Chapter VI. Inner Product Spaces.

VI.1. Basic Definitions and Examples.

In Calculus you encountered Euclidean coordinate spaces $\mathbb{R}^n$ equipped with additional structure: an **inner product** $B : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$.

Euclidean Inner Product: $B(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} x_i y_i$

which is often abbreviated to $B(x, y) = (x, y)$. Associated with it we have the **Euclidean norm**

$$\|\mathbf{x}\| = \sum_{i=1}^{n} |x_i|^2 = (\mathbf{x}, \mathbf{x})^{1/2}$$

which represents the "length" of a vector, and a **distance function**

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

which gives the Euclidean distance from $\mathbf{x}$ to $\mathbf{y}$. Note that $\mathbf{y} = \mathbf{x} + (\mathbf{y} - \mathbf{x})$.

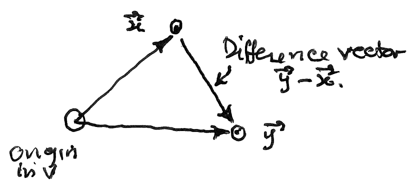

Figure 6.1. The distance between points $\mathbf{x}$, $\mathbf{y}$ in an inner product space is interpreted as the norm (length) $\|\mathbf{y} - \mathbf{x}\|$ of the difference vector $\Delta\mathbf{x} = \mathbf{y} - \mathbf{x}$.

This inner product on $\mathbb{R}^n$ has the following geometric interpretation

$$(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\| \cdot \|\mathbf{x}\| \cdot \cos\big(\theta(\mathbf{x}, \mathbf{y})\big)$$

where θ is the angle between x and y, measured in the plane $M = \mathbb{R}\text{-span}\{\mathbf{x}, \mathbf{y}\}$, the 2-dimensional subspace in $\mathbb{R}^n$ spanned by $\mathbf{x}$ and $\mathbf{y}$. **Orthogonality** of two vectors is then interpreted to mean $(\mathbf{x}, \mathbf{y}) = 0$; the zero vector is orthogonal to everybody, by definition. These notions of *length, distance*, and *orthogonality* do not exist in unadorned vector spaces.

We now generalize the notion of inner product to arbitrary vector spaces, even if they are infinite-dimensional.

1.1. Definition. *If V is a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, an* **inner product** *is a map $B : V \times V \rightarrow \mathbb{K}$ taking ordered pairs of vectors to scalars $B(v_1, v_2) \in \mathbb{K}$ with the following properties*

1. Separate Additivity in each Entry. *B is additive in each input if the other input is held fixed:*
   - $B(v_1 + v_2, w) = B(v_1, w) + B(v_2, w)$
   - $B(v, w_1 + w_2) = B(v, w_1) + B(v, w_2)$.

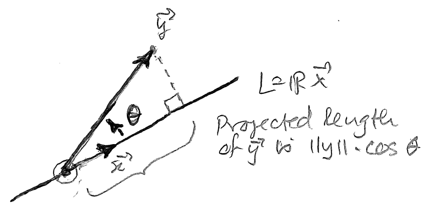

Figure 6.2. Geometric interpretation of the inner product $(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\|\,\|\mathbf{y}\|\cdot\cos(\theta(\mathbf{x}, \mathbf{y}))$ in $\mathbb{R}^n$. The projected length of a vector $\mathbf{y}$ onto the line $L = \mathbb{R}\mathbf{x}$ is $\|\mathbf{y}\|\cdot\cos(\theta)$. The angle $\theta(\mathbf{x}, \mathbf{y})$ is measured within the two-dimensional subspace $M = \mathbb{R}\text{-span}\{\mathbf{x}, \mathbf{y}\}$. Vectors are *orthogonal* when $\cos\theta = 0$, so $(\mathbf{x}, \mathbf{y}) = 0$. The zero vector is orthogonal to everybody.

for v, v_i, w, w_i in V.

$|rm2$. Positive Definite. *For all $v \in V$,*

$$B(v, v) \geq 0 \quad \text{and} \quad B(v, v) = 0 \text{ if and only if } v = 0$$

3. Hermitian Symetric. *For all $v, w \in V$,*

$$B(v, w) = \overline{B(w, v)} \qquad \text{when inputs are interchanged.}$$

Conjugation does nothing for $x \in \mathbb{R}$ ($\overline{x} = x$ for $x \in \mathbb{R}$), so an inner product on a real vector space is simply symmetric, with $B(w, v) = B(v, w)$.

1. Hermitian. *For $\lambda \in \mathbb{K}$, $v, w \in V$,*

4. $B(\lambda v, w) = \lambda B(v, w)$ *and,*
- $B(v, \lambda w) = \bar{\lambda} B(v, w)$.

An inner product on a real vector space is just a **bilinear map** *– one that is $\mathbb{R}$-linear in each input when the other is held fixed – because conjugation does nothing in $\mathbb{R}$.*

The Euclidean inner product in $\mathbb{R}^n$ is a special case of the standard Euclidean inner product in complex coordinate space $V = \mathbb{C}^n$,

$$(\mathbf{z}, \mathbf{w}) = \sum_{j=1}^{n} z_j \overline{w_j} \quad ,$$

which is easily seen to have properties (1.)–(4.) The corresponding Euclidean norm and distance functions on $\mathbb{C}^n$ are then

$$\|\mathbf{z}\| = (\mathbf{z}, \mathbf{z})^{1/2} = \Big[\sum_{j=1}^{n} |z_j|^2\Big]^{1/2} \qquad \text{and} \qquad d(\mathbf{z}, \mathbf{w}) = \|\mathbf{z} - \mathbf{w}\| = \Big[\sum_{j=1}^{n} |z_j - w_j|^2\Big]^{1/2}$$

Again, properties (1.) - (4.) are easily verified.

For an arbitrary inner product B we define the corresponding norm and distance functions

$$\|v\|_B = B(v, v)^{1/2} \qquad d_B(v_1, v_2) = \|v_1 - v_2\|_B$$

which are no longer given by such formulas.

1.2. Example. Here are two important examples of inner product spaces.

1. On $V = \mathbb{C}^n$ (For $\mathbb{R}^n$) we can define "nonstandard" inner products by assigning different positive weights $\alpha_j > 0$ to each coordinate direction, taking

$$B_\alpha(\mathbf{z},\mathbf{w}) = \sum_{j=1}^{n} \alpha_j \cdot z_j \overline{w_j} \qquad \text{with norm} \qquad \|\mathbf{z}\|_\alpha = \Big[\sum_{j=1}^{n} \alpha_j \cdot |z_j|^2\Big]^{1/2}$$

This is easily seen to be an inner product. Thus the standard Euclidean inner product on $\mathbb{R}^n$ or $\mathbb{C}^n$, for which $\alpha_1 = \ldots = \alpha_n = 1$, is part of a much larger family.

2. The space $\mathcal{C}[a,b]$ of continuous complex-valued functions $f : [a,b] \to \mathbb{C}$ becomes an inner product space if we define

$$(f,h)_2 = \int_a^b f(t)\overline{h(t)}\,dt \qquad \text{(Riemann integral)}$$

The corresponding "**$\mathbf{L}^2$-norm**" of a function is then

$$\|f\|_2 = \Big[\int_a^b |f(t)|^2\,dt\Big]^{1/2} ;$$

the inner product axioms follow from simple properties of the Riemann integral. This infinite-dimensional inner product space arises in many applications, particularly Fourier analysis. $\square$

1.3. Exercise. Verify that both inner products in the last example actually satisfy the inner product axioms. In particular, explain why the L^2-inner product $(f,h)_2$ has $\|f\|_2 > 0$ when f is not the zero function ($f(t) \equiv 0$ for all t).

We now take up the basic properties common to all inner product spaces.

1.4. Theorem. *On any inner product space V the associated norm has the following properties*

(a) $\|x\| \geq 0$*;*

(b) $\|\lambda x\| = |\lambda| \cdot \|x\|$ (*and in particular,* $\|-x\| = \|x\|$)*;*

(c) (Triangle Inequality) *For* $x, y \in V$*,* $\|x \pm y\| \leq \|x\| + \|y\|$.

Proof: The first two are obvious. The third is important because it implies that the distance function $d_B(x,y) = \|x-y\|$ satisfies the "*geometric triangle inequality*"

$$d_B(x,y) \leq d_B(x,z) + d_B(z,y), \qquad \text{for all } x,y,z \in V$$

as indicated in Figure 6.3. This follows directlly from (3.) because

$$d_B(x,y) = \|x-y\| = \|(x-z)+(z-y)\| \leq \|x-z\| + \|z-y\| = d_B(x,z) + d_B(z,y)$$

The version of (3.) involving a $(-)$ sign follows from that featuring a $(+)$ because $v - w = v + (-w)$ and $\|-w\| = \|w\|$.

The proof of (3.) is based on an equally important inequality:

1.5. Lemma (Schwartz Inequality). *If B is an inner product on a real or complex vector space then*

$$|B(x,y)| \leq \|x\|_B \cdot \|y\|_B$$

for all $x, y \in V$.

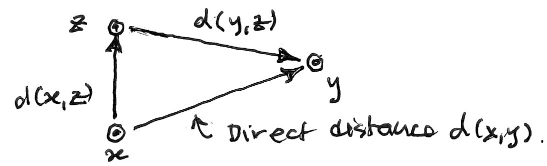

Figure 6.3. The meaning of the Triangle Inequality: direct distance from $\mathbf{x}$ to $\mathbf{y}$ is always $\leq$ the sum of distances $d(\mathbf{x},\mathbf{z}) + d(\mathbf{z},\mathbf{y})$ to any third vector $\mathbf{z} \in V$.

Proof: For all real t we have $\phi(t) = \|x+ty\|_B^2 \geq 0$. By the axioms governing B we can rewrite $\phi(t)$ as

$$\begin{aligned}\phi(t) &= B(x+ty, x+ty)\\ &= B(x,x) + B(ty,x) + B(x,ty) + B(ty,ty)\\ &= \|x\|_B^2 + t\,B(x,y) + t\,\overline{B(x,y)} + t^2\|y\|_B^2\\ &= \|x\|_B^2 + 2t\,\mathrm{Re}(B(x,y)) + t^2\|y\|_B^2\end{aligned}$$

because $B(tx,y) = tB(x,y)$ and $B(x,ty) = tB(x,y)$ (since $t \in \mathbb{R}$), and $z + \overline{z} = 2\,\mathrm{Re}(z) = 2x$ for $z = x+iy$ in $\mathbb{C}$. Now $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is a quadratic function whose minimum value occurs at t_0 where

$$\frac{d\phi}{dt}(t_0) = 2t_0\|y\|_B^2 + \mathrm{Re}(B(x,y)) = 0$$

or

$$t_0 = \frac{-\mathrm{Re}(B(x,y))}{2\|y\|_B^2}$$

Inserting this into ϕ we find the actual minimum value of ϕ:

$$0 \leq \min\{\phi(t) : t \in \mathbb{R}\} = \frac{\|x\|_B^2 \cdot \|y\|_B^2 - 2|\mathrm{Re}(B(x,y))|^2 + |\mathrm{Re}(B(x,y))|^2}{\|y\|_B^2}$$

Thus

$$0 \leq \|x\|_B^2 \cdot \|y\|_B^2 - |\mathrm{Re}(B(x,y))|^2$$

which in turn implies

$$|\mathrm{Re}\,B(x,y)| \leq \|x\|_B \cdot \|y\|_B \qquad \text{for all } x,y \in V.$$

If we replace $x \mapsto e^{i\theta}x$ this does not change $\|x\|$ since $|e^{i\theta}| = |\cos(\theta) + i\sin(\theta)| = 1$ for real θ; in the inner product on the left we have $B(e^{i\theta}x,y) = e^{i\theta}B(x,y)$. We may now take $\theta \in \mathbb{R}$ so that $e^{i\theta} \cdot B(x,y) = |B(x,y)|$. For this particular choice of θ we get

$$\begin{aligned}0 \leq |\mathrm{Re}(B(e^{i\theta}x,y))| &= |\mathrm{Re}(e^{i\theta}B(x,y))|\\ &= \mathrm{Re}(|B(x,y)|) \;=\; |B(x,y)| \leq \|x\|_B \cdot \|y\|_B \quad .\end{aligned}$$

That proves the Schwartz inequality. $\square$

Proof (Triangle Inequality): The algebra is easier if we prove the (equivalent) inequality obtained when we square both sides:

$$\begin{aligned}0 \;\leq\; \|x+y\|^2 &\leq \big(\|x\| + \|y\|\big)^2\\ &= \|x\|^2 + 2\|x\|\cdot\|y\| + \|y\|^2\end{aligned}$$

In proving the Schwartz inequality we saw that

$$\|x+y\|^2 = (x+y\,,\,x+y) = \|x\|^2 + 2\,\mathrm{Re}(x,y) + \|y\|^2$$

so our proof is finished if we can show $2\,\mathrm{Re}(x,y) \leq 2\|x\|\cdot\|y\|$. But

$$\mathrm{Re}(z) \leq |\,\mathrm{Re}(z)| \leq |z| \qquad \text{for all } z \in \mathbb{C}$$

and then the Schwartz inequality yields

$$\mathrm{Re}\big(B(x,y)\big) \leq |B(x,y)| \leq \|x\|_B \cdot \|y\|_B$$

as desired. $\square$

1.6. Example. On $V = \mathrm{M}(n,\mathbb{K})$ we define the **Hilbert-Schmidt** inner product and norm for matrices:

$$(A,B)_{\text{HS}} = \mathrm{Tr}(B^*A) \qquad \text{and} \qquad \|A\|^2_{\text{HS}} = \sum_{i,j=1} |a_{ij}|^2 = \mathrm{Tr}(A^*A) \tag{44}$$

It is easily verified that this is an inner product. First note that the trace map from $\mathrm{M}(n,\mathbb{K}) \rightarrow \mathbb{K}$

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}$$

is a complex linear map and $\mathrm{Tr}(\,\overline{A}\,) = \overline{\mathrm{Tr}(A)}$; then observe that

$$\|A\|_2^2 = (A,A)_{\text{HS}} = \sum_{i,j=1}^{n} |a_{ij}|^2 \text{ is } > 0 \text{ unless } A \text{ is the zero matrix.}$$

Alternatively, consider what happens when we identify $\mathrm{M}(n,\mathbb{C}) \cong \mathbb{C}^{n^2}$ as complex vector spaces. The Hilbert-Schmidt norm becomes the usual Euclidean norm on $\mathbb{C}^{n^2}$, and likewise for the inner products; obviously $(A,B)_{\text{HS}}$ is then an inner product on matrix space.

The norm $\|A\|_{\text{HS}}$ and the sup-norm $\|A\|_\infty$ discussed in Chapter V are different ways to measure the "size" of a matrix; the HS-norm turns out to be particularly well adapted to applications in statistics, starting with "least-squares regression" and moving on into "analysis of variance." Each of these norms determines a notion of matrix convergence $A_n \rightarrow A$ as $n \rightarrow \infty$ in $\mathrm{M}(N,\mathbb{C})$.

$$\|\cdot\|_2\text{-Convergence:} \qquad \|A_n - A\|_{\text{HS}} = \Big[\sum_{i,j} |a^{(n)}_{ij} - a_{ij}|^2\Big]^{1/2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

$$\|\cdot\|_\infty\text{-Convergence:} \qquad \|A_n - A\|_\infty = \max_{i,j}\{\,|a^{(n)}_{ij} - a_{ij}|\,\} \rightarrow 0 \text{ as } n \rightarrow \infty$$

However, despite their differences both norms *determine the same notion of matrix convergence.*

$$A_n \rightarrow A \text{ in } \|\cdot\|_2\text{-norm} \quad \Leftrightarrow \quad A_n \rightarrow A \text{ in } \|\cdot\|_\infty\text{-norm}$$

The reason is explained in the next exercise. $\square$

1.7. Exercise. Show that there exist bounds $M_2, M_\infty > 0$ such that the $\|\cdot\|_2$ and $\|\cdot\|_\infty$ norms mutually dominate each other

$$\|x\|_2 \leq M_\infty\,\|x\|_\infty \qquad \text{and} \qquad \|x\|_\infty \leq M_2\,\|x\|_2$$

for all $x \in \mathbb{C}^n$. Explain why this leads to the conclusion that $A_n \to A$ in $\|\cdot\|_2$-norm if and only if $A_n \to A$ in $\|\cdot\|_\infty$-norm.
Hint: The Schwartz inequality might be helpful in one direction.

