\mathbb{C}$ . We therefore assume  $V$  is real. Choose any orthonormal basis  $\gamma$  of  $V$ , let  $A = [T]_\gamma \in M_n(\mathbb{R})$ , and define  $S := L_A \in \mathcal{L}(\mathbb{C}^n)$ . Then;

- The characteristic polynomial of  $S$  splits over  $\mathbb{C}$ , whence there exists an eigenvalue  $\lambda \in \mathbb{C}$ .
- The characteristic polynomials of  $S$  and  $T$  are *identical* (to that of  $A$ ).
- $S$  is self-adjoint and thus (part 2)  $\lambda \in \mathbb{R}$ .

It follows that  $T$  has the same *real* eigenvalue  $\lambda$ . ■

We are now able to prove the spectral theorem for self-adjoint operators on a finite-dimensional inner product space  $V$ . The argument applies regardless of whether  $V$  is real or complex.

*Proof of the Spectral Theorem (self-adjoint case).* We prove by induction on  $\dim V$ .

(Base case) If  $\dim V = 1$ , then  $V = \text{Span}\{\mathbf{x}\}$  and  $T(\mathbf{x}) = \lambda\mathbf{x}$  for some unit vector  $\mathbf{x}$  and scalar  $\lambda \in \mathbb{R}$ ; plainly  $\{\mathbf{x}\}$  is an orthonormal eigenbasis for  $T$ .

(Induction step) Fix  $n \in \mathbb{N}$  and assume that *every* self-adjoint operator on *every* inner product space of dimension  $n$  satisfies the spectral theorem. Let  $\dim V = n + 1$  and  $T \in \mathcal{L}(V)$  be self-adjoint.

By part 3 of the Lemma,  $T$  has an eigenpair  $(\lambda, \mathbf{x})$  where we may assume  $\mathbf{x}$  has unit length. Let  $W = \text{Span}\{\mathbf{x}\}^\perp$ . If  $\mathbf{w} \in W$ , then

$$\langle \mathbf{x}, T(\mathbf{w}) \rangle = \langle T(\mathbf{x}), \mathbf{w} \rangle = \lambda \langle \mathbf{x}, \mathbf{w} \rangle = 0 \tag{*}$$

whence  $W$  is  $T$ -invariant. Plainly  $\dim W = n$ . By part 1 of the Lemma,  $T_W$  is self-adjoint. By the induction hypothesis,  $T_W$  is diagonalized by some orthonormal basis  $\gamma$  of  $W$ . But then  $T$  is diagonalized by the orthonormal basis  $\beta = \gamma \cup \{\mathbf{x}\}$  of  $V$ . ■

## Proving the Spectral Theorem for Normal Operators

What changes for normal operators on complex inner product spaces? Not much! Indeed the proof is almost identical when  $T$  is merely normal.

- We don't need parts 2 and 3 of Lemma 2.39: every linear operator on a finite-dimensional complex inner product space has an eigenvalue and we no longer care whether eigenvalues are real.
- Two parts of the induction step need fixed:
  - $W$  being  $T$ -invariant: This isn't quite as simple as (\*), but thankfully part 3 of the next result provides the needed correction.
  - $T_W$  being normal: We need a replacement for part 1 of Lemma 2.39; this is a little more involved.

Rather than write out all the details, we leave this to Exercises 6 and 7.

For completeness, and as an analogue/extension of Lemma 2.39, we summarize some of the basic properties of normal operators. These also apply to self-adjoint operators as a special case.

**Lemma 2.40 (Basic properties of normal operators).** *Let  $T$  be normal on  $V$ . Then:*

1.  $\forall \mathbf{x} \in V, \|T(\mathbf{x})\| = \|T^*(\mathbf{x})\|$ .
2.  $T - tI$  is normal for any scalar  $t$ .
3.  $T(\mathbf{x}) = \lambda \mathbf{x} \iff T^*(\mathbf{x}) = \bar{\lambda} \mathbf{x}$  so that  $T$  and  $T^*$  have the same eigenvectors and conjugate eigenvalues. This recovers the previously established fact that  $\lambda \in \mathbb{R}$  if  $T$  is self-adjoint.
4. Distinct eigenvalues of  $T$  have orthogonal eigenvectors.

*Proof.* 1.  $\|T(\mathbf{x})\|^2 = \langle T(\mathbf{x}), T(\mathbf{x}) \rangle = \langle T^*T(\mathbf{x}), \mathbf{x} \rangle = \langle TT^*(\mathbf{x}), \mathbf{x} \rangle = \langle T^*(\mathbf{x}), T^*(\mathbf{x}) \rangle = \|T^*(\mathbf{x})\|^2$ .

2.  $\langle \mathbf{x}, (T - tI)(\mathbf{y}) \rangle = \langle \mathbf{x}, T(\mathbf{y}) \rangle - t \langle \mathbf{x}, \mathbf{y} \rangle = \langle T^*(\mathbf{x}), \mathbf{y} \rangle - \langle \bar{t} \mathbf{x}, \mathbf{y} \rangle = \langle (T^* - \bar{t}I)(\mathbf{x}), \mathbf{y} \rangle$  shows that  $T - tI$  has adjoint  $T^* - \bar{t}I$ . It is trivial to check that these commute.

3.  $T(\mathbf{x}) = \lambda \mathbf{x} \iff \|(T - \lambda I)(\mathbf{x})\| = 0 \xrightarrow[1\&2]{\text{parts}} \|(T^* - \bar{\lambda}I)(\mathbf{x})\| = 0 \iff T^*(\mathbf{x}) = \bar{\lambda} \mathbf{x}$ .

4. In part this follows from the spectral theorem, but we can also prove more straightforwardly. Suppose  $T(\mathbf{x}) = \lambda \mathbf{x}$  and  $T(\mathbf{y}) = \mu \mathbf{y}$  where  $\lambda \neq \mu$ . By part 3,

$$\begin{aligned} \lambda \langle \mathbf{x}, \mathbf{y} \rangle &= \langle \lambda \mathbf{x}, \mathbf{y} \rangle = \langle T(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, T^*(\mathbf{y}) \rangle = \langle \mathbf{x}, \bar{\mu} \mathbf{y} \rangle \\ &= \bar{\mu} \langle \mathbf{x}, \mathbf{y} \rangle \end{aligned}$$

This is a contradiction unless  $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ . ■

## Schur's Lemma

It is reasonable to ask how useful an orthonormal basis can be in general. Here is one answer.

**Lemma 2.41 (Schur).** *Suppose  $T$  is a linear operator on a finite-dimensional inner product space  $V$ . If the characteristic polynomial of  $T$  splits, then there exists an orthonormal basis  $\beta$  of  $V$  such that  $[T]_\beta$  is upper-triangular.*

The spectral theorem is a special case; since the proof is similar, we leave it to the exercises.

The conclusion of Schur's lemma is weaker than the spectral theorem, though it applies to more operators: indeed if  $V$  is complex, it applies to *any*  $T$ ! Every example of the spectral theorem is also an example of Schur's lemma. Example 2.38.2 provides another, since the matrix  $A$  is already upper triangular with respect to the standard orthonormal basis. Here is another example.

**Example 2.42.** Consider  $T(f) = 2f'(x) + xf(1)$  as a linear map  $T \in \mathcal{L}(P_1(\mathbb{R}))$  with respect to the  $L^2$  inner product  $\langle f, g \rangle = \int_0^1 f(t)g(t) dt$ . We have

$$T(a + bx) = 2b + (a + b)x$$

If  $[T]_\beta$  is to be upper-triangular, the first vector in  $\beta$  must be an eigenvector of  $T$ . It is easily checked that  $f_1 = 1 + x$  is such with eigenvalue 2. To find a basis satisfying Schur's lemma, we need only find  $f_2$  orthogonal to this and then normalize. This can be done by brute force since the problem is small, but for the sake of practice we apply Gram-Schmidt to the polynomial 1:

$$1 - \frac{\langle 1, 1+x \rangle}{\|1+x\|^2}(1+x) = 1 - \frac{1 + \frac{1}{2}}{1 + 1 + \frac{1}{3}}(1+x) = \frac{1}{14}(5 - 9x) \implies f_2 = 5 - 9x$$

Indeed we obtain an upper-triangular matrix for  $T$ :

$$T(f_2) = -18 - 4x = -13(1+x) - (5-9x) = -13f_1 - f_2 \implies [T]_{\{f_1, f_2\}} = \begin{pmatrix} 2 & -13 \\ 0 & -1 \end{pmatrix}$$

We can also work with the corresponding orthonormal basis as posited in the theorem, though the matrix is messier:

$$\beta = \{g_1, g_2\} = \left\{ \sqrt{\frac{3}{7}}(1+x), \frac{1}{\sqrt{7}}(5-9x) \right\} \implies [T]_\beta = \begin{pmatrix} 2 & -\frac{13}{\sqrt{3}} \\ 0 & -1 \end{pmatrix}$$

Alternatively, we could have started with the other eigenvector  $h_1 = 2 - x$ : an orthogonal vector to this is  $h_2 = 4 - 9x$ , with respect to which

$$[T]_{\{h_1, h_2\}} = \begin{pmatrix} -1 & -13 \\ 0 & 2 \end{pmatrix}$$

In both cases the eigenvalues are down the diagonal, as must be for an upper-triangular matrix.