The **polarization identities** below show that inner products over $\mathbb{R}$ or $\mathbb{C}$ can be reconstructed if we only know the norms of vectors in V. Over $\mathbb{C}$ we have

$$(45) \quad B(x,y) = \frac{1}{4}\sum_{k=0}^{3} \frac{1}{i^k} B(x+i^k y\,,\, x+i^k y) = \frac{1}{4}\sum_{k=0}^{3} \frac{1}{i^k}\,\|x+i^k y\|^2, \qquad \text{where } i=\sqrt{-1}$$

Over $\mathbb{R}$ we only need 2 terms:

$$B(x,y) = \frac{1}{4}\Big(B(x+y\,,\,x+y) + (-1)B(x-y\,,\,x-y)\Big)$$

1.8. Exercise. Expand

$$(x+i^k y\,,\, x+i^k y) = \|x+i^k y\|^2$$

to verify the polarization identities.

Orthonormal Bases in Inner Product Spaces.

A set $\mathfrak{X} = \{e_i : i \in I\}$ of vectors is **orthogonal** if $(e_i, e_j) = 0$ for $i \neq j$; it is **orthonormal** if

$$(e_i, e_j) = \delta_{ij} \quad \text{(Kronecker delta)} \qquad \text{for all } i,j \in I\ .$$

An orthonormal set can be infinite (in infinite dimensional inner product spaces), and all vectors in it are nonzero; an orthogonal family could have $v_i = 0$ for some indices since $(v, 0) = 0$ for any v. The set $\mathfrak{X}$ is an **orthonormal basis** (ON basis) if it is orthonormal and V is spanned by $\{\mathfrak{X}\}$.

1.9. Proposition. *Orthonormal sets have the following properties.*

1. *Orthonormal sets are independent;*
2. *If $\mathfrak{X} = \{e_i : i \in I\}$ is a finite orthonormal set and v is in $M = \mathbb{K}$-span$\{\mathfrak{X}\}$ then by* (1.) *$\mathfrak{X}$ is a basis for M and the expansion of any v in M with respect to this basis is just*

$$v = \sum_{i\in I} (v, e_i)\, e_i$$

   (*Finiteness of $\mathfrak{X}$ required for $\sum_{i\in I}(\ldots)$ to make sense; otherwise the right side is an infinite series*).

In particular if $\mathfrak{X} = \{e_1, ..., e_n\}$ is an orthonormal basis for a finite-dimansional inner product space V, the coefficients in the expansion

$$v = \sum_{i=1}^{n} (v, e_i)\, e_i, \qquad \textit{for every } v \in V$$

are easily computed by taking inner products.

Proof: For (1.), if a finite sum $\sum_i c_i e_i$ equals 0 we have

$$0 = (v, e_k) = \sum_i c_i (e_i, e_k) = \sum_i c_i \delta_{ik} = c_k$$

for each k, so the e_i are independent. Part (2.) is an immediate consequence of (1.): we know $\{e_i\}$ is a basis, and if $v = \sum_i c_i e_i$ is its expansion the inner product with a typical basis vector is

$$(v, e_k) = \sum_i c_i (e_i, e_k) = \sum_i c_i \delta_{ik} = c_k \ . \quad \Box$$

1.10. Corollary. *If vectors $\{v_1, ..., v_n\}$ are nonzero, orthogonal, and a vector basis in V, then the renormalized vectors*

$$e_i = \frac{v_i}{\|v_i\|} \qquad \textit{for } 1 \leq i \leq n$$

are an orthonormal basis. $\Box$

Entries in the matrix $[T]_{\mathfrak{Y}\mathfrak{X}}$ of a linear operator are easily computed by taking inner products if the bases are orthonormal (but not for arbitrary bases).

1.11. Exercise. Let $T : V \to W$ be a linear operator between finite-dimensional inner product spaces and let $\mathfrak{X} = \{e_i\}$, $\mathfrak{Y} = \{f_i\}$ be orthonormal bases. Prove that the entries in $[T]_{\mathfrak{Y}\mathfrak{X}}$ are given by

$$T_{ij} = \big(T(e_j), f_i\big)_W = \overline{\big(f_i, T(e_j)\big)_W}$$

for $1 \leq i \leq \dim(W),\ 1 \leq j \leq \dim(V)$.

The fundamental fact about ON bases is that the coefficients in $v = \sum_{k=1}^n (v, e_i)\, e_i$ determine the norm $\|v\|$ via a generalization of Pythagoras' Formula for $\mathbb{R}^n$,

$$\textsc{Pythagoras:} \qquad \text{If } \mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i \qquad \text{then} \qquad \|\mathbf{x}\|^2 = \sum_{i=1}^n |x_i|^2$$

We start by proving a fundamental inequality.

1.12. Theorem (Bessel's Inequality). *Let $\mathfrak{X} = \{e_1, \ldots, e_m\}$ be any finite orthonormal set in an inner product space V (possibly infinite-dimensional). Then*

$$\sum_{i=1}^n |(v, e_i)|^2 \leq \|v\|^2 \qquad \textit{for all } v \in V \tag{46}$$

Furthermore, if $v' = v - \sum_{i=1}^n (v, e_i)\, e_i$, this vector is orthogonal to each e_j and hence is orthogonal to all the vectors in the linear span $M = \mathbb{K}$-span$\{\mathfrak{X}\}$.

Note: The inequality (46) becomes an equality if $\mathfrak{X}$ is an orthonormal basis for V because then $v' = 0$.

Proof: Since inner products are conjugate bilinear, we have

$$\begin{aligned}
0 \;\leq\; \|v'\|^2 \;=\; (v', v') \;&=\; \Big(v - \sum_{i=1}^m (v, e_i)\, e_i \,,\, v - \sum_{j=1}^m (v, e_j)\, e_j \Big) \\
&= (v, v) - \big(\sum_i (v, e_i)\, e_i \,,\, v\big) - \big(v \,,\, \sum_j (v, e_j)\, e_j\big) + \big(\sum_i (v, e_i)\, e_i \,,\, \sum_j (v, e_j)\, e_j\big) \\
&= \|v\|^2 - \sum_i (v, e_i) \cdot (e_i, v) - \sum_j \overline{(v, e_j)} \cdot (v, e_j) + \sum_{i,j} (v, e_i) \cdot \overline{(v, e_j)} \cdot (e_i, e_j) \\
&= \|v\|^2 - \sum_i |(v, e_i)|^2 - \sum_j |(v, e_j)|^2 + \sum_i |(v, e_i)|^2 \quad (\text{since } (e_k, v) = \overline{(v, e_k)}\ ) \\
&= \|v\|^2 - \sum_i |(v, e_i)|^2
\end{aligned}$$

Therefore
$$\sum_{i=1} |(v, e_i)|^2 \leq \|v\|^2$$
as required.

The second statement now follows easily because

$$\begin{aligned}(v', e_k) &= \Big(v - \sum_j (v, e_j)\, e_j\,,\, e_k\Big) \;=\; (v, e_k) - \sum_j (v, e_j)\cdot(e_j, e_k) \\ &= (v, e_k) - (v, e_k) = 0 \qquad \text{for all } k\end{aligned}$$

Furthermore, if $w = \sum_{k=1}^m c_k e_k$ is any vector in M we also have

$$(v', w) = \sum_k c_k (v', e_k) = 0\ ,$$

so v' is orthogonal to M as claimed. $\square$

1.13. Corollary (Pythagoras). *If $\mathfrak{X}$ is an orthonormal basis in a finite dimensional inner product space, then*
$$\|v\|^2 = \sum_{i=1}^m |(v, e_i)|^2$$
(*sum of squares of the coefficients in the basis expansion $v = \sum_i (v, e_i)\, e_i$*).

1.14. Theorem. *Orthonormal bases exist in any finite dimensional inner product space.*

Proof: We argue by induction on $n = \dim(V)$; the result is trivial if $n = 1$ (any vector of length 1 is an orthonormal basis). If $\dim(V) = n + 1$, let v_0 be any nonzero vector. The linear functional $\ell_0 : v \rightarrow (v, v_0)$ is nonzero, and as in Example 1.3 of Chapter III its kernel
$$M = \{v : (v, v_0) = 0\} = \big(\mathbb{K}v_0\big)^\perp$$
is a hyperplane of dimension $\dim(V) - 1 = n$. By the induction hypothesis there is an ON basis $\mathfrak{X}_0 = \{e_1, , \ldots, e_n\}$ in M, and every vector in M is orthogonal to v_0. If we rescale v_0 and adjoin $e_{n+1} = v_0/\|v_0\|$ to $\mathfrak{X}_0$ the enlarged set $\mathfrak{X} = \{e_1, \ldots, e_n, e_{n+1}\}$ is obviously orthonormal; it is also a basis for V. [By Lemma 4.4 of Chapter III, $\mathfrak{X}$ is a basis for $W = \mathbb{K}\text{-span}\{\mathfrak{X}\} \subseteq V$, and since $\dim(W) = |\mathfrak{X}| = n + 1 = \dim(V)$ we must have $W = V$.] $\square$

VI.2. Orthogonal Complements and Projections.

If M is a subspace of a (possibly infinite-dimensional) inner product space V, its **orthogonal complement** $M^\perp$ is the set of vectors orthogonal to every vector in M,

$$M^\perp = \{\, v \in V : (v, m) = 0, \text{ for all } m \in M\,\} = \{\, v : (v, M) = \{0\}\,\}\ .$$

Obviously $\{0\}^\perp = V$ and $V^\perp = \{0\}$ from the Axioms for inner product.

2.1. Exercise. Show that $M^\perp$ is again a subspace of V, and that

$$M_1 \subseteq M_2 \;\Rightarrow\; M_2^\perp \subseteq M_1^\perp\ .$$

2.2. Proposition. *If M is a finite dimensional subspace of a (possibly infinite-dimensional) inner product space V, then*

1. *$M \cap M^\perp = \{0\}$ and $M + M^\perp = V$, so we have a direct sum decomposition $V = M \oplus M^\perp$.*

2. *If* $\dim(V) < \infty$ *we also have* $(M^\perp)^\perp = M$*; if* $|V| = \infty$ *we can only say that* $M \subseteq (M^\perp)^\perp$.

Proof: If $v \in M \cap M^\perp$ then $\|v\|^2 = (v, v) = 0$ so $v = 0$ and $M \cap M^\perp = \{0\}$. Now let $\{e_1, ..., e_n\}$ be an orthonormal basis for M. If $v \in V$ write

$$v = \Big(v - \sum_{i=0}^{m}(v, e_i)\, e_i\Big) + \sum_{i=1}^{m}(v, e_i)\, e_i \;=\; v_\perp + v_\|$$

in which $v_\perp$ is orthogonal to M and $v_\|$ is the component of v "parallel to" the subspace M (because it lies in M). Then for all $v \in V$ we have

$$(v, v_\perp) \;=\; (v_\perp + v_\| \,,\, v_\perp) \;=\; (v_\perp, v_\perp) + (v_\|, v_\perp) = \|v_\perp\|^2 + 0 \;=\; \|v_\perp\|^2$$

If $v \in (M^\perp)^\perp$, so $(v, v_\perp) = 0$, we conclude that $\|v_\perp\| = 0$ and hence $v = v_\perp + v_\| = 0 + v_\|$ is in M. That proves the reverse inclusion $M^{\perp\perp} \subseteq M$. $\square$

The situation is illustrated in Figure 6.4.

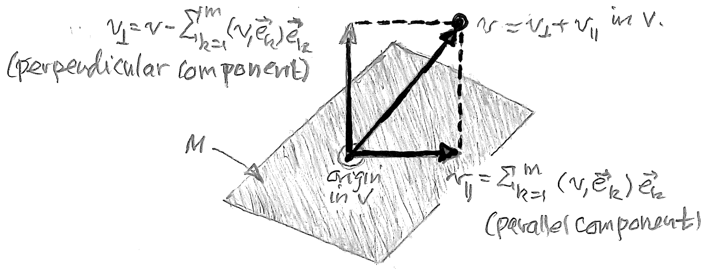

Figure 6.4. Given an ON basis $\{e_i, \ldots, e_m\}$ in a finite dimensional subspace $M \subseteq V$, the vector $v_\| = \sum_{k=1}^{m}(v, e_k)\, e_k$ is in M and $v_\perp = v - v_\|$ is orthogonal to M. These are the components of $v \in V$ "parallel to$M$" and "perpendicular to M," with $v = v_\perp + v_\|$.

Orthogonal Projections on Inner Product Spaces.

If an inner product space is a direct sum $V = V_1 \oplus \ldots \oplus V_r$ we call this an **orthogonal direct sum** if the subspaces are mutually orthogonal.

$$(V_i, V_j) = 0 \qquad \text{if } i \neq j$$

We indicate this by writing $V = V_1 \dot{\oplus} \ldots \dot{\oplus} V_r = \dot{\bigoplus}_{i=1}^{r} V_i$. The decomposition $V = M \dot{\oplus} M^\perp$ of Proposition 2.2 was an orthogonal decomposition.

In equation Exercise 3.5 of Chapter II we defined the linear projection operators $P_i : V \to V$ associated with an ordinary direct sum decomposition $V = V_1 \oplus \ldots \oplus V_r$, and showed that such operators are precisely the linear operators that have the *idempotent property* $P^2 = P$. In fact there is a bijective correspondence

$$(\textit{idempotent linear operators}) \;\longleftrightarrow\; (\textit{direct sum decompositions } V = R \oplus K)\ ,$$

described in Proposition 3.7 of Chapter II, and reprised below.

Theorem. *If a linear operator* $P : V \to V$ *is idempotent operator, so* $P^2 = P$*, there is a direct sum decomposition* $V = R \oplus P$ *such that* P *projects* V *onto* R *along* K*. In particular,*

$$R = R(P) = \text{range}(P) \qquad \textit{and} \qquad K = K(P) = \ker(P)$$

Furthermore $Q = I - P$ is also idempotent and

$$R(Q) = K(P) \qquad \textit{and} \qquad K(Q) = R(P)$$

When V is an inner product space we will see that the projections associated with an *orthogonal* direct sum $V = E\dot{\oplus}F$ have special properties. They are also easy to compute using the inner product. (Compare what follows with the calculations in Example 3.6 of Chapter II, of projections associated with an ordinary direct sum decomposition $V = E \oplus F$ in a space without inner product.)

Projections associated with an orthogonal direct sum decomposition $V = V_1\dot{\oplus}\ldots\dot{\oplus}V_r$ are called **orthogonal projections**.

2.3. Lemma. *If $V = E\dot{\oplus}F$ is an orthogonal direct sum decomposition of a finite dimensional inner product space, then*

$$E^{\perp} = F \quad \text{and} \quad F^{\perp} = E \qquad\qquad E^{\perp\perp} = E \quad \text{and} \quad F^{\perp\perp} = F$$

Proof: The argument for F is the same as that for E. We proved that $E^{\perp\perp} = E$ in Proposition 2.2 and we know that $E \subseteq F^{\perp}$ by definition; based on this we will prove the reverse inequality $E \supseteq F^{\perp}$.

Since $|V| < \infty$ we have $V = F \oplus F^{\perp}$, so that $|V| = |F| + |F^{\perp}|$; since $V = E \oplus F$ we also have $|V| = |F| + |E|$. Therefore $|E| = |F^{\perp}|$. But $E \subseteq F^{\perp}$ in an orthogonal direct sum $E\dot{\oplus}F$, so we conclude that $E = F^{\perp}$. $\square$

2.4. Exercise. Let $V = V_1\dot{\oplus}\ldots\dot{\oplus}V_r$ be an orthogonal direct sum decomposition of an inner product space (not necessarily finite dimensional).

(a) If W_i is the linear span $\sum_{j\neq i} V_j$, prove that $W_i \perp V_i$ for each i, and $V = V_i\dot{\oplus}W_i$.

(b) If $v = v_1 + \ldots + v_r$ is the unique decomposition into pairwise orthogonal vectors $v_i \in V_i$, prove that $\|v\|^2 = \sum_i \|v_i\|^2$.

The identity (2.) is yet another version of Pythagoras' formula.