In general, it is difficult to quickly find a suitable basis satisfying Schur's lemma. After trying the proof in the exercises, you should be able to describe a method, though it is impractically slow!

**Exercises 2.4** 1. For each linear operator  $T$  on an inner product space  $V$ , decide whether  $T$  is normal, self-adjoint, or neither. If the spectral theorem permits, find an orthonormal eigenbasis.

(a)  $V = \mathbb{R}^2$  and  $T \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2a-b \\ -2a+5b \end{pmatrix}$

(b)  $V = \mathbb{R}^3$  and  $T \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -a-b \\ 5b \\ 4a-2b+5c \end{pmatrix}$

(c)  $V = \mathbb{C}^2$  and  $T \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} 2z+iw \\ z+2iw \end{pmatrix}$

(d)  $V = \mathbb{R}^4$  with  $T : (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4) \mapsto (\mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_1, \mathbf{e}_2)$

(e)  $V = P_2(\mathbb{R})$  with  $\langle f, g \rangle = \int_0^1 f(t)g(t) dt$  and  $T(f) = f'$

(Hint: Don't compute  $T^*$ ! Instead assume  $T$  is normal and aim for a contradiction...)

2. Let  $T(f(x)) = f'(x) + 4xf(0)$  where  $T \in \mathcal{L}(P_1(\mathbb{R}))$  and  $\langle f, g \rangle = \int_{-1}^1 f(t)g(t) dt$ . Find an orthonormal basis of  $P_1(\mathbb{R})$  with respect to which the matrix of  $T$  is upper-triangular.

3. Suppose  $S, T$  are self-adjoint operators on an inner product space  $V$ . Prove that  $ST$  is self-adjoint if and only if  $ST = TS$ .

(Hint: recall Theorem 2.27)

4. Let  $T$  be normal on a finite-dimensional inner product space  $V$ . Prove that  $\mathcal{N}(T^*) = \mathcal{N}(T)$  and that  $\mathcal{R}(T^*) = \mathcal{R}(T)$ .

(Hint: Use Lemma 2.40 and the Fundamental Subspaces Theorem 2.30)

5. Let  $T$  be self-adjoint on a finite-dimensional inner product space  $V$ . Prove that

$$\forall \mathbf{x} \in V, \quad \|\mathbf{T}(\mathbf{x}) \pm i\mathbf{x}\|^2 = \|\mathbf{T}(\mathbf{x})\|^2 + \|\mathbf{x}\|^2$$

Hence prove that  $T - iI$  is invertible and that  $[(T - iI)^{-1}]^* = (T + iI)^{-1}$ .

6. Let  $W$  be a  $T$ -invariant subspace of an inner product space  $V$  and let  $T_W \in \mathcal{L}(W)$  be the restriction of  $T$  to  $W$ . Prove:

(a)  $W^\perp$  is  $T^*$ -invariant.

(b) If  $W$  is both  $T$ - and  $T^*$ -invariant, then  $(T_W)^* = (T^*)_W$ .

(c) If  $W$  is both  $T$ - and  $T^*$ -invariant and  $T$  is normal, then  $T_W$  is normal.

7. Use the previous question to complete the proof of the spectral theorem for a normal operator on a finite-dimensional complex inner product space.

8. (a) Suppose  $S$  is a normal operator on a finite-dimensional *complex* inner product space, all of whose eigenvalues are *real*. Prove that  $S$  is self-adjoint.

(b) Let  $T$  be a normal operator on a finite-dimensional *real* inner product space  $V$  whose characteristic polynomial splits. Prove that  $T$  is self-adjoint and that there exists an orthonormal basis of  $V$  of eigenvectors of  $T$ .