2.5. Exercise. In a finite dimensional inner prodcut space, prove that the *Parseval formula*

$$(v, w) = \sum_{i=1}^{n} (v, e_i)\cdot(e_i, w)$$

holds for every orthonormal basis $\{e_1, \ldots, e_n\}$.

The Gram-Schmidt Construction.

We now show how any independent set of vectors $\{v_1, \ldots, v_n\}$ in an inner product space can be modified to obtain an orthonormal set of vectors $\{e_1, \ldots, e_n\}$ with the same linear span. This **Gram-Schmidt construction** is recursive, and at each step we have

1. $e_k \in \mathbb{K}\text{-span}\{v_1, ..., v_k\}$
2. $M_k = \mathbb{K}\text{-span}\{e_1, ..., e_k\}$ is equal to $\mathbb{K}\text{-span}\{v_1, .., v_k\}$ for each $1 \leq k \leq n$.

The result is an orthonormal basis $\{e_1, ..., e_n\}$ for $M = \mathbb{K}\text{-span}\{v_1, .., v_n\}$ (and for all of V if the $\{v_i\}$ span V). The construction procedes inductively by constructing two sequences of vectors $\{u_i\}$ and $\{e_i\}$.

Step 1: Take

$$u_1 = v_1 \qquad \text{and} \qquad e_1 = \frac{v_1}{\|v_1\|}$$

Conditions (1.) and (2.) obviously hold and $\mathbb{K}\cdot v_1 = \mathbb{K} \cdot u_1 = \mathbb{K}\cdot e_1$.

STEP 2: Define

$$u_2 = v_2 - (v_2|e_1)\cdot e_1 \qquad \text{and} \qquad e_2 = \frac{u_2}{\|u_2\|}.$$

Obviously $u_2 \in \mathbb{K}\text{-span}\{v_1, v_2\}$ and $u_2 \neq 0$ because $v_2 \notin \mathbb{K}v_1 = \mathbb{K}e_1 = M_1$; thus e_2 is well defined. Furthermore

1. $u_2 \perp M_1$ because

$$(u_2, e_1) = \big(v_2 - (v_2, e_1)e_1 \,,\, e_1\big) = (v_2, e_1) - (v_2, e_1)\cdot(e_1, e_1) = 0 \;\Rightarrow\; e_2 \perp M_1$$

   hence $\{e_1, e_2\}$ is an orthonormal set of vectors;

2. $M_2 = \mathbb{K}\text{-span}\{e_1, e_2\} = \mathbb{K}u_2 + \mathbb{K}e_1 = \mathbb{K}v_2 + \mathbb{K}e_1 = \mathbb{K}v_2 + \mathbb{K}v_1 = \mathbb{K}\text{-span}\{v_1, v_2\}$.

If $n = 2$ we're done; otherwise continue with

STEP 3: Define

$$u_3 = v_3 - \sum_{i=1}^{2}(v_3, e_i)\cdot e_i = v_3 - \sum_{i=1}^{2}\frac{(v_3, u_i)}{\|u_i\|^2}\,u_i$$

Then $u_3 \neq 0$ because the sum is in $\mathbb{K}\text{-span}\{v_1, v_2\}$ and the v_i are independent; thus $e_3 = \frac{u_3}{\|u_3\|}$ is well defined. We have $u_3 \perp M_2$ because

$$\begin{aligned}
(u_3, e_1) &= \Big(v_3 - \sum_{i=1}^{2}(v_3, e_i)e_i \,,\, e_1\Big) \\
&= (v_3, e_1) - \sum_{i=1}^{2}(v_3, e_i)\cdot(e_i, e_1) \\
&= (v_3, e_1) - (v_3, e_1) = 0\ ,
\end{aligned}$$

and similarly $(u_3, e_2) = 0$, hence $e_3 \perp M_2 = \mathbb{K}\text{-span}\{e_1, e_2\}$. Finally,

$$\begin{aligned}
\mathbb{K}\text{-span}\{e_1, e_2, e_3\} &= \mathbb{K}u_3 + \mathbb{K}\text{-}\{e_1, e_2\} \;=\; \mathbb{K}v_3 + \mathbb{K}\text{-}\{e_1, e_2\} \\
&= \mathbb{K}v_3 + \mathbb{K}\text{-}\{v_1, v_2\} \;=\; \mathbb{K}\text{-}\{v_1, v_2, v_3\}
\end{aligned}$$

At the k^{th} step we have produced orthonormal vectors $\{e_1, ..., e_k\}$ with $\mathbb{K}\text{-span}\{e_1, ..., e_k\} = \mathbb{K}\text{-span}\{v_1, ..., v_k\} = M_k$. Now for the induction step:

STEP $k+1$: Define

$$u_{k+1} = v_{k+1} - \sum_{i=1}^{k}(v_{k+1}, e_i)\,e_i = v_{k+1} - \sum_{i=1}^{k}\frac{(v_{k+1}, u_i)}{\|u_i\|^2}\,u_i$$

and

$$e_{k+1} = \frac{u_{k+1}}{\|u_{k+1}\|}.$$

Again $u_{k+1} \neq 0$ because $v_{k+1} \notin M_k = \mathbb{K}\text{-span}\{v_1, ..., v_k\} = \mathbb{K}\text{-span}\{e_1, ..., e_k\}$, so e_{k+1} is well defined. Furthermore $u_{k+1} \perp M_k$ because

$$\begin{aligned}
(u_{k+1}, e_j) &= \Big(v_{k+1} - \sum_{i=1}^{k}(v_{k+1}, e_i)\,e_i \,,\, e_j\Big) \\
&= (v_{k+1}, e_j) - \sum_{i=1}^{k}(v_{k+1}, e_i)\cdot(e_i, e_j) \\
&= (v_{k+1}, e_j) - (v_{k+1}, e_j) = 0
\end{aligned}$$

hence also $e_{k+1} \perp M_k$. Then

$$\begin{aligned}\mathbb{K}\text{-}\{e_1, ..., e_{k+1}\} &= \mathbb{K}u_{k+1} + \mathbb{K}\text{-}\{e_1, ..., e_k\} = \mathbb{K}v_{k+1} + \mathbb{K}\text{-}\{e_1, ..., e_k\} \\ &= \mathbb{K}\text{-}\{v_1, ..., v_{k+1}\} \ .\end{aligned}$$

By induction, $\{e_1, ..., e_n\}$ has the properties claimed. $\square$

Note that the outcome of Step(k+1) depends only on the $\{e_1, ..., e_k\}$ and the new vector v_{k+1}; the original vectors $\{v_1, ..., v_k\}$ play no further role in the inductive process.

2.6. Example. The standard inner product in $\mathcal{C}[-1, 1]$ is the L^2 inner product

$$(f, h)_2 = \int_{-1}^{1} f(t)\overline{h(t)}\, dt$$

for functions $f : [-1, 1] \to \mathbb{C}$. Regarding $v_1 = \mathbb{1}$, $v_2 = x$, $v_3 = x^2$ as functions from $[-1, 1] \to \mathbb{C}$, these vectors are independent. Find the orthonormal set $\{e_1, e_2, e_3\}$ produced by the Gram-Schmidt process.

Solution: We have $u_1 = v_1 = \mathbb{1}$ and since $\|u_1\|^2 = \int_{-1}^{1} \mathbb{1}\, dx = 2$, we get $e_1 = \frac{1}{\sqrt{2}} \cdot \mathbb{1}$. At the next step

$$u_2 = v_2 - (v_2, e_1)\, e_1 = v_2 - \frac{(v_2, u_1)}{\|u_1\|^2}\, u_1 = x - \frac{\int_{-1}^{1} x \cdot \mathbb{1}\, dx}{\|u_1\|^2} \cdot \mathbb{1} = x - 0 = x$$

and

$$\|u_2\|^2 = \int_{-1}^{1} x^2\, dx = 2\int_{0}^{1} x^2\, dx = 2\left[\frac{1}{3}x^3\Big|_0^1\right] = \frac{2}{3}$$

The second basis vector is

$$e_2 = \frac{u_2}{\|u_2\|} = \sqrt{\frac{3}{2}} \cdot x$$

At the next step:

$$\begin{aligned}u_3 &= v_3 - \big((v_3|e_1)e_1 + (v_3, e_2)e_2\big) \\ &= v_3 - \Big(\frac{(v_3, u_2)}{\|u_2\|^2} \cdot u_2 + \frac{(v_3, u_1)}{\|u_1\|^2} \cdot u_1\Big) \\ &= x^2 - \frac{\int_{-1}^{1} x^2 \cdot x\, dx}{\frac{2}{3}} \cdot x - \frac{\int_{-1}^{1} x^2 \cdot \mathbb{1}\, dx}{2} \cdot \mathbb{1} \\ &= x^2 - 0 - \frac{1}{3}\mathbb{1} = x^2 - \frac{1}{3}\end{aligned}$$

Then

$$\begin{aligned}\|u_3\|^2 &= \int_{-1}^{1} |u_3(x)|^2\, dx = \int_{-1}^{1} \big(x^2 - \frac{1}{3}\big)^2\, dx \\ &= \int_{-1}^{1} \Big(x^4 - \frac{2}{3}x^2 + \frac{1}{9}\Big)dx \\ &= 2 \cdot \left[\frac{x^5}{5} - \frac{2}{9}x^3 + \frac{1}{9}x\,\Big|_0^1\right] = \frac{8}{45}\end{aligned}$$

and the third orthonormal basis vector is

$$e_3 = \frac{u_3}{\|u_3\|} = \sqrt{\frac{45}{8}}\Big(x^2 - \frac{1}{3}\Big) = \sqrt{\frac{5}{8}}(3x^2 - 1) \quad \square$$

If we extend the original list to include $v_4 = x^4$ we may compute e_4 knowing only e_1, e_2, e_3 (or u_1, u_2, u_3) and v_4; there is no need to repeat the previous calculations!

2.7. Exercise. Find u_4 and e_4 in the above situation.

This process can be continued indefinitely to produce the orthonormal family of **Legendre polynomials** $e_1(t), e_2(t), \ldots, e_n(t)\ldots$ in the space of polynomials $\mathbb{C}[x]$ restricted to the interval $[-1, 1]$. (This is also true for $\mathbb{R}[x]$ restricted to $[-1, 1]$ since the Legendre polynomials all have real coefficients.) Clearly the $(n + 1)$-dimensional subspace M_n obtained by restricting the space of polynomials of degree $\leq n$

$$\mathcal{P}_n = \mathbb{K}\text{-span}\{e_1, ..., e_{n+1}\} = \mathbb{K}\text{-span}\{\mathbb{1}, x, \ldots, x^n\}$$

to the interval (so $M_n = \mathcal{P}_n|[-1, 1]$) has $\{e_1, \ldots, e_{n+1}\}$ as an ON basis with respect to the usual inner product on $\mathcal{C}[-1, 1]$

$$(f, h)_2 = \int_{-1}^{1} f(t)\overline{h(t)}\, dt\ .$$

Restricting the full set of Legendre polynomials $e_1(t), \ldots, e_{n+1}(t), \ldots$ to $[-1, 1]$ yields an orthonormal set of vectors in the infinite-dimensional inner product space $\mathcal{C}[-1, 1]$. The orthogonal projection $P_n : \mathcal{C}[-1, 1] \rightarrow M_n \subseteq \mathcal{C}[-1, 1]$ associated with the orthogonal direct sum decomposition $V = M_n \oplus (M_n)^{\perp}$ (in which $\dim(M_n)^{\perp} = \infty$) is given by the explicit formula

$$\begin{aligned}
P_n f(t) &= \sum_{k=1}^{n+1} (f, e_k)\, e_k(t) \qquad (-1 \leq t \leq 1) \\
&= \sum_{k=1}^{n+1} \Big(\int_{-1}^{1} f(x)\overline{e_k(x)}\, dx \Big) \cdot e_k(t) \\
&= \sum_{k=0}^{n} c_k\, t^k \qquad (c_k \in \mathbb{C})
\end{aligned}$$

for any continuous function on $[-1, 1]$. The projected image $P_n f(t)$ is a polynomial of degree $\leq n$ even though $f(t)$ is continuous and need not be differentiable.

A standard result from analysis shows that the partial sums of the infinite series $\sum_{k=0}^{\infty} c_k\, t^k$ converge in the L^2-norm to the original function $f(t)$ throughout the interval $-1 \leq t \leq 1$,

$$\|f - P_n f\|_2 = \Big[\int_{-1}^{1} |f(t) - P_n f(t)|^2\, dt \Big]^{1/2} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

for all $f \in \mathcal{C}[-1, 1]$.

It must be noted that this series expansion of $f(t) \sim \sum_{k=0}^{\infty} c_k\, t^k$ is not at all the same thing as a Taylor series expansion about $t = 0$, which in any case would not make sense because $f(t)$ is only assumed continuous (the derivatives used to compute Taylor coefficients might not exist!) In fact, convergence of this series in the L^2-norm is much more robust than convergence of Taylor series, which is why it is so useful in applications.

Fourier Series Expansions.

The *complex trig polynomials* $E_n(t) = e^{2\pi i n t}$ $(n \in \mathbb{Z})$ are periodic complex-valued functions on $\mathbb{R}$; each has period $\Delta t = 1$ since

$$e^{2\pi i n(t+1)} = e^{2\pi i n t} \cdot e^{2\pi i n} = e^{2\pi i n t} \qquad \text{for all } t \in \mathbb{R} \text{ and } n \in \mathbb{Z}.$$

If $e_n(t)$ is the restriction of $E_n(t)$ to the "period-interval" $I = [0, 1]$ we get an ON family of vectors with respect to the usual inner product $(f, h) = \int_0^1 f(t)\overline{h(t)}\,dt$ on $\mathcal{C}[0, 1]$, because

$$\begin{aligned}
\|e_n\|^2 &= \int_0^1 |e_n(t)|^2\,dt \;=\; \int_0^1 1\,dt = 1 \\
(e_m, e_n) &= \int_0^1 e_m(t)\overline{e_n(t)}\,dt \;=\; \int_0^1 e^{2\pi i(m-n)t}dt \\
&= \left[\frac{e^{2\pi i(m-n)t}}{2\pi i(m-n)}\Big|_0^1\right] = 0 \qquad \text{if} \;\; m \neq n.
\end{aligned}$$

Thus $\{e_n : n \in \mathbb{Z}\}$ is an orthonormal family in $\mathcal{C}[0, 1]$.

For $N \geq 0$ let $M_N = \mathbb{K}\text{-span}\{e_k : -N \leq k \leq N\}$. For f in this subspace we have the basis expansion:

$$f = \sum_{k=-N}^{N} (f, e_k)\, e_k = \sum_{k=-N}^{N} c_k e^{2\pi i k t}$$

where c_k is the k^{th} **Fourier coefficient**

$$c_k = (f, e_k) = \int_0^1 f(t)e^{-2\pi i k t}dt. \tag{47}$$

By Bessel's inequality:

$$\|f\|_2^2 = \int_0^1 |f(t)|^2 dt \geq \sum_{k=-N}^{N} |c_k|^2 = \sum_{k=-N}^{N} |(f, e_k)|^2$$

and this is true for $N = 1, 2, ...$ The projection P_N of $\mathcal{C}[0, 1]$ onto M_N along $M_N^{\perp}$ is then given by

$$P_N f(t) = \sum_{k=-N}^{N} c_k e_k(t) = \sum_{k=-N}^{N} (f, e_k)\, e^{2\pi i k t}, \qquad N = 0, 1, 2, ...$$

because $P_N(f) \in M_N$ by definition, and $\big(f - P_N f, e_k\big) = 0$ for $-N \leq k \leq N$.

The **Fourier series** of a continuous (or bounded Riemann integrable) complex-valued function $f : [0, 1] \to \mathbb{C}$ is the infinite series

$$f \sim \sum_{k \in \mathbb{Z}} (f, e_k) \cdot e^{2\pi i k t} \tag{48}$$

whose coefficients $c_k = (f, e_n)$ are the Fourier coefficients defined in (47).

It is not immediately clear when this series converges, but when convergence is suitably interpreted it can be proved that the series does converge, and to the initial function $f(t)$. This expansion has proved to be extremely useful in applications. Its significance is best described as follows.