(Hint: Mimic the proof of Lemma 2.39 part 3 and use part (a))

9. Prove Schur's lemma by induction, similarly to the proof of the spectral theorem.

(Hint:  $T^*$  has an eigenvector  $\mathbf{x}$ ; why? Now show that  $W = \text{Span}\{\mathbf{x}\}^\perp$  is  $T$ -invariant...)



## 2.5 Unitary and Orthogonal Operators and their Matrices

In this section we focus on length-preserving transformations of an inner product space.

**Definition 2.43.** A linear<sup>6</sup> isometry of an inner product space  $V$  is a linear map  $T$  satisfying

$$\forall \mathbf{x} \in V, \|T(\mathbf{x})\| = \|\mathbf{x}\|$$

Every eigenvalue of an isometry must have modulus 1: if  $T(\mathbf{w}) = \lambda\mathbf{w}$ , then

$$\|\mathbf{w}\|^2 = \|T(\mathbf{w})\|^2 = \|\lambda\mathbf{w}\|^2 = |\lambda|^2 \|\mathbf{w}\|^2$$

**Example 2.44.** Let  $T = L_A \in \mathcal{L}(\mathbb{R}^2)$ , where  $A = \frac{1}{5} \begin{pmatrix} 4 & -3 \\ 3 & 4 \end{pmatrix}$ . Then

$$\left\| T \begin{pmatrix} x \\ y \end{pmatrix} \right\|^2 = \left\| \frac{1}{5} \begin{pmatrix} 4x - 3y \\ 3x + 4y \end{pmatrix} \right\|^2 = \frac{1}{25} ((4x - 3y)^2 + (3x + 4y)^2) = x^2 + y^2 = \left\| \begin{pmatrix} x \\ y \end{pmatrix} \right\|^2$$

This matrix is very special in that its inverse equals its transpose:

$$A^{-1} = \frac{1}{\frac{16}{25} + \frac{9}{25}} \begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 4 & 3 \\ -3 & 4 \end{pmatrix} = A^T$$

We call such matrices *orthogonal*. The simple version of what follows is that every linear isometry on  $\mathbb{R}^n$  is multiplication by an orthogonal matrix.

**Definition 2.45.** A unitary operator  $T$  on an inner product space  $V$  is an invertible linear map satisfying  $T^*T = I = TT^*$ . A unitary matrix is a (real or complex) matrix satisfying  $A^*A = I$ .

If  $V$  is real, we usually call these *orthogonal operators/matrices*; this isn't necessary, since *unitary* encompasses both real and complex spaces. An orthogonal matrix satisfies  $A^T A = I$ .

**Example 2.46.** The matrix  $A = \frac{1}{3} \begin{pmatrix} i & 2+2i \\ 2-2i & i \end{pmatrix}$  is unitary:

$$A^*A = \frac{1}{9} \begin{pmatrix} -i & 2+2i \\ 2-2i & -i \end{pmatrix} \begin{pmatrix} i & 2+2i \\ 2-2i & i \end{pmatrix} = \frac{1}{9} \begin{pmatrix} -i^2 + 4 + 4 & (-i+i)(2+2i) \\ (i-i)(2-2i) & 4 + 4 - i^2 \end{pmatrix} = I$$

If  $V$  is finite-dimensional, the operator/matrix notions correspond straightforwardly. By Theorem 2.27, if we choose any orthonormal basis  $\beta$  of  $V$ , then

$$T \in \mathcal{L}(V) \text{ is unitary/orthogonal} \iff [T]_\beta \text{ is unitary/orthogonal}$$

Moreover, we need only assume  $T^*T = I$  (or  $TT^* = I$ ) if  $V$  is finite-dimensional: if  $\beta$  is an orthonormal basis, then

$$T^*T = I \iff [T^*]_\beta [T]_\beta = I \iff [T]_\beta [T^*]_\beta = I \iff TT^* = I$$

In infinite dimensions, we need  $T^*$  to be both the left- and right-inverse of  $T$ . This isn't an empty requirement (see Exercise 13).

<sup>6</sup>There also exist *non-linear* isometries: for instance *translations* ( $T(\mathbf{x}) = \mathbf{x} + \mathbf{a}$  for any constant  $\mathbf{a}$ ) and *complex conjugation* ( $T(\mathbf{x}) = \bar{\mathbf{x}}$ ) on  $\mathbb{C}^n$ . Together with linear isometries, these essentially comprise all isometries in finite dimensions.

We now tackle the correspondence between unitary operators and isometries.