If t is regarded as a time variable, and $F(t)$ is some sort of periodic "signal" or "waveform" such that $F(t + 1) = F(t)$ for all t, then F is completely determined by its restriction $f = F|\,[0, 1]$ to the basic period interval $0 \leq t \leq 1$. The Fourier series expansion of f on this interval can in turn be regarded as a representation of the original waveform as a "superposition," with suitable weights, of the basic periodic waveforms $E_n(t) = e^{2\pi i n t}$ $(t \in \mathbb{R})$.

$$F(t) \sim \sum_{n=-\infty}^{+\infty} c_n \cdot E_n(t) \qquad \text{for all } t \in \mathbb{R}$$

For instance, this implies that any periodic sound wave $F(t)$ with period $\Delta t = 1$ can be reconstructed by superposing scalar multiples of the "pure tones" $E_n(t)$, which have frequencies $\omega_n = n$ cycles per second. This is precisely how sound synthesizers work. It is remarkable, that the correct "weight" assigned to each pure tone is the Fourier coefficient $c_n = (f, e_n)$; even more remarkable is the fact that *complex-valued* weights $c_k \in \mathbb{C}$ must be allowed, even if the signal is real-valued, because the functions $E_n(t) = \cos(2\pi nt) + i\sin(2\pi nt)$ are complex-valued.

If f is piecewise differentiable the infinite series (48) converges (except at points of discontinuity) to the original periodic function $f(t)$. Furthermore the following results can be proved for any continuous (or Riemann integrable) function on $[0, 1]$.

Theorem. *If $f(t)$ is bounded and Riemann integrable for $0 \leq t \leq 1$, then*

1. L^2-Norm Convergence: *The partial sums of the Fourier series* (48) *converge to $f(t)$ in the L^2-norm.*

$$\Big\|f - \sum_{k=-N}^{N} (f, e_k)\, e_k\Big\|_2 \to 0 \;\; \textit{as } N \to \infty$$

2. Extended Bessel: $\|f\|^2 = \int_0^1 |f(t)|^2\, dt$ *is equal to* $\sum_{k\in\mathbb{Z}} |(f, e_k)|^2$.

The norm $\|f - h\|_2 = \Big[\int |f - h|^2\, dt\Big]^{1/2}$ is often referred to as the "RMS = *Root Mean Square*" distance between f and h.

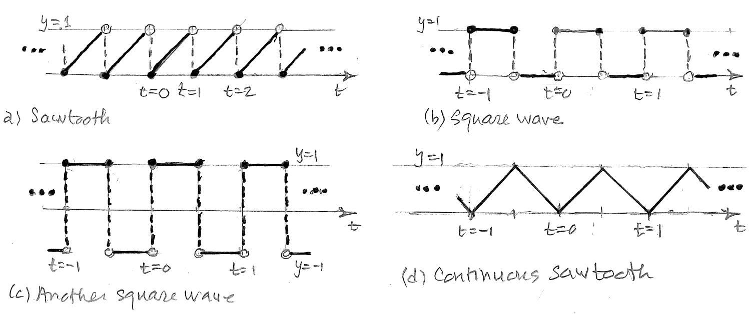

Figure 6.5. Various waveforms with period $\Delta t = 1$, whose Fourier transforms can be computed by Calculus methods.

2.8. Example. Let

$$f(t) = \begin{cases} t & \text{for } 0 \leq t < 1 \\ 0 & \text{for } t = 1 \end{cases}$$

This is the restriction to $[0, 1]$ of the periodic "sawtooth" waveform in Figure 6.5(a). Find its Fourier series.

Solution: If $k \neq 0$ integration by parts yields

$$\begin{aligned}
c_k &= \int_0^1 t e^{-2\pi i k t}\, dt \\
&= \left[\frac{-1}{2\pi i k} e^{-2\pi i k t} \cdot t \Big|_0^1 \right] - \int_0^1 \frac{-1}{2i\pi k} e^{-2\pi i k t}\, dt \\
&= \frac{-1}{2\pi i k} + \frac{1}{2\pi i k}(e_k, e_0) \qquad (\text{where } e_0(t) \equiv 1 \text{ for all } t) \\
&= \frac{-1}{2\pi i k} \qquad \text{if } k \neq 0 \ .
\end{aligned}$$

For $k = 0$ we get a different result:

$$c_0 = \int_0^1 t\, dt = \tfrac{1}{2}$$

By Bessel's Inequality we have

$$\begin{aligned}
\|f\|_2^2 &= \int_0^1 |f(t)|^2\, dt \ = \int_0^1 t^2\, dt = \tfrac{1}{3} \quad (\text{by direct calculation}) \\
&\geq \sum_{k=-N}^{N} |(f, e_k)|^2 \ = \sum_{k=-N}^{N} |c_k|^2 \\
&= \frac{1}{4} + \sum_{k \neq 0, -N \leq k \leq N} \frac{1}{4\pi^2 k^2}
\end{aligned}$$

for any $N = 1, 2, ...$ If we multiply both sides by $4\pi^2$, then for all N we get

$$\begin{aligned}
\frac{4}{3}\pi^2 &\geq \sum_{0<|k|\leq N} \frac{1}{k^2} + \pi^2 \\
\frac{1}{3}\pi^2 &\geq 2 \cdot \sum_{k=1}^{N} \frac{1}{k^2} \\
\frac{\pi^2}{6} &\geq \sum_{k=1}^{N} \frac{1}{k^2} \quad \text{for all } N = 1, 2, \ldots \ \Rightarrow \ \frac{\pi^2}{6} \geq \sum_{k=1}^{\infty} \frac{1}{k^2}
\end{aligned}$$

(the infinite series converges by the Integral Test). Once we know that $\|f\|^2 = \sum_{k\in\mathbb{Z}} |c_k|^2$ we get the famed formula

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

The Fourier series associated with the sawtooth function $f(t)$ is

$$f(t) \sim \sum_{k=-\infty}^{\infty} (f, e_k)\, e_k(t) = \frac{1}{2} \cdot \mathbb{1} + \sum_{k \neq 0} \frac{-1}{2\pi i k} e^{2\pi i k t} \ ,$$

which converges pointwise for all $t \in \mathbb{R}$ except the "jump points" $t \in \mathbb{Z}$, where the series converges to the middle value $\frac{1}{2}$. $\square$

2.9. Exercise. Compute the Fourier transforms of the periodic functions whose graphs are shown in Figure 6.5 (b) – (d).

A Geometry Problem. The following result provides further insight into the meaning of the projection $P_N(v) = \sum_{i=1}^{N}(v, e_i)\, e_i$ where $\{e_i\}$ is an orthonormal family in an inner product space V.

2.10. Theorem. *If $\{e_1, \ldots, e_n\}$ is an orthonormal family in an inner product space, and $P_M(v) = \sum_{i=1}^{n}(v, e_i)\, e_i$ the projection of v onto $M = \mathbb{K}$-span$\{e_1, ..., e_n\}$ along $M^{\perp}$, then the image $P_M(v)$ is the point in M closest to v,*

$$\|P_M(v) - v\| = \min\{\, \|u - v\| : u \in M\}$$

for any $v \in V$. In particular the minimum is achieved at the unique point $P_M(v) \in M$.

Proof: Write $v = v_{\|} + v_{\perp}$ where $v_{\|} = P_M(v) = \sum_{i=1}^{N}(v, e_i)\, e_i$ and $v_{\perp} = v - \sum_i (v, e_i)\, e_i$. Obviously $v_{\|} \perp v_{\perp}$ and if z is any point in M we have $(v_{\|} - z) \in M$ and $(v - v_{\|}) \perp M$, so by Pythagoras

$$\begin{aligned} \|v - z\|^2 &= \|(v - v_{\|}) + (v_{\|} - z)\|^2 \\ &= \|v - v_{\|}\|^2 + \|v_{\|} - z\|^2 \end{aligned}$$

Thus

$$\|v - z\|^2 \geq \|v - v_{\|}\|^2$$

for all $z \in M$, so $\|v - z\|^2$ is minimized at $z = v_{\|} = \sum_{i=1}^{N}(v, e_i)\, e_i$. Figure 6.6 shows why the formula $\|v\|^2 = \|v_{\|}\|^2 + \|v_{\perp}\|^2$ really is equivalent to Pythagora's formula for right triangle (see the shaded triangle). $\square$

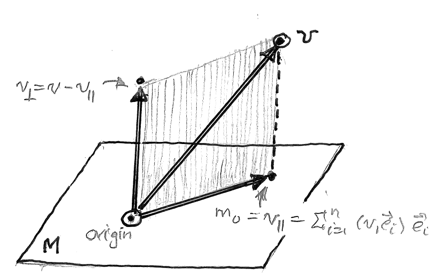

Figure 6.6. If M is a finite dimensional subspace of inner product space V and $v \in V$, the unique point in M closest to v is $m_0 = v_{\|} = \sum_i (v, e_i)\, e_i$, and the minimized distance is $\|v - m_{)}\|$. The shaded plane is spanned by the orthogonal vectors $v_{\|}$ and $v_{\perp}$ and we have $\|v\|^2 = \|v_{\|}\|^2 + \|v_{\perp}\|^2$ (Pythagoras' formula).

V.3. Adjoints and Orthonormal Decompositions.

Let V be a finite dimensional inner product space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. Recall that a linear operator $T : V \to V$ is *diagonalizable* if there is a basis $\{e_1, \ldots, e_n\}$ of eigenvectors (so $T(e_i) = \mu_i e_i$ for some $\mu_i \in \mathbb{K}$). We have seen that this happens if and only if $V = \bigoplus_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$ where

$$\begin{aligned} \mathrm{sp}(T) &= (\text{the distinct eigenvalues of } T \text{ in } \mathbb{K}) = \{\lambda \in \mathbb{K} : E_\lambda(T) \neq (0)\} \\ E_\lambda(T) &= \{v \in V : (T - \lambda I)v = 0\} = \ker(T - \lambda I) \end{aligned}$$

We say T is **orthogonally diagonalizable** if there is an orthonormal basis $\{e_1, \ldots, e_n\}$ of eigenvectors, so $T(e_i) = \mu_i e_i$ for some $\mu_i \in \mathbb{K}$.

3.1. Lemma. *A linear operator $T : V \to V$ on a finite dimensional inner product space is orthogonally diagonalizable if and only if the eigenspaces span V and are pairwise* *orthogonal, so* $E_\lambda(T) \perp E_\mu(T)$ *for* $\lambda \neq \nu$ *in* $\mathrm{sp}(T)$.

Proof ($\Leftarrow$): is easy. We have seen that the span $W = \sum_{\lambda\in\mathrm{sp}(T)} E_\lambda(T)$ is a direct sum whether or not $W = V$. If $W = V$ and the E_λ are orthogonal then we have an orthogonal direct sum decomposition $V = \dot{\bigoplus}_\lambda E_\lambda(T)$. Taking an orthonormal basis in each E_λ we get a diagonalizing orthonormal basis for all of V.

Proof ($\Rightarrow$): If $\mathfrak{X} = \{e_1, \ldots, e_n\}$ is a diagonalizing orthonormal basis with $T(e_i) = \mu_i e_i$, each μ_i is an eigenvalue. Define

$$\mathrm{sp}' = \{\lambda \in \mathrm{sp}(T) : \lambda = \mu_i \text{ for some } i\} \subseteq \mathrm{sp}(T)$$

and for $\lambda \in \mathrm{sp}(T)$ let

$$M_\lambda = \sum \{\mathbb{K}e_i : \mu_i = \lambda\} \subseteq E_\lambda(T)$$

(which will $= (0)$ if λ does not appear among the scalars μ_i). Obviously $|M_\lambda| \leq |E_\lambda|$; furthermore, each e_i lies in some eigenspace E_λ, so

$$V = \mathbb{K}\text{-span}\{e_1, \ldots, e_n\} \subseteq \sum_{\lambda\in\mathrm{sp}(T)} E_\lambda \subseteq V$$

and these subspaces coincide. Thus

$$|V| = \sum_{\lambda\in\mathrm{sp}(T)} |E_\lambda| \geq \sum_{\lambda\in\mathrm{sp}'} |E_\lambda| \geq \sum_{\lambda\in\mathrm{sp}'} |M_\lambda| \ \geq\ |V|$$

and all sums are equal. (The last inequality holds because $\sum_{\lambda\in\mathrm{sp}'} M_\lambda \supseteq \sum_{j=1}^n \mathbb{K}e_j = V$.)

Now if $\mathrm{sp}(T) \neq \mathrm{sp}'$ the first inequality would be strict, and if $M_\lambda \subsetneqq E_\lambda$ the second the second would be strict, both impossible. We conclude that $|M_\lambda| = |E_\lambda(T)|$ so $M_\lambda = E_\lambda(T)$. But the M_λ are mutually orthogonal by definition, so the eigenspaces E_λ are pairwise orthogonal as desired. $\square$

Simple examples (discussed later) show that a linear operator on an inner product space can be diagonalizable in the ordinary sense but fail to be orthogonally diagonalizable. To explore this distinction further we need additional background, particularly the definition of *adjoints* of linear operators.

Dual Spaces of Inner Product Spaces. There is a natural identification of any finite dimensional inner product space V with its dual space V^*. It is implemented by a map $J : V \to V^*$ where $J(v) =$ the functional $\ell_v \in V^*$ such that

$$\langle \ell_v, x\rangle = (x, v) \qquad \text{for all } x \in V\ .$$

Each map ℓ_v is a linear functional because the inner product $(*, *)$ is $\mathbb{K}$-linear in its left hand entry (but conjugate linear in the right hand entry unless $\mathbb{K} = \mathbb{R}$). The map J is one-to-one because

$$J(v_1) = J(v_2) \ \Rightarrow\ 0 = \langle \ell_{v_1}, x\rangle - \langle \ell_{v_2}, x\rangle = (x, v_1) - (x, v_2) = (x, v_1 - v_2)$$

for all $x \in V$. Taking $x = v_1 - v_2$, we get $0 = \|v_1 - v_2\|^2$ which implies $v_1 - v_2 = 0$ and $v_1 = v_2$ by positive definiteness of the inner product. To see J is also surjective we invoke:

3.2. Lemma. *If V is finite dimensional inner product space, $\{e_1, \ldots, e_n\}$ an orthonormal basis, and $\ell \in V^*$, then*

$$\ell = J(v_0) \qquad \textit{where} \qquad v_0 = \sum_{i=1}^n \overline{\langle \ell, e_i\rangle}\, e_i$$

(*proving J surjective*).

Proof: For any $x \in V$ we have $x = \sum_i (x, e_i)\, e_i$. Hence by conjugate-linearity of $(*, *)$

$$\begin{aligned} \langle J(v_0), x\rangle &= (x, v_0) = \Big(\sum_i x_i e_i\,,\, \sum_j \overline{\langle \ell, e_j\rangle}\, e_j\Big) = \sum_{i,j} x_i \langle \ell, e_j\rangle \cdot (e_i, e_j) \\ &= \sum_i x_i \langle \ell, e_i\rangle = \Big\langle \ell, \sum_i x_i e_i\Big\rangle = \ell(x) \qquad \text{for all } x \in V. \end{aligned}$$

Therefore $J(v_0) = \ell$ as elements of V^*. $\square$

3.3. Exercise. Prove that $J : V \to V^*$ is a *conjugate linear* bijection: it is additive, with $J(v + v') = J(v) + J(v')$ for all $v, v' \in V$, but $J(\lambda v) = \overline{\lambda} J(v)$ for $v \in V$, $\lambda \in \mathbb{C}$.

The Adjoint Operator $\mathbf{T^*}$. If $T : V \to W$ is a linear operator between finite dimensional vector spaces we showed that there is a natural transpose $T^{\mathrm{t}} : W^* \to V^*$. Since $V \cong V^*$ for inner product spaces, it follows that there is a natural *adjoint operator* $T^* : V \to W$ between the original vector spaces, rather than their duals.

3.4. Theorem (Adjoint Operator). *Let V, W be finite dimensional inner product spaces and $T : V \to W$ a $\mathbb{K}$-linear operator. Then there is a unique $\mathbb{K}$-linear* **adjoint operator** $T^* : W \to V$ *such that*

$$(T(v), w)_W = (v, T^*(w))_V \qquad \text{for all } v \in V,\ w \in W \quad , \tag{49}$$

or equivalently $(T^*(w), v)_V = (w, T(v))_W$ owing to Hermitian symmetry of the inner product.

Proof: We define $T^*(w)$ for $w \in W$ using our observations about dual spaces. Given $w \in W$, we get a well defined linear functional ϕ_w on V if define

$$\langle\, \phi_w, v\rangle = (\, T(v), w)_W$$

(w is fixed; the variable is v).

Obviously $\phi_w \in V^*$ because $(*, *)_W$ is linear in its left-hand entry. By the previous discussion there is a unique vector in V, which we label $T^*(w)$, such that $J(T^*(w)) = \phi_w$ in V^*, hence

$$(T(x), w)_W = \langle \phi_w, x\rangle = \langle J(T^*(w)), x\rangle = (x, T^*(w))_V$$

We obtain a well defined map $T^* : W \to V$.

Once we know a map T^* satisfying (49) exists, it is easy to use these scalar identities to verify that T^* is a linear operator, and verify its important properties. For linearity we first observe that two vectors v_1, v_2 are equal in V if and only if $(v_1, x) = (v_2, x)$, for all $x \in V$ because the inner product is positive definite.

Then $T^*(w_1 + w_2) = T^*(w_1) + T^*(w_2)$ in V follows: for all $v \in V$ we have

$$\begin{aligned} (T^*(w_1 + w_2), v)_V &= (w_1 + w_2, T(v))_W = (w_1, T(v))_W + (w_2, T(v))_W \\ &= (T^*(w_1), v)_V + (T^*(w_2, v)_V \qquad \text{(definition of } T^*(w_k)) \\ &= (T^*(w_1) + T^*(w_2), v)_V \quad \text{(linearity of } (*|*) \text{ in first entry)} \end{aligned}$$

Similarly, $T^*(\lambda w) = \lambda T^*(w)$, for all $\lambda \in \mathbb{K}$, $w \in W$ (check that λ comes forward instead of $\overline{\lambda}$). $\square$

Note: A general philosophy regarding calculations with adjoints: Don't look at $T^*(v)$; look at $(T^*(v), w)$ instead, for all $v \in V, w \in W$.

3.5. Lemma. *On an inner product space $(T^*)^* = T$ as linear maps from $V \to W$.*

Proof: It suffices to check the scalar identities $(T^{**}(v), w)_W = (T(v), w)_W$, for all $v \in V$, $w \in W$. But by definition,

$$\big(T^{**}(v), w\big)_W = \big(v, T^*(w)\big)_V = (T(v), w)_W$$

Done. $\square$

The adjoint $T^* : W \to V$ of a linear operator $T : V \to W$ between inner product space is analogous to the transpose $T^{\mathrm{t}} : W^* \to V^*$. In fact, if V, W are inner product spaces and we identify $V = V^*$, $W = W^*$ via the maps $J_V : V \to V^*$, $J_W : W \to W^*$ then T^* *becomes* the transpose $T^{\mathrm{t}} : W^* \to V^*$ in the sense that the following diagram commutes:

$$\begin{array}{ccc} W & \xrightarrow{T^*} & V \\ J_W \downarrow & & \downarrow J_V \\ W^* & \xrightarrow{T^{\mathrm{t}}} & V^* \end{array}$$

That is ,

$$T^t \circ J_W = J_V \circ T^* \qquad (\text{ or } T^* = J_V^{-1} \circ T^{\mathrm{t}} \circ J_W)$$

3.6. Exercise. Prove this last identity from the definitions.

Furthermore, as remarked earlier, when V is just a vector space, there is a natural identification of $V \cong V^{**}$

$$j : V \to V^{**} \qquad \langle j(v), \ell \rangle = \ell(v) \quad \text{for all } \ell \in V^*, v \in V$$

We remarked that under this identification of $V^{**} \cong V$ we have $T^{\mathrm{tt}} = T$ for any linear operator $T : V \to W$, in the sense that the following diagram commutes

$$\begin{array}{ccc} V^{**} & \xrightarrow{T^{\mathrm{tt}}} & W^{**} \\ j_V \uparrow & & \uparrow j_W \\ V & \xrightarrow{T} & W \end{array}$$

If V, W are inner product spaces, we may actually identify $V \simeq V^*$ (something that cannot be done in any natural way in the absence of the extra structure an inner product provides). Then we may identify $V \cong V^* \cong V^{**} \cong V^{***} \cong ...$ and $W \cong W^* \cong W^{**} \cong W^{***} \cong ...$; when we do, T^{t} becomes T^* and T^{tt} becomes $T^{**} = T$.

3.7. Exercise (Basic Properties of Adjoints). Use (49) to prove:

(a) $I^* = I$ and $(\lambda I)^* = \overline{\lambda} I$,

(b) $(T_1 + T_2)^* = T_1^* + T_2^*$,

(c) $(\lambda T)^* = \bar{\lambda} T^*$ (conjugate-linearity)

3.8. Exercise. Given linear operators $V \xrightarrow{S} W \xrightarrow{T} Z$ between finite dimensional inner product spaces, prove that

$$(T \circ S)^* = S^* \circ T^* : Z \to V \ .$$

Note the reversal of order when we take adjoints.

3.9. Exercise. If $A \in \mathrm{M}(n, \mathbb{C})$ and $(A^*)_{ij} = \overline{A_{ji}}$ is the usual adjoint matrix, consider the operator $L_A : \mathbb{C}^n \to \mathbb{C}^n$ such that $L_A(\mathbf{z}) = A \cdot \mathbf{z}$. If $\mathbb{C}^n$ is given the standard inner product prove that

(a) If $\mathfrak{X} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the standard orthonormal basis then $[L_A]_{\mathfrak{X}\mathfrak{X}} = A$.

(b) $\big(L_A\big)^* = L_{A^*}$ as operators on $\mathbb{C}^n$.

3.10. Example (Self-Adjointness of Orthogonal Projections). On an unadorned vector space V the "idempotent" relation $P^2 = P$ identifies the linear operators that are projections associated with an ordinary direct sum decomposition $V = M \oplus N$. The same is true of an inner product space, but if we only know $P = P^2$ the subspaces M, N are not necessarily orthogonal. We now show that an idempotent operator P on an inner product space corresponds to an *orthogonal* direct sum decomposition $V = M \dot{\oplus} N$ if and only if it is **self-adjoint** ($P^* = P$), so that

$$P^2 = P = P^* \tag{50}$$

Discussion: If $M \perp N$ it is fairly easy to verify (Exercise 3.11) that the associated projection P_M of V onto $M = \mathrm{range}(P_M)$ along $N = \ker(P_M)$ is self-adjoint. If $v, w \in V$, let us indicate the components by writing $v = v_M + v_N$, $w = w_M + w_N$. With (49) in mind, self-adjointness of P_M emerges from the following calculation.

$$\begin{aligned}
\big(v, P_M^*(w)\big) &= \big(P_M(v), w\big) = (v_M, w_M + w_N) \quad (\text{definition of } P_M(v) = v_M) \\
&= (v_M, w_M) \qquad\qquad (\text{since } w_N \perp w_M) \\
&= (v_M + v_N, w_M) = (v, w_M) = \big(v, P_M(w)\big)
\end{aligned}$$

Since the is true for all $v \in V$ we get $P_M^*(w) = P_M(w)$ for all w, whence $P_M^* = P_M$ as operators.

For the converse we must prove: If the projection P_M associated with an ordinary direct sum decomposition $V = M \oplus N$ is self-adjoint, so that $P_M^* = P_M$, then the subspaces must be orthogonal. We leave this proof as an exercise. $\square$

3.11. Exercise. If $P : V \to V$ is a linear operator on a vector space such that $P^2 = P$ it is the projection operator associated with the decomposition

$$V = R \oplus K \qquad \text{where} \qquad R = \mathrm{range}(P),\ K = \ker(P)$$

If V is an inner product space prove that the subspaces must be orthogonal ($R \perp K$) if the projection is self-adjoint, so $P^2 = P = P^*$. $\square$

Matrix realizations of adjoints are easily computed, provided we restrict attention to *orthonormal* bases in both V and W. With respect to arbitrary bases the computation of $[T^*]_{\mathfrak{X}\mathfrak{Y}}$ can be quite a mess.

3.12. Proposition. *Let $T : V \to W$ be a linear operator between finite dimensional inner product spaces and let $\mathfrak{X} = \{e_i\}$, $\mathfrak{Y} = \{f_j\}$ be orthonormal bases in V, W. Then*

$$[T^*]_{\mathfrak{X}\mathfrak{Y}} = \big([T]_{\mathfrak{Y}\mathfrak{X}}\big)^* \qquad (\textit{taking matrix adjoint on the right}) \tag{51}$$

where A^* is the usual $m \times n$ "adjoint matrix," the conjugate-transpose of A such that $(A^*)_{ij} = \overline{A_{ji}}$ for $A \in \mathrm{M}(n \times m, \mathbb{K})$.

Proof: By definition, the entries of $[T]_{\mathfrak{Y}\mathfrak{X}}$ are determined by the vector identities

$$T(e_i) = \sum_{k=1}^{n} T_{ki}\, f_k \quad \text{which imply} \quad \big(T(e_i), f_j\big)_W = \sum_{k=1}^{n} T_{ki}\, (f_k, f_j)_W = T_{ji},$$

for all i, j. Hence

$$T^*(f_i) = \sum_{k=1}^{n} [T^*]_{ki} e_k \;\Rightarrow\; (T^*(f_i), e_j) = [T^*]_{ji}\ ,$$

from which we see that

$$
\begin{aligned}
[T^*]_{ij} &= \big(T^*(f_j), e_i\big)_V \\
&= (f_j, T(e_i))_W \;=\; \overline{(T(e_i), f_j)_W} \;=\; \overline{[T]_{ji}} = \Big([T]^*\Big)_{ij}
\end{aligned}
$$

where $(A^*)_{ij} = \overline{A_{ji}}$ for any matrix. $\square$

3.13. Exercise. Let V_N be the restrictions to $[0,1]$ of polynomials $f \in \mathbb{C}[x]$ having degree $\leq N$. Give this $(N+1)$-dimensional space of $\mathcal{C}[0,1]$ the usual L^2 inner product $(f,h)_2 = \int_0^1 f(t)\overline{h(t)}\,dt$ inherited from the larger space of continuous functions. Let $D : V_N \to V_N$ be the differentiation operator

$$
D\big(a_0 + a_1 t + a_2 t^2 + \ldots + a_N t^N\big) = a_1 + 2a_2\,t + 3a_3\,t^2 + \ldots + Na_n\,t^{N-1}
$$

(a) Is D one-to-one? Onto? What are range(D) and $\ker(D)$?

(b) Determine the matrix $[D]_{\mathfrak{X}\mathfrak{X}}$ with respect to the vector basis $\mathfrak{X} = \{1\!\!\!-, x, x^2, \ldots, x^N\}$.

(c) Determine the eigenvalues of $D : V_N \to V_N$ and their multiplicities.

(d) Compute the L^2-inner product $(f,h)_2$ in terms of the coefficients a_k, b_k that determine f and h.

(e) Is D a self-adjoint operator? Skew-adjoint?

3.14. Exercise. If D^* is the adjoint of the differentiation operator $D : V_N \to V_N$, entries D^*_{ij} in its matrix $[D^*]_{\mathfrak{X}}$ with respect to the basis $\mathfrak{X} = \{1\!\!\!-, x, x^2, \ldots, x^N\}$ are determined by the vector identities $D^*(x^i) = \sum_{k=0}^N D^*_{ki}\,x^k$. By definition of the adjoint D^* we have

$$
\big(x^i\,,\,D(x^j)\big)_2 = \big(D^*(x^i)\,,\,x^j\big)_2 \;=\; \sum_{k=0}^{N} D^*_{ik}\,\big(x^k\,,\,x^j\big)_2 \qquad \text{for } 0 \leq i,j \leq N
$$

and since $\mathfrak{X}$ is a basis these identities implicitly determine the D^*_{ij}. Compute explicit matrices B and C such that $[D^*]_{\mathfrak{X}} = C{\cdot}B^{-1}$. As in the preceding problem, $D(x^k) = kx^{k-1}$ and inner products in V_N are integrals

$$
\big(f\,,\,h\big)_2 = \int_0^1 f(x){\cdot}h(x)\,dx
$$

for polynomials $f, h \in V_N$.

Hint: BEWARE: The powers x^i are NOT an orthonormal basis, so you will have to use some algebraic brute force instead of (51). This could get complicated. For something more modest, just compute the action of D^* on the three-dimensional space $V = \mathbb{C}$-span$\{1\!\!\!-, t, t^2\}$.

3.15. Exercise. Let $V = \mathcal{C}^\infty_c(\mathbb{R})$ be the space of real-valued functions $f(t)$ on the real line that have continuous derivatives $D^k f$ of all orders, and have "bounded support" – each f is zero off of some bounded interval (which is allowed to vary with f). Because all such functions are "zero near ∞" there is a well defined inner product

$$
(f,h)_2 = \int_{-\infty}^{\infty} f(t)\overline{h(t)}\,dt
$$

The derivative $Df = df/dt$ is a linear operator on this infinite dimensional space.

(a) Prove that the adjoint of D is *skew-adjoint*, with $D^* = -D$.

(b) Prove that the second derivative $D^2 = d^2/dt^2$ is *self-adjoint*.

Hint: Integration by parts.

Normal and Self-Adjoint Operators. Various classes of operators $T : V \to V$ can be defined on an finite dimensional inner product space.

1. Self-adjoint: $T^* = T$
2. Skew-adjoint: $T^* = -T$
3. Unitary: $T^*T = I$ (which implies $TT^* = I$ because $T : V \to V$ is one-to-one $\Leftrightarrow$ onto $\Leftrightarrow$ bijective.) Thus "unitary" is equivalent to saying that $T^* = T^{-1}$, at least when V is finite dimensional. (In the infinite-dimensional case we need both identities $TT^* = T^*T = I$ to get $T^* = T^{-1}$.)
4. Normal: $T^*T = TT^*$ (T commutes with T^*)

The **spectrum** $\mathrm{sp}_{\mathbb{K}}(T) = \{\lambda \in \mathbb{K} : E_\lambda(T) \neq (0)\}$ of T is closely related to that of T^*.

3.16. Lemma. *On any inner product space*

$$\mathrm{sp}(T^*) = \overline{\mathrm{sp}(T)} = \{\,\overline{\lambda} : \lambda \in \mathrm{sp}(T)\}$$

Proof: If $(T - \lambda I)(v) = 0$ for some $v \neq 0$, then $0 = \det(T - \lambda I) = \det\big([T]_{\mathfrak{X}} - \lambda I_{n\times n}\big)$ for any basis $\mathfrak{X}$ in V. If $\mathfrak{X}$ is an orthonormal basis we get $[T^*]_{\mathfrak{X}} = [T]^*_{\mathfrak{X}} = \overline{[T]}^{t}_{\mathfrak{X}}$. Then

$$\begin{aligned} \det\big([T^*]_{\mathfrak{X}} - \overline{\lambda} I_{n\times n}\big) &= \overline{\det\big([T]^{\mathrm{t}}_{\mathfrak{X}} - \lambda I_{n\times n}\big)} = \overline{\det\big([T]_{\mathfrak{X}} - \lambda I_{n\times n}\big)^{\mathrm{t}}} \\ &= \overline{\det\big([T]_{\mathfrak{X}} - \lambda I_{n\times n}\big)} = 0 \end{aligned}$$

because

$$\det(A^{\mathrm{t}}) = \det(A) \qquad \text{and} \qquad \det(\overline{A}) = \overline{\det(A)}\ .$$

Hence $\overline{\lambda} \in \mathrm{sp}(T^*)$. Since $T^{**} = T$, we get

$$\mathrm{sp}(T) = \mathrm{sp}(T^{**}) \subseteq \overline{\mathrm{sp}(T^*)} \subseteq \overline{\overline{\mathrm{sp}(T)}} = \mathrm{sp}(T) \quad \square$$

3.17. Exercise. If $A \in \mathrm{M}(n, \mathbb{K})$ prove that its matrix adjoint $(A^*)_{ij} = \overline{A_{ji}}$ has determinant

$$\det(A^*) = \overline{\det(A)}.$$

If $T : V \to V$ is a linear map on an inner product space, prove that $\det(T^*) = \overline{\det(T)}$.