**Theorem 2.47.** *Let  $T$  be a linear operator on an inner product space  $V$ .*

1. *If  $T$  is a unitary/orthogonal operator, then it is a linear isometry.*
2. *If  $T$  is a linear isometry and  $V$  is finite-dimensional, then  $T$  is unitary/orthogonal.*

*Proof.* 1. If  $T$  is unitary, then

$$\forall \mathbf{x}, \mathbf{y} \in V, \quad \langle \mathbf{x}, \mathbf{y} \rangle = \langle T^*T(\mathbf{x}), \mathbf{y} \rangle = \langle T(\mathbf{x}), T(\mathbf{y}) \rangle \quad (\dagger)$$

In particular taking  $\mathbf{x} = \mathbf{y}$  shows that  $T$  is an isometry.

2.  $(I - T^*T)^* = I^* - (T^*T)^* = I - T^*T$  is self-adjoint. By the spectral theorem, there exists an orthonormal basis of  $V$  of eigenvectors of  $I - T^*T$ . For any such  $\mathbf{x}$  with (real) eigenvalue  $\lambda$ ,

$$\begin{aligned} 0 &= \|\mathbf{x}\|^2 - \|T(\mathbf{x})\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle - \langle T(\mathbf{x}), T(\mathbf{x}) \rangle = \langle \mathbf{x}, (I - T^*T)\mathbf{x} \rangle = \lambda \|\mathbf{x}\|^2 \\ &\implies \lambda = 0 \end{aligned}$$

Since  $I - T^*T = 0$  on a basis,  $T^*T = I$ . Since  $V$  is finite-dimensional, we also have  $TT^* = I$  whence  $T$  is unitary. ■

The finite-dimensional restriction is important in part 2: we use the existence of adjoints, the spectral theorem, and that a left-inverse is also a right-inverse. See Exercise 13 for an example of a *non-unitary* isometry in infinite dimensions.

The proof shows a little more:

**Corollary 2.48.** *On a finite-dimensional space, being unitary is equivalent to each of the following:*

- (a) *Preservation of the inner product<sup>7</sup> ( $\dagger$ ).*
- (b) *The existence of an orthonormal basis  $\beta = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$  such that  $T(\beta) = \{T(\mathbf{w}_1), \dots, T(\mathbf{w}_n)\}$  is also orthonormal.*
- (c) *That every orthonormal basis  $\beta$  of  $V$  is mapped to an orthonormal basis  $T(\beta)$ .*

While (a) is simply ( $\dagger$ ), claims (b) and (c) are also worth proving explicitly: see Exercise 9. If  $\beta$  is the standard orthonormal basis of  $\mathbb{F}^n$  and  $T = L_A$ , then the columns of  $A$  form the orthonormal set  $T(\beta)$ . This makes identifying unitary/orthogonal matrices easy:

**Corollary 2.49.** *A matrix  $A \in M_n(\mathbb{R})$  is orthogonal if and only if its columns form an orthonormal basis of  $\mathbb{R}^n$  with respect to the standard (dot) inner product.*

*A matrix  $A \in M_n(\mathbb{C})$  is unitary if and only if its columns form an orthonormal basis of  $\mathbb{C}^n$  with respect to the standard (Hermitian) inner product.*

<sup>7</sup>In particular, in a real inner product space isometries also preserve the *angle*  $\theta$  between vectors since  $\cos \theta = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}$ .

**Examples 2.50.** 1. The matrix  $A_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \in M_2(\mathbb{R})$  is orthogonal for any  $\theta$ . Example 2.44 is this with  $\theta = \tan^{-1} \frac{3}{4} = \sin^{-1} \frac{3}{5} = \cos^{-1} \frac{4}{5}$ . More generally (Exercise 6), it can be seen that every real orthogonal  $2 \times 2$  matrix has the form  $A_\theta$  or

$$B_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}$$

for some angle  $\theta$ . The effect of the  $L_{A_\theta}$  is to rotate counter-clockwise by  $\theta$ , while that of  $L_{B_\theta}$  is to reflect across the line making angle  $\frac{1}{2}\theta$  with the positive  $x$ -axis.

2.  $A = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \\ -\sqrt{2} & \sqrt{3} & -1 \end{pmatrix} \in M_3(\mathbb{R})$  is orthogonal: check the columns!

3.  $A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \in M_2(\mathbb{C})$  is unitary: indeed it maps the standard basis to the orthonormal basis

$$T(\beta) = \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix} \right\}$$

It is also easy to check that the characteristic polynomial is

$$p(t) = \det \begin{pmatrix} \frac{1}{\sqrt{2}} - t & \frac{i}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} - t \end{pmatrix} = \left(t - \frac{1}{\sqrt{2}}\right)^2 + \frac{1}{2} \implies t = \frac{1}{\sqrt{2}}(1 \pm i) = e^{\pm\pi i/4}$$

whence the eigenvalues of  $T$  both have modulus 1.

4. Here is an example of an infinite-dimensional unitary operator. On the space  $C[-\pi, \pi]$ , the function  $T(f(x)) = e^{ix}f(x)$  is linear. Moreover

$$\langle e^{ix}f(x), g(x) \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ix}f(x)\overline{g(x)} dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-ix}\overline{g(x)} dx = \langle f(x), e^{-ix}g(x) \rangle$$

whence  $T^*(f(x)) = e^{-ix}f(x)$ . Indeed  $T^* = T^{-1}$  and so  $T$  is a unitary operator.

Since  $C[-\pi, \pi]$  is infinite-dimensional, we don't expect all parts of the Corollary to hold:

(a)  $T$  does preserve the inner product.

(b), (c)  $C[-\pi, \pi]$  doesn't have an orthonormal basis; there is no orthonormal set  $\beta = \{f_k\}$  so that every continuous function is a *finite* linear combination.<sup>8</sup>We cannot therefore claim that  $T$  maps orthonormal bases to orthonormal bases!

<sup>8</sup>An *infinite* orthonormal set  $\beta = \{f_k : k \in \mathbb{Z}\}$  can be found so that every function  $f$  'equals' an *infinite series* in the sense that  $\|f - \sum a_k f_k\| = 0$ . Since these are not finite sums,  $\beta$  isn't strictly a basis, though it isn't uncommon for it to be so described. Moreover, given that the norm is defined by an integral, this also isn't a claim that  $f$  and  $\sum a_k f_k$  are equal as functions. Indeed the infinite series need not be continuous! For these reasons, when working with Fourier series, one tends to consider a broader class than the continuous functions.

## Unitary and Orthogonal Equivalence

Suppose  $A \in M_n(\mathbb{R})$  is symmetric (self-adjoint)  $A^T = A$ . By the spectral theorem,  $A$  has an orthonormal eigenbasis  $\beta = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ :  $A\mathbf{w}_j = \lambda_j\mathbf{w}_j$ . Arranging the eigenbasis as the columns of a matrix, we see that the columns of  $U = (\mathbf{w}_1 \cdots \mathbf{w}_n)$  are orthonormal and so  $U$  is an orthogonal matrix. We can therefore write

$$A = UDU^{-1} = U \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} U^T$$

A similar approach works if  $A \in M_n(\mathbb{C})$  is normal: we now have  $A = UDU^*$  where  $U$  is unitary.

**Example 2.51.** The matrix  $A = \begin{pmatrix} 1+i & 1+i \\ -1-i & 1+i \end{pmatrix}$  is normal as can easily be checked. Its characteristic polynomial is

$$p(t) = t^2 - 2(1+i)t + 4i = (t-2i)(t-2)$$

with corresponding orthonormal eigenvectors

$$\mathbf{w}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad \mathbf{w}_{2i} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$$

We conclude that

$$A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2i \end{pmatrix} \left[ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \right]^{-1} = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}$$

This is an example of unitary equivalence.

**Definition 2.52.** Square matrices  $A, B$  are *unitarily equivalent* if there exists a unitary matrix  $U$  such that  $B = U^*AU$ . *Orthogonal equivalence* is similar:  $B = U^T AU$ .

The above discussion proves half the following:

**Theorem 2.53.**  $A \in M_n(\mathbb{C})$  is normal if and only if it is unitarily equivalent to a diagonal matrix (the matrix of its eigenvalues).

$A \in M_n(\mathbb{R})$  is self-adjoint (symmetric) if and only if it is orthogonally equivalent to a diagonal matrix.

*Proof.* We've already observed the  $(\Rightarrow)$  direction.

For the converse, let  $D$  be diagonal,  $U$  unitary, and  $A = U^*DU$ . Then

$$\begin{aligned} A^*A &= (U^*DU)^*U^*DU = U^*D^*UU^*DU = U^*\bar{D}DU = U^*D\bar{D}U = U^*DUU^*\bar{D}U \\ &= U^*DU(U^*DU)^* = AA^* \end{aligned}$$

since  $U^* = U^{-1}$  and because diagonal matrices commute:  $\bar{D}D = D\bar{D}$ .