3.18. Exercise. *If $T : V \to V$ is a linear map on an inner product space, show that the characteristic polynomial satisfies*

$$p_{T^*}(\lambda) = \overline{p_T(\overline{\lambda})} \qquad \textit{or equivalently} \qquad p_T(\overline{\lambda}) = \overline{p_T(\lambda)}$$

for all $\lambda \in \mathbb{K}$. In particular,

$$\mathrm{sp}_{\mathbb{K}}(T^*) = \overline{\mathrm{sp}_{\mathbb{K}}(T)} = \{\,\overline{\lambda} : \lambda \in \mathrm{sp}_{\mathbb{K}}(T)\}.$$

Proof: Since $I^* = I$ and $(\lambda I)^* = \overline{\lambda} I$ we get

$$\begin{aligned} p_{T^*}(\lambda) &= \det(T^* - \lambda I) = \det\big(T^* - (\overline{\lambda} I)^*\big) \\ &= \det\big(\,(T - \overline{\lambda} I)^*\big) = \overline{\det(T - \overline{\lambda} I)} = \overline{p_T(\overline{\lambda})} \end{aligned}$$

Recall that $\mu \in \mathrm{sp}_{\mathbb{K}}(T) \Leftrightarrow p_T(\mu) = 0$. $\square$

VI.4. Diagonalization in Inner Product Spaces.

If M is a T-invariant subspace of inner product space V it does not follow that $T^*(M) \subseteq M$. The true relationship between invariance under T and under T^* is:

4.1. Exercise. If V is any inner product space and $T : V \to V$ a linear map, prove that

(a) A subspace $M \subseteq V$ is T-invariant (so $T(M) \subseteq M$) $\Rightarrow M^\perp$ is T^*-invariant.

(b) If $\dim_{\mathbb{K}}(V) < \infty$ (so $M^{\perp\perp} = M$) then $T(M) \subseteq M \Leftrightarrow T^*(M^\perp) \subseteq M^\perp$.

4.2. Proposition. *If $T : V \to W$ is a linear map between finite dimensional inner product spaces, let $R(T) = \mathrm{range}(T)$, $K(T) = \ker(T)$. Then $T^* : W \to V$ and*

$$\begin{array}{rcl} K(T^*) & = & R(T)^\perp \text{ in } W \\ R(T^*) & = & K(T)^\perp \text{ in } V \end{array}$$

In particular if T is self-adjoint then $\ker(T) \perp \mathrm{range}(T)$ *and we have an orthogonal direct sum decomposition $V = K(T) \dot{\oplus} R(T)$.*

Proof: If $w \in W$ then

$$\begin{array}{rcl} T^*(w) = 0 & \Leftrightarrow & \big(v, T^*(w)\big)_V = 0 \quad \text{for all } v \in V \\ & \Leftrightarrow & 0 = (v, T^*(w))_V = (T(v), w)_W, \quad \text{for all } v \in V \\ & \Leftrightarrow & w \perp R(T)\ . \end{array}$$

Hence $w \in K(T^*)$ if and only if $w \perp R(T)$. The second part follows because $T^{**} = T$ and $M^{\perp\perp} = M$ for any subspace. $\square$

We will often invoke this result.

Orthogonal Diagonalization. Not all linear operators $T : V \to V$ are diagonalizable, let alone orthogonally diagonalizable, but if V is an inner product space we can always find a basis that at least puts it into *upper-triangular* form, which can be helpful. In fact, this can be achieved via an othonormal basis provided the characteristic polynomial splits into linear factors over $\mathbb{K}$ (always true if $\mathbb{K} = \mathbb{C}$).

4.3. Theorem (Schur Normal Form). *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space over $K = \mathbb{R}$ or $\mathbb{C}$ such that $p_T(x) = \det(T - xI)$ splits over $\mathbb{K}$. Then there are scalars $\lambda_1, \ldots, \lambda_n$ and an orthonormal basis* $\mathfrak{X}$ in V such that

$$[T]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

Proof: Work by induction on $n = \dim_{\mathbb{K}}(V)$; the case $n = 1$ is trivial. For $n > 1$, since p_T splits there is an eigenvalue λ in $\mathbb{K}$ and a vector $v_0 \neq 0$ such that $T(v_0) = \lambda v_0$. Then $\overline{\lambda}$ is an eigenvalue for T^*, so there is some $w_0 \neq 0$ such that $T^*(w_0) = \overline{\lambda} w_0$.

Let $M = \mathbb{K}w_0$; this one-dimensional space is T^*-invariant, so $M^\perp$ is invariant under $(T^*)^* = T$ and has dimension $n - 1$. Scale w_0 if necessary to make $\|w_0\| = 1$. By the Induction Hypothesis there there is an orthonormal basis $\mathfrak{X}_0 = \{e_1, ..., e_{n-1}\}$ in $M^\perp$ such that

$$[T|M^\perp]_{\mathfrak{X}_0} = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_{n-1} \end{pmatrix}$$

Then letting $e_n = w_0$ (norm $= 1$) we get an orthonormal basis for V such that $[T]_{\mathfrak{X}\mathfrak{X}}$ has the form:

$$[T]_{\mathfrak{X}\mathfrak{X}} = \left(\begin{array}{ccc|c} \lambda_1 & & * & c_1 \\ & \ddots & & \vdots \\ 0 & & \lambda_{n-1} & c_{n-1} \\ \hline 0 & & 0 & \lambda_n \end{array} \right)$$

where

$$T(e_n) = T(w_0) = \lambda_n e_n + \sum_{j=1}^{n-1} c_j\, e_j$$

(Remember: $M = \mathbb{K}w_0$ need not be invariant under T.) $\square$

4.4. Exercise. Explain why the diagonal entries in the Schur normal form must be the roots in $\mathbb{K}$ of the characteristic polynomial $p_T(x) = \det(T - xI)$, each counted according to its algebraic multiplicity.

Note: Nevertheless, it might not be possible to find an orthonormal basis such that all occurrences of a particular eigenvalue $\lambda \in \mathrm{sp}_{\mathbb{K}}(T)$ appear in a consecutive string $\lambda, \ldots, \lambda$ on the diagonal. $\square$

Recall that a linear operator $T : V \to V$ on an inner product space is *normal* if it commutes with its adjoint, so that $T^*T = TT^*$. We will eventually show that when $\mathbb{K} = \mathbb{C}$ (or when $\mathbb{K} = \mathbb{R}$ and the characteristic polynomial of T splits into linear factors: $p_T(x) = \prod_{i=1}^{n}(x - \alpha_i)$ with $\alpha_i \in K$), then T is orthogonally diagonalizable *if and only if* T is normal. Note carefully what this does not say: T might be (non-orthogonally) diagonalizable over $\mathbb{K} = \mathbb{C}$ even if T is not normal. This latter issue can only be resolved by determining the pattern of eigenspaces $E_\lambda(T)$ and demonstrating that they span all of V.

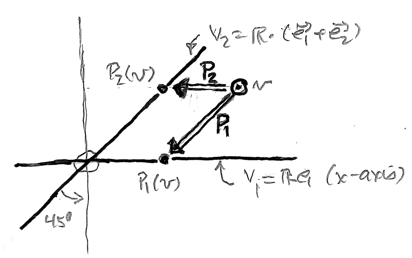

Figure 6.7. The (non-orthogonal) basis vectors $\mathbf{u}_1 = \mathbf{e}_1$ and $\mathbf{u}_2 = \mathbf{e}_1 + \mathbf{e}_2$ in Exercise 4.5.

4.5. Exercise. Let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be the standard orthonormal basis vectors in $V = \mathbb{K}^2$, and consider the ordinary direct sum decomposition

$$V = V_1 \oplus V_2 = \mathbb{K}\mathbf{e}_1 \oplus \mathbb{K}(\mathbf{e}_1 + \mathbf{e}_2) = \mathbb{K}\,\mathbf{f}_1 \oplus \mathbb{K}\,\mathbf{f}_2 \qquad \text{where} \qquad \mathbf{f}_1 = \mathbf{e}_1, \mathbf{f}_2 = \mathbf{e}_1 + \mathbf{e}_2 \ .$$

These subspaces are not orthogonal with respect to the standard Euclidean inner product

$$\big(x_1\mathbf{e}_1 + x_2\mathbf{e}_1\, ,\, y_1\mathbf{e}_1 + y_2\mathbf{e}_2\big) \;=\; x_1\overline{y_1} + x_2\overline{y_2}$$

Define a $\mathbb{K}$-linear map $T : V \to V$, letting

$$T(\mathbf{e}_1) = 2\mathbf{e}_1 \qquad T(\mathbf{e}_1 + \mathbf{e}_2) = \tfrac{1}{2}(\mathbf{e}_1 + \mathbf{e}_2)$$

(see Figure 6.7). Then T is diagonalized by the basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ with $\mathbf{f}_1 = \mathbf{e}_1$ and $\mathbf{f}_2 = \mathbf{e}_1 + \mathbf{e}_2$ (which is obviously not orthonormal), with

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

(a) Determine the action of T on the orthonormal basis vectors $\mathfrak{X} = \{\mathbf{e}_1, \mathbf{e}_2\}$ and find $[T]_{\mathfrak{X}\mathfrak{X}}$;

(b) Describe the operator T^* by determining its action on the standard orthonormal basis $\mathfrak{X}$, and find $[T^*]_{\mathfrak{X}\mathfrak{X}}$;

(c) Explain why T is not a normal operator on V. Explain why no orthonormal basis $\{\mathbf{f}_1, \mathbf{f}_2\}$ in V can possibly diagonalize T.

Hint: The discussion is exactly the same for $\mathbb{K} = \mathbb{R}$ and $\mathbb{C}$, so assume $\mathbb{K} = \mathbb{R}$ if that makes you more comfortable.

Diagonalizing Self-Adjoint and Normal Operators. We now show that a linear operator $T : V \to V$ on a finite dimensional inner product space is orthogonally diagonalizable if and only if T is normal. First, we analyze the special case of self-adjoint operators ($T^* = T$), which motivates the more subtle proof needed for normal operators.

4.6. Theorem (Diagonalizing Self-Adjoint T). *On a finite dimensional inner product space any self-adjoint linear operator $T : V \to V$ is orthogonally diagonalizable.*

Proof: If $\mu, \lambda \in \mathrm{sp}_{\mathbb{K}}(T)$, we first observe that:

1. If $T = T^*$ all eigenvalues are real, so $\mathrm{sp}_{\mathbb{K}}(T) \subseteq \mathbb{R} + i0$.

   Proof: If $v \in E_\lambda(T)$, $v \neq 0$, we have

   $$\lambda \|v\|^2 = (Tv, v) = (v, T^*v) = (v, Tv) = (v, \lambda v) = \overline{\lambda}\, \|v^2\|^2$$

   which implies $\lambda = \overline{\lambda}$.

2. If $\lambda \neq \mu$ in $\mathrm{sp}(T)$ the eigenspaces $E_\lambda(T)$ and $E_\mu(T)$ must be orthogonal.

   Proof: If $v \in E_\lambda(T)$, $w \in E_\mu(T)$ then

   $$\lambda(v, w) = (Tv, w) = (v, T^*w) = (v, \mu w) = \overline{\mu}\,(v, w) = \mu\,(v, w)$$

   since eigenvalues are real when $T^* = T$. But $\mu \neq \lambda$, hence $(v, w) = 0$ and $E_\lambda(T) \perp E_\mu(T)$. Thus the linear span $E = \sum E_\lambda(T)$ (which is always a direct sum) is actually an orthogonal sum $E = \dot{\bigoplus}_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$.

3. If $T^* = T$ the span of the eigenspaces is all of V, hence T is orthogonally diagonalizable.

   Proof: If $\lambda \in \mathrm{sp}_{\mathbb{K}}(T)$, then $E_\lambda(T) \neq (0)$ and $M = E_\lambda(T)^\perp$ has $\dim(M) < \dim(V)$. By Exercise 4.1 the orthogonal complement is T^*-invariant, hence T-invariant because $T^* = T$. It is easy (see Exercise 4.7 below) to check that if $W \subseteq V$ is T-invariant and $T^* = T$ on V, then the restriction $T|_W : W \to W$ is self-adjoint on W if one equips W with the restricted inner product from V.

4.7. Exercise. If $T : V \to V$ is linear and $T^* = T$, prove that

$$(T|_W)^* = (T^*|_W)$$

for any T-invariant subspace $W \subseteq V$ equipped with the restricted inner product.

To complete our discussion we show that self-adjoint operators are orthogonally diagonalizable, arguing by induction on $n = \dim(V)$. This is clear if $\dim(V) = 1$, so assume it true whenever $\dim(V) \leq n$ and consider a space of dimension $n+1$. Since all eigenvalues (roots of the characteristic polynomial) are real there is a nontrivial eigenspace $M = E_\lambda(T)$, and if this is all of V we're done: $T = \lambda I$. Otherwise, M has lower dimension and by Exercise 4.7 it has an orthonormal basis that diagonalizes $T|_M$. But $V = M \dot{\oplus} M^\perp$ (an orthogonal direct sum), and $M = E_\lambda$ obviously has an orthonormal basis of eigenvectors. Combining these bases we get an orthonormal diagonalizing basis for all of V. $\square$

We now elaborate the basic properties of normal operators on an inner product space.

4.8. Proposition. *A normal linear operator $T : V \to V$ on a finite dimensional inner product space has the following properties.*

1. *If $T : V \to V$ is normal, $\|T(v)\| = \|T^*(v)\|$ for all $v \in V$.*
   Proof: *We have*
   $$\begin{aligned} \|T(v)\|^2 &= (Tv, Tv) \;=\; (T^*T(v), v) \;=\; (TT^*(v), v) \\ &= (T^*v, T^*v) \;=\; \|T^*(v)\|^2 \end{aligned}$$

2. *For any $c \in \mathbb{K}$, $T - cI$ is also normal because $(T - cI)^* = T^* - \overline{c}I$ and cI commutes with all operators.*

3. *If $T(v) = \lambda v$ then for the same vector v we have $T^*(v) = \overline{\lambda}v$. In particular, $E_{\overline{\lambda}}(T^*) = E_\lambda(T)$. (This is a much stronger statement than our earlier observation that $\mathrm{sp}_{\mathbb{K}}(T^*) = \overline{\mathrm{sp}_{\mathbb{K}}(T)} = \{\,\overline{\lambda} : \lambda \in \mathrm{sp}_{\mathbb{K}}(T)\}\,$).*
   Proof: *$(T - \lambda)$ is also normal. Therefore if $v \in V$ and $T(v) = \lambda v$, we have*
   $$T(v) = \lambda v \;\Rightarrow\; \big((T-\lambda)^*(T-\lambda)\,v\,,\,v\big) = \|(T-\lambda)\,v\|^2 = 0$$
   which implies that
   $$0 = \big((T-\lambda)(T-\lambda)^*\,v\,,\,v\big) = \|(T^* - \overline{\lambda}I)\,v\|^2 \;\Rightarrow\; T^*(v) = \overline{\lambda}v$$

4. *If $\lambda \neq \mu$ in $\mathrm{sp}_{\mathbb{K}}(T)$, then $E_\lambda \perp E_\mu$.*
   Proof: *If v, w are in E_λ, E_μ then*
   $$\lambda(v, w) = (\lambda v, w) = (Tv, w) = (v, T^*w) = \big(v, \overline{\mu}\,w\big) = \mu(v, w)$$
   since $T^*(w) = \overline{\mu}\,w$ if $T(w) = \mu\,w$. Therefore $(v, w) = 0$ if $\mu \neq \lambda$.

If $M = \sum_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$ for a normal operator T, it follows that this is a direct sum of orthogonal subspaces $M = \dot{\bigoplus}_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$, and that there is an orthonormal basis $\{e_1, \ldots, e_n\} \subseteq M$ consisting of eigenvectors.