In the special case where  $A$  is real and  $U$  is orthogonal, then  $A$  is symmetric:

$$A^T = (U^T DU)^T = U^T D^T U = U^T D U = A$$

■

**Exercises 2.5** 1. For each matrix  $A$  find an orthogonal or unitary  $U$  and a diagonal  $D = U^*AU$ .

(a)  $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$     (b)  $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$     (c)  $\begin{pmatrix} 2 & 3-3i \\ 3+3i & 5 \end{pmatrix}$     (d)  $\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$

2. Which of the following pairs are unitarily/orthogonally equivalent? Explain your answers.

(a)  $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  and  $B = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$     (b)  $A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  and  $B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(c)  $A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  and  $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$

3. Let  $a, b \in \mathbb{C}$  be such that  $|a|^2 + |b|^2 = 1$ . Prove that every  $2 \times 2$  matrix of the form  $\begin{pmatrix} a & -e^{i\theta}\bar{b} \\ b & e^{i\theta}a \end{pmatrix}$  is unitary. Are these all the unitary  $2 \times 2$  matrices? Prove or disprove.

4. If  $A, B$  are orthogonal/unitary, prove that  $AB$  and  $A^{-1}$  are also orthogonal/unitary.

(This proves that orthogonal/unitary matrices are groups under matrix multiplication)

5. Check that  $A = \frac{1}{3} \begin{pmatrix} 5 & -4i \\ 4i & 5 \end{pmatrix} \in M_2(\mathbb{C})$  satisfies  $A^T A = I$  (it is a complex orthogonal matrix).

(These don't have the same nice relationship with inner products, and are thus less useful to us)

6. Supply the details of Exercise 2.50.1.

(Hints:  $\beta = \{\mathbf{i}, \mathbf{j}\}$  is orthonormal, whence  $\{A\mathbf{i}, A\mathbf{j}\}$  must be orthonormal. Now draw pictures to compute the result of rotating and reflecting the vectors  $\mathbf{i}$  and  $\mathbf{j}$ .)

7. Show that the linear map in Example 2.50.4 has no eigenvectors.

8. Prove that  $A \in M_n(\mathbb{C})$  has an orthonormal basis of eigenvectors whose eigenvalues have modulus 1, if and only if  $A$  is unitary.

9. Prove parts (b) and (c) of Corollary 2.48 for a finite-dimensional inner product space:

(a) If  $\beta$  is an orthonormal basis such that  $T(\beta)$  is orthonormal, then  $T$  is unitary.

(b) If  $T$  is unitary, and  $\eta$  is an orthonormal basis, then  $T(\eta)$  is an orthonormal basis.

10. Let  $T$  be a linear operator on a finite-dimensional inner product space  $V$ . If  $\|T(\mathbf{x})\| = \|\mathbf{x}\|$  for all  $\mathbf{x}$  in some orthonormal basis of  $V$ , must  $T$  be unitary? Prove or disprove.

11. Let  $T$  be a unitary operator on an inner product space  $V$  and let  $W$  be a finite-dimensional  $T$ -invariant subspace of  $V$ . Prove:

(a)  $T(W) = W$  (Hint: show that  $T_W$  is injective);

(b)  $W^\perp$  is  $T$ -invariant.

12. Let  $W$  a subspace of an inner product space  $V$  such that  $V = W \oplus W^\perp$ . Define  $T \in \mathcal{L}(V)$  by  $T(\mathbf{u} + \mathbf{w}) = \mathbf{u} - \mathbf{w}$  where  $\mathbf{u} \in W$  and  $\mathbf{w} \in W^\perp$ . Prove that  $T$  is unitary and self-adjoint.

13. In the inner product space  $\ell^2$  of square-summable sequences, consider the linear operator  $T(x_1, x_2, \dots) = (0, x_1, x_2, \dots)$ . Prove that  $T$  is an isometry and compute its adjoint. Check that  $T$  is non-invertible and non-unitary.

14. Prove Schur's Lemma for matrices. Every  $A \in M_n(\mathbb{R})$  is orthogonally equivalent and every  $A \in M_n(\mathbb{C})$  is unitarily equivalent to an upper triangular matrix.