4.9. Corollary. *If $T : V \to V$ is normal and $\mathbb{K} = \mathbb{C}$ (or if $\mathbb{K} = \mathbb{R}$ and the characteristic polynomial p_T splits over $\mathbb{R}$), there is a diagonalizing orthonormal basis $\{e_i\}$ and V is an orthogonal direct sum $\dot{\bigoplus}_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$.*

Proof: The characteristic polynomial $p_T(x) = \det(T - xI)$ splits in $\mathbb{C}[x]$, so there is an eigenvalue λ_0 such that $T(v_0) = \lambda_0 v_0$ for some $v_0 \neq 0$. The one-dimensional space $M = \mathbb{C}v_0$ is T-invariant, but is also T^*-invariant since $T^*(v_0) = \overline{\lambda_0}v_0$ by (3.). Then

$$T^*(M) \subseteq M \;\Rightarrow\; T^{**}(M^\perp) \;=\; T(M^\perp) \subseteq M^\perp\ .$$

We also have $T^*(M^\perp) \subseteq M^\perp$ because $T(M) \subseteq M \Leftrightarrow T^*(M^\perp) \subseteq M^\perp$. $\square$

4.10. Exercise. If N is a subspace in an inner product space that is invariant under both T and T^*, prove that $T|_N$ satisfies

$$(T|_N)^* = (T^*|_N)$$

Note: Here we do not assume $T^* = T$, which was assumed in Exercise 4.7.

Since $T|_{M^\perp}$ is again a normal operator with respect to the inner product $M^\perp$ inherits from the larger space V, but $\dim(M^\perp) < \dim(V)$, we may argue by induction to get an orthonormal basis of eigenvectors. $\square$

4.11. Theorem (Orthogonal Diagonalization). *Let $T : V \to V$ be a linear operator on a finite dimensional inner product space. Assume that the characteristic polynomial $p_T(x)$ splits over $\mathbb{K}$ (certainly true for $\mathbb{K} = \mathbb{C}$). There is an orthonormal basis that diagonalizes T if and only if T is normal: $T^*T = TT^*$*

Note: It follows that $V = \bigoplus_{\lambda \in \mathrm{sp}_{\mathbb{K}}(T)} E_\lambda(T)$; in particular, the eigenspaces are mutually orthogonal. Once the eigenspaces are determined it is easy to construct the diagonalizing orthonormal basis for T.

Proof: ($\Rightarrow$) has just been done.

Proof: ($\Leftarrow$). If there is an orthonormal basis $\mathfrak{X} = \{e_i\}$ that diagonalizes T then

$$[T]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

But $[T^*]_{\mathfrak{X}\mathfrak{X}}$ is the adjoint of the matrix $[T]_{\mathfrak{X}\mathfrak{X}}$,

$$[T^*]_{\mathfrak{X}\mathfrak{X}} = \overline{[T]^{\mathrm{t}}_{\mathfrak{X}\mathfrak{X}}} = \begin{pmatrix} \overline{\lambda_1} & & & 0 \\ & \overline{\lambda_2} & & \\ & & \ddots & \\ 0 & & & \overline{\lambda_n} \end{pmatrix}$$

Obviously these diagonal matrices commute (all diagonal matrices do), so

$$[T^*T]_{\mathfrak{X}\mathfrak{X}} = [T^*]_{\mathfrak{X}\mathfrak{X}}[T]_{\mathfrak{X}\mathfrak{X}} = [T]_{\mathfrak{X}\mathfrak{X}}[T^*]_{\mathfrak{X}\mathfrak{X}} = [TT^*]_{\mathfrak{X}\mathfrak{X}}$$

which implies $T^*T = TT^*$ as operators on V. $\square$

4.12. Example. Let $L_A : \mathbb{C}^2 \to \mathbb{C}^2$ be the multiplication operator determined by

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}$$

so that $L_A(\mathbf{e}_1) = \mathbf{e}_1$ and $L_A(\mathbf{e}_1 + \mathbf{e}_2) = 2 \cdot (\mathbf{e}_1 + \mathbf{e}_2)$, where $\mathfrak{X} = \{\mathbf{e}_1, \mathbf{e}_2\}$ is the standard orthonormal basis. As we saw in Chapter 2, $[L_A]_{\mathfrak{X}\mathfrak{X}} = A$. But L_A is obviously diagonalizable with respect to the *non-orthonormal* basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ where $\mathbf{f}_1 = \mathbf{e}_1$, $\mathbf{f}_2 = \mathbf{e}_1 + \mathbf{e}_2$. The $\mathbf{f}_i$ are basis vectors for the (one-dimensional) eigenspaces of L_A, which are uniquely determined without any reference to the inner product in $V = \mathbb{C}^2$; if there were an orthonormal basis that diagonalized L_A the eigenspaces would be orthogonal. which they are not. This operator cannot be orthogonally diagonalized with respect to the standard inner product in $\mathbb{C}^2$. $\square$

4.13. Exercise. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be L_A for the matrix

$$A = A^* = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

in M(2.$\mathbb{C}$). Determine the eigenvalues in $\mathbb{C}$ and the eigenspaces, and exhibit an orthonormal basis $\mathfrak{Y} = \{\mathbf{f}_1, \mathbf{f}_2\}$ that diagonalizes T.

4.14. Exercise. Prove that $|\lambda| = 1$ for all eigenvalues $\lambda \in \mathrm{sp}(T)$ of a unitary operator (so λ lies on the unit circle if $\mathbb{K} = \mathbb{C}$, or $\lambda = \pm 1$ if $\mathbb{K} = \mathbb{R}$).

4.14A. Exercise. If P is a projection on a finite dimensional vector space (so $P^2 = P$),

(a) Explain why P is diagonalizable, over any field $\mathbb{K}$. What are the eigenvalues and eigenspaces?

(b) Give an explicit example of a projection operator on a finite dimensional inner product space that is not *orthogonally diagonalizable*.

4.14B. Exercise. If P is a projection operator (so $P^2 = P$) on a finite dimensional inner product space, prove that P is a normal operator $\Leftrightarrow$ $K(P) = \ker(P)$ and $R(P) = \mathrm{range}(P)$ are orthogonal subspaces.
Note: ($\Rightarrow$) is trivial since $K(P) = E_{\lambda=0}(P)$ and $R(P) = E_{\lambda=1}(P)$.

4.14C. Exercise. A projection operator P (with $P^2 = P$) on an inner product space is fully determined once we know its kernel $K(P)$ and range $R(P)$, since $V = R(P) \oplus K(P)$. The adjoint P^* is also a projection operator because $(P^*P^*) = (PP)^* = P^*$.

(a) In an inner product space, how are $K(P)$ and $R(P)$ related to $K(P^*)$ and $R(P^*)$?

(b) For the non-orthogonal direct sum decomposition of Exercise VI-4.5 give explicit descriptions of the subspaces $K(P^*)$ and $R(P^*)$. (Find bases for each.)

If $T : V \to V$ is an arbitrary linear operator on an inner product space we showed in IV.3.16 that $\mathrm{sp}(T^*)$ is equal to $\overline{\mathrm{sp}(T)}$; in VI-3.48 we showed that

$$E_{\overline{\lambda}}(T^*) = E_\lambda(T) \qquad (\lambda \in \mathrm{sp}(T))$$

for *normal* operators. Unfortunately the latter property is not true in general.

4.14D. Exercise. If $T : V \to V$ is a linear operator on an inner product space and $\lambda \in \mathrm{sp}(T)$, prove that

(a) $E_{\overline{\lambda}}(T^*) = K\big(T^* - \overline{\lambda} I\big)$ is equal to $R(T - \lambda I)^\perp$.

(b) $\dim E_{\bar{\lambda}}(T^*) = \dim E_\lambda(T)$.

(c) T diagonalizable $\Rightarrow$ T^* is diagonalizable.

As the next example shows, $E_{\overline{\lambda}}(T^*) = K\big(T^* - \overline{\lambda} I\big)$ is not always equal to $E_\lambda(T)$ unless T is normal.

4.14E. Exercise. If $P : V \to V$ is an idempotent operator on a finite dimensional vector space (so $P^2 = P$), explain why P must be diagonalizable over any field. If $P \neq 0$ and $P \neq I$, what are its eigenvalues and its eigenspaces.

4.14F. Exercise. Let P be the projection operator on an inner product space V corresponding to a *non-orthogonal* direct sum decomposition $V = R(P) \oplus K(P)$. Its adjoint P^* is also a projection, onto $R(P^*)$ along $K(P^*)$.

(a) What are the eigenvalues and eigenspaces for P and P^*?

(b) For $\lambda = 1$, is $E_{\overline{\lambda}}(T^*) = K\big(T^* - \overline{\lambda}I\big)$ is equal to $E_\lambda(T)$?

Hint: See Exercise VI-4.14C and D.

Unitary Equivalence of Operators. We say that two operators T, T' on a vector space V are **similar**, written as $T' \sim T$, if there is an invertible linear operator S such that $T' = SAS^{-1}$; this means they are represented by the same matrix $[T']_{\mathfrak{Y}\mathfrak{Y}} = [T]_{\mathfrak{X}\mathfrak{X}}$ with respect to suitably chosen bases in V. We say T' is **unitarily equivalent to** T if there is a unitary operator U such that $T' = UTU^*(= UTU^{-1})$. This relation, denoted $T' \cong T$, is an RST equivalence relation between operators on an inner product space, but is more stringent than mere similarity. We now show $T' \cong T$ if and only if there are *orthonormal* bases $\mathfrak{X}, \mathfrak{Y}$ such that $[T']_{\mathfrak{Y}\mathfrak{Y}} = [T]_{\mathfrak{X}\mathfrak{X}}$.

4.15. Definition. *A* **linear isometry** *is a linear operator $U : V \rightarrow W$ between inner product spaces that preserve distances in mapping points from V into W,*

$$\|Uv - Uv'\|_W = \|U(v - v')\|_W = \|v - v'\|_V \ ; \tag{52}$$

in particular $\|U(v)\|_W = \|v\|_V$ for all $v \in V$. Isometries are one-to-one but need not be bijections unless $\dim V = \dim W$ *(see exercises below).*

A linear map $U : V \rightarrow W$ is **unitary** *if $U^*U = \mathrm{id}_V$ and $UU^* = \mathrm{id}_W$, which means U is invertible with $U^{-1} = U^*$ (hence* $\dim V = \dim W$*). Obviously the inverse map $U^{-1} : W \rightarrow V$ is also unitary. Unitary operators $U : V \rightarrow W$ are also isometries since*

$$\|Ux\|_W^2 = \big(Ux, Ux\big)_W = (x, U^*Ux)_V = \|x\|_V^2 \ ,$$

Thus unitary maps are precisely the bijective linear isometries from V to W.

If V is finite dimensional and we restrict attention to the case $V = W$, either of the conditions $UU^* = \mathrm{id}_V$ or $U^*U = \mathrm{id}_V$ implies U is invertible with $U^{-1} = U^*$ because

$$U \text{ one-to-one} \quad \Leftrightarrow \quad U \text{ is surjective} \quad \Leftrightarrow \quad U \text{ is bijective,}$$

for any linear operator $U : V \rightarrow V$ when $\dim(V) < \infty$.

4.16. Exercise. If V, W are inner product spaces of the same *finite* dimension, explain why there must exist a bijective linear isometry $T : V \rightarrow W$. Is T unique? Is the adjoint $T^* : W \rightarrow V$ also an isometry?

4.17. Exercise. Let $V = \mathbb{C}^m, W = \mathbb{C}^n$ with the usual inner products. Exhibit examples of linear operators $U : V \rightarrow W$ such that

(a) $UU^* = \mathrm{id}_W$ but $U^*U \neq \mathrm{id}_V$.
(b) $U^*U = \mathrm{id}_V$ but $UU^* \neq \mathrm{id}_W$.

Note: This might not be possible for all choices of m, n (for instance $m = n$).

4.18. Exercise. If $m < n$ and the coordinate spaces $\mathbb{K}^m$, $\mathbb{K}^n$ are equipped with the standard inner products, consider the linear operator

$$T : \mathbb{K}^m \rightarrow \mathbb{K}^n \qquad T(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, , 0)$$

This is an isometry from $\mathbb{K}^m$ into $\mathbb{K}^n$, with trivial kernel $K(T) = (0)$ and range $R(T) = \mathbb{K}^m \times (\mathbf{0})$ in $\mathbb{K}^n = \mathbb{K}^m \oplus \mathbb{K}^{n-m}$.

(a) Provide an explicit description of the adjoint operator $T^* : \mathbb{K}^n \rightarrow \mathbb{K}^m$ and determine $K(T^*), R(T^*)$.

(b) Compute the matrices of $[T]$ and $[T^*]$ with respect to the standard orthonormal bases in $\mathbb{K}^m, \mathbb{K}^n$.

(c) How is the action of T^* related to the subspaces $K(T), R(T^*)$ in $\mathbb{K}^m$ and $R(T), K(T^*)$ in $\mathbb{K}^n$? Can you give a geometric description of this action?

Unitary operators can be described in several different ways, each with its own advantages in applications.

4.19. Theorem. *The statements below are equivalent for a linear operator $U : V \to W$ between finite dimensional inner product spaces.*

(a) $UU^* = \mathrm{id}_W$ *and* $U^*U = \mathrm{id}_V$ (*so* $U^* = U^{-1}$ *and* $\dim V = \dim W$).

(b) *U maps* SOME *orthonormal basis $\{e_i\}$ in V to an orthonormal basis $\{f_i = U(e_i)\}$ in W.*

(c) *U maps* EVERY *orthonormal basis $\{e_i\}$ in V to an orthonormal basis $\{f_i = U(e_i)\}$ in W.*

(d) *U is a surjective isometry, so distances are preserved:*

$$\|U(x) - U(y)\|_W = \|x - y\|_V \qquad \text{for } x, y \in V$$

(*Then U is invertible and U^{-1} is also an isometry*).

(e) *U is a bijective map that preserves inner products, so that*

$$\big(U(x), U(y)\big)_W = (x, y)_V \qquad \text{for all } x, y \in V.$$

$$\begin{tikzcd} & (b) \arrow[d, Rightarrow] & (c) \arrow[l, Rightarrow] \\ (a) \arrow[r, Leftrightarrow] & (e) \arrow[ur, Rightarrow] \arrow[r, Leftrightarrow] & (d) \end{tikzcd}$$

Figure 6.8. The pattern of implications in proving Theorem 4.19.

Proof: We prove the implications shown in Figure 6.8.

Proof: (d) $\Leftrightarrow$ (e). Clearly (e) $\Rightarrow$ (d). For the converse, (d) implies U preserves lengths of vectors, with $VertUx\|_W = \|x\|_V$ for all x. Then by the Polarization Identity for inner products

$$(x, y) = \frac{1}{4}\sum_{k=0}^{3} \frac{1}{i^k}\,\|x + i^k y\|^2\ ,$$

so inner products are preserved, proving (d) $\Rightarrow$ (e) when $\mathbb{K} = \mathbb{C}$; same argument but with only 2 terms if $\mathbb{K} = \mathbb{R}$.

Proof: (e) $\Rightarrow$ (c) $\Rightarrow$ (b). These are obvious since "orthonormal basis" is defined in terms of the inner product. For instance if (e) holds and $\mathfrak{X} = \{e_i\}$ is an orthonormal basis in V then $\mathfrak{Y} = \{f_i = U(e_j)\}$ is an orthonormal family in W because

$$(f_i, f_j)_W = \big(U(e_i), U(e_j)\big)_W = (e_i, U^*Ue_j)_V = (e_i, e_j)_V = \delta_{ij} \quad \text{(Kronecker delta).}$$

But

$$\mathbb{K}\text{-span}\{f_j = U(e_j)\} = U\big(\mathbb{K}\text{-span}\{e_j\}\big) = U(V) = W \ ,$$

so $\mathfrak{Y}$ spans W and therefore is a basis.

Proof (a) $\Leftrightarrow$ (e). We have

$$U^*U = \mathrm{id}_V \Leftrightarrow U^*Ux = x \ \text{ for all } x \Leftrightarrow (Ux, Uy)_W = (x, U^*Uy)_V = (x,y)_V$$

for all $x, y \in V$.

Proof: (b) $\Rightarrow$ (e). Given an orthonormal basis $\mathfrak{X} = \{e_i\}$ in V such that the vectors $\mathfrak{Y} = \{f_i = U(e_i)\}$ are an orthonormal basis in W, we may write $x, y \in V$ as $x = \sum_i (x, e_i)\, e_i$, etc. Then

$$U(x) = \sum_i (x, e_i)_V\, U(e_i) = \sum_i (x, e_i)_V\, f_i, \qquad \text{etc} \ ,$$

hence by orthonormality

$$\begin{aligned}(Ux, Uy)_W &= \Big(\sum_i (x,e_i)_V\, f_i \,,\, \sum_j (y,e_j)_V\, f_j\Big)_W \ = \sum_{i,j} (x,e_i)_V\, \overline{(y,e_j)_V}\, (f_i, f_j)_W \\ &= \sum_k (x, e_k)_V\, (e_k, y)_V \ = \ (x,y)_V \quad \square\end{aligned}$$

Here we applied a formula worth remembering (*Parseval's identity*).

4.20. Lemma (Parseval). *If $x = \sum_i a_i e_i$, $y = \sum b_j e_j$ with respect to an orthonormal basis in a finite dimensional inner product space then $(x,y) = \sum_{k=1}^n a_k \overline{b_k}$. Equivalently, since $a_i = (x, e_i)$, ... etc, we have*

$$(x,y) = \sum_{k=1}^n (x, e_k)(e_k, y) \qquad \textit{for all } x, y$$

in any finite dimensional inner product space, since $\overline{(y, e_k)} = (e_k, y)$. $\square$

Unitary Operators vs Unitary Matrices.

4.21. Definition. *A matrix $A \in \mathrm{M}(n, \mathbb{K})$ is* **unitary** *if $AA^* = I$ (which holds $\Leftrightarrow AA^* = I \Leftrightarrow A^* = A^{-1}$), where A^* is the* **adjoint matrix** *such that $(A^*)_{ij} = \overline{A_{ji}}$. The set of all unitary matrices is a group since products and inverses of such matrices are again unitary. When $\mathbb{K} = \mathbb{C}$ this is the* **unitary group**

$$\mathrm{U}(n) = \{A \in \mathrm{M}(n, \mathbb{C}) : A^*A = I\} = \{A \in \mathrm{M}(n, \mathbb{C}) : A^* = A^{-1}\} \ .$$

But when $\mathbb{K} = \mathbb{R}$ and $A^* = A^{\mathrm{t}}$ (the transpose matrix), it goes by another name and is called the **orthogonal group**,

$$\mathrm{O}(n) = \{A \in \mathrm{M}(n, \mathbb{R}) : A^{\mathrm{t}}A = I\} = \{A \in \mathrm{M}(n, \mathbb{R}) : A^{\mathrm{t}} = A^{-1}\}$$

Both groups lie within the **general linear group** *of nonsingular matrices $\mathrm{GL}(n, \mathbb{K}) = \{A : \det(A) \neq 0\}$, and both contain noteworthy subgroups*

$$\begin{aligned}\text{Special Unitary Group:} \ \ \mathrm{SU}(n) &= \{A : A^*A = I \textit{ and } \det(A) = +1\} \\ \text{Special Orthogonal Group:} \ \ \mathrm{SO}(n) &= \{A : A^{\mathrm{t}}A = I \textit{ and } \det(A) = +1\}\end{aligned}$$

The group SU(3), for instance, seems to be the symmetry group that governs the relations between electromagnetic forces and the weak and strong forces of nuclear physics. As we will see in the next section, SO(3) is the group of rotations in Euclidean space $\mathbb{R}^3$, by any angle about any oriented line through the origin (with a similar interpretation for $\mathrm{SO}(n)$ in higher dimensional spaces $\mathbb{R}^n$).

Given a matrix $A \in \mathrm{M}(n,\mathbb{K})$ it is important to know when the operator $L_A : \mathbb{K}^n \rightarrow \mathbb{K}^n$ is unitary with respect to the standard inner product. The answer extends the list of condititions (a) – (e) of Theorem VI-4.19 describing when an operator is unitary, and is quite useful in calculations.

4.22. Proposition. *If $A \in \mathrm{M}(n,\mathbb{K})$ the following conditions are equivalent.*

1. *$L_A : \mathbb{K}^n \rightarrow \mathbb{K}^n$ is unitary;*
2. *A is a unitary matrix, so $A^*A = AA^* = I$ in $\mathrm{M}(n,\mathbb{K})$*
3. *The rows in A form an orthonormal basis in $\mathbb{K}^n$.*
4. *The columns in A form an orthonormal basis in $\mathbb{K}^n$.*

Proof: With respect to the standard basis $\mathfrak{X} = \{e_1, \ldots, e_n\}$ in $\mathbb{K}^n$ we know that $[L_A]_{\mathfrak{X}} = A$, but since $\mathfrak{X}$ is an orthonormal basis we also have $[(L_A)^*]_{\mathfrak{X}} = [L_A]^*_{\mathfrak{X}} = A^*$ (the adjoint matrix), by Exercise 3.12. Next observe that

$$L_{A^*} = \big(L_A\big)^* \qquad \text{as operators on } \mathbb{K}^n$$

(This may sound obvious, but it actually needs to be proved keeping in mind how the various "adjoints" are defined – see Exercise 4.24 below.) Then we get

$$\begin{aligned} A^*A = I \quad &\Leftrightarrow \quad \mathrm{id}_{\mathbb{K}^n} = [L_{A^*A}]_{\mathfrak{X}} = [L_{A^*}]_{\mathfrak{X}} \cdot [L_A]_{\mathfrak{X}} = [(L_A)^*]_{\mathfrak{X}} \cdot [L_A]_{\mathfrak{X}} \\ &\Leftrightarrow \quad (L_A)^* L_A = \mathrm{id}_{\mathbb{K}^n} \;\Leftrightarrow\; (L_A \text{ is a unitary operator}) \quad , \end{aligned}$$

proving (1.) $\Leftrightarrow$ (2.)

By definition of row-column matrix multiplication we have

$$\delta_{ij} = (AA^*)_{ij} = \sum_k A_{ik}(A^*)_{kj} = \sum_k A_{ik}\overline{A_{jk}} = \Big(\mathrm{Row}_i(A)\,,\,\mathrm{Row}_j(A)\Big)_{\mathbb{K}^n}$$

This says precisely that the rows are an orthonormal basis with respect to the standard inner product in $\mathbb{K}^n$. Thus (2.) $\Leftrightarrow$ (3.), and similarly $A^*A = I \Leftrightarrow$ the columns form an orthonormal basis in $\mathbb{K}^n$. $\square$

A similar criterion allows us to decide when a general linear operator is unitary.

4.23. Proposition. *A linear operator $T : V \rightarrow V$ on a finite dimensional inner product space is unitary $\Leftrightarrow$ its matrix $A = [T]_{\mathfrak{X}}$ with respect to any orthonormal basis is a unitary matrix (so $AA^* = A^*A = I$).*

Proof: For any orthonormal basis we have

$$I = [\mathrm{id}_V]_{\mathfrak{X}} = [T^*T]_{\mathfrak{X}} = [T^*]_{\mathfrak{X}}\,[T]_{\mathfrak{X}} = \big([T]_{\mathfrak{X}}\big)^*\,[T]_{\mathfrak{X}} = A^*A$$

and similarly $AA^* = I$, so A is a unitary matrix.

Conversely, if $A = [T]_{\mathfrak{X}}$ is a unitary matrix we have

$$\begin{aligned} (Te_i, Te_j) \;&=\; \Big(\sum_k A_{ki}\,e_k\,,\,\sum_\ell A_{\ell j}\,e_j\Big) \;=\; \sum_{k,\ell} A_{ki}\overline{A_{\ell j}}\,\delta_{k\ell} \\ &=\; \sum_k A_{ki}(A^*)_{jk} \;=\; (AA^*)_{ji} \;=\; \delta_{ji} = (e_i, e_j) \end{aligned}$$

Thus T maps orthonormal basis $\mathfrak{X}$ to a new orthonormal basis $\mathfrak{Y} = \{T(e_i)\}$, and T is unitary by Theorem 4.19(c). $\square$

4.24. Exercise. Prove that $L_{A^*} = \big(L_A\big)^*$ when $\mathbb{K}^n$ is given the standard inner product. **Hint:** Show that $(A^*\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y})$ for the standard inner product.

This remains true when A in an $n \times m$ matrix, $L_A : \mathbb{K}^m \to \mathbb{K}^n$, and $(L_A)^* : \mathbb{K}^n \to \mathbb{K}^m$.

4.24A. Exercise. If $A \in \mathrm{M}(n, \mathbb{C})$ give a careful proof that $A^*A = I \Leftrightarrow AA^* = I$.

4.25. Exercise. Given two orthonormal bases $\{e_i\}$, $\{f_j\}$ in finite dimensional inner product spaces V, W of the same dimension, construct a unitary operator $U : V \to W$ such that $U(e_i) = f_i$ for all i.

Change of Orthonormal Basis. If $T : V \to V$ is a linear operator on a finite dimensional inner product space, and we know its matrix $[T]_{\mathfrak{X}\mathfrak{X}}$ with respect to one orthonormal basis, what is its matrix realization with respect to a different orthonormal basis $\mathfrak{Y}$?

4.26. Definition. *Matrices $A, B \in \mathrm{M}(n, \mathbb{K})$ are* **unitarily equivalent**, *indicated by writing $A \cong B$, if there is some unitary matrix $S \in \mathrm{M}(n, \mathbb{K})$ such that $B = SAS^* = SAS^{-1}$.* $\square$

4.27. Theorem (Change of Orthonormal Basis). *If $\mathfrak{X} = \{e_i\}$ and $\mathfrak{Y} = \{f_j\}$ are orthonormal bases in a finite dimensional inner product space and $T : V \to V$ is any linear operator, the corresponding matrices $A = [T]_{\mathfrak{X}\mathfrak{X}}$ and $B = [T]_{\mathfrak{Y}\mathfrak{Y}}$ are unitarily equivalent: there is some unitary matrix S such that*

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = S\,[T]_{\mathfrak{X}\mathfrak{X}}S^* = S\,[T]_{\mathfrak{X}\mathfrak{X}}S^{-1} \qquad \textit{where } S = [\mathrm{id}_V]_{\mathfrak{Y}\mathfrak{X}} = [\mathrm{id}_V]^{-1}_{\mathfrak{X}\mathfrak{Y}} = [\mathrm{id}_V]^*_{\mathfrak{X}\mathfrak{Y}} \tag{53}$$

The identity (53) *remains true if the transition matrix S is multiplied by any scalar such that $|\lambda|^2 = \lambda\overline{\lambda} = 1$.*

Proof: For *arbitrary* vector bases $\mathfrak{X}, \mathfrak{Y}$ in V we have $[\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = [\mathrm{id}]^{-1}_{\mathfrak{Y}\mathfrak{X}}$ and

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}} \cdot [T]_{\mathfrak{X}\mathfrak{X}} \cdot [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}} = S\,[T]_{\mathfrak{X}\mathfrak{X}}S^{-1} \tag{54}$$

where $S = [\mathrm{id}]_{\mathfrak{Y}\mathfrak{X}}$ is given by the vector identities $e_i = \mathrm{id}(e_i) = \sum_j S_{ji} f_j$. But we also have $e_i = \sum_j (e_i, f_j)\, f_j$, so $S_{ij} = (e_j, f_i)$, for $1 \le i, j \le n$.

The transition matrix S in (54) is unitary because $S_{ij} = (e_j, f_i) \Rightarrow$

$$\begin{aligned}
\big(\mathrm{Row}_i(S)\,,\, \mathrm{Row}_j(S)\big)_{\mathbb{K}^n} &= \sum_k S_{ik}\overline{S_{jk}} \;=\; \sum_k (e_k, f_i)\,\overline{(e_k, f_j)} \\
&= \sum_k (f_j, e_k)\,(e_k, f_i) \;=\; (f_j, f_i) = \delta_{ij}
\end{aligned}$$

by Parseval's identity. Then $S^* = S^{-1} = [\mathrm{id}]^{-1}_{\mathfrak{Y}\mathfrak{X}} = [\mathrm{id}]_{\mathfrak{X}\mathfrak{Y}}$ by Theorem 4.22, and

$$[T]_{\mathfrak{Y}\mathfrak{Y}} = S\,[T]_{\mathfrak{X}\mathfrak{X}}S^* = S\,[T]_{\mathfrak{X}\mathfrak{X}}S^{-1} \quad \square$$

We conclude that the various matrix realizations of T with respect to orthonormal bases in V are related by unitary equivalence (similarity modulo a unitary matrix) rather than similarity modulo a matrix that is merely invertible. Unitary equivalence is therefore a more stringent condition on two matrices than similarity (as defined in Chapter V).

Elements U in the unitary group $\mathrm{U}(n)$ act on matrix space $X = \mathrm{M}(n, \mathbb{C})$ by conjugation, sending

$$A \mapsto UAU^{-1} = UAU^*\,.$$

This group action $\mathrm{U}(n) \times X \to X$ partitions X into disjoint orbits

$$\mathcal{O}_A = \mathrm{U}(n) \cdot A = \{UAU^* : U \in \mathrm{U}(n)\}\ ,$$

which are the the **unitary equivalence classes** in matrix space. There is a similar group action $\mathrm{O}(n) \times \mathrm{M}(n, \mathbb{R}) \to \mathrm{M}(m, \mathbb{R})$ of the orthogonal group on real matrices. Recall that the *similarity class* of an $n \times n$ matrix A is its orbit $\mathrm{GL}(n, \mathbb{K}) \cdot A = \{EAE^{-1} : E \in \mathrm{GL}(n, \mathbb{K})\}$ under the action of the general linear group $\mathrm{GL}(n, \mathbb{K}) = \{A : \det(A) \neq 0\}$, which is considerably larger than $\mathrm{U}(n)$ or $\mathrm{O}(n)$ and has larger orbits.

Diagonalization over $\mathbb{K} = \mathbb{C}$: A Summary. We recall that the spectra $\mathrm{sp}_{\mathbb{C}}(T)$ of operators over $\mathbb{C}$ and their adjoints have the following properties.

1. For any T, $\mathrm{sp}(T^*) = \overline{\mathrm{sp}(T)}$ and $\dim E_{\overline{\lambda}}(T^*) = \dim E_\lambda(T)$. But as we will see in 4.14E below, the $\bar{\lambda}$ eigenspace $E_{\bar{\lambda}}(T^*)$ is not always equal to $E_\lambda(T)$ unless T is normal.

2. If $T = T^*$ then T is orthogonally diagonalizable, and all eigenvalues are *real* because $T(v) = \lambda \Rightarrow$
$$\lambda \|v\|^2 = (T(v), v) = \big(v, T^*(v)\big) = (v, \lambda v) = \overline{\lambda}\|v\|^2$$

3. If T is unitary then all eigenvalues satisfy $|\lambda| = 1$ (they lie on the unit circle in $\mathbb{C}$), because
$$\begin{aligned} T(v) = \lambda \cdot v \quad &\Rightarrow \quad \|v\|^2 = (T^*Tv, v) = (Tv, Tv) = (\lambda v, \lambda v) = |\lambda|^2 \cdot \|v\|^2 \\ &\Rightarrow \quad |\lambda|^2 = 1 \text{ if } v \neq 0 \end{aligned}$$

4. If T is skew-adjoint, so $T^* = -T$, then all eigenvalues are pure imaginary because
$$\lambda \|v\|^2 = (Tv, v) = (v, T^*v) = (v, -T(v)) = (v, -\lambda v) = -\overline{\lambda}\|v\|^2$$
Consequently, $\overline{\lambda} = -\lambda$ and $\lambda \in 0 + i\mathbb{R}$ in $\mathbb{C}$.

5. A general normal operator is orthogonally diagonalizable, but there are no restrictions on the pattern of eigenvalues.

In Theorem 4.11 we proved the following necessary and sufficient condition for a linear operator on a complex inner product space to be diagonalizable.

4.28. Theorem (Orthogonal Diagonalization). *A linear operator $T : V \to V$ on a finite dimensional complex inner product space is orthogonally diagonalizable $\Leftrightarrow$ T is normal (so $T^*T = TT^*$).* $\square$

VI.5. Some Operators on Real Inner Product Spaces: Reflections, Rotations and Rigid Motions.

All this works over $\mathbb{K} = \mathbb{R}$ except that in this context unitary operators are referred to as **orthogonal transformations**. The corresponding matrices $A = [T]_{\mathfrak{X},\mathfrak{X}}$ with respect to orthonormal bases satisfy $A^{\mathrm{t}}A = I = AA^{\mathrm{t}}$, so $A^{\mathrm{t}} = A^{-1}$ in $\mathrm{M}(n, \mathbb{R})$. An orthogonal transformation might not have enough real eigenvalues to be diagonalizable, which happens $\Leftrightarrow$ the eigenspaces $E_\lambda(T)$ ($\lambda \in \mathbb{R}$) fail to span V. In fact there might not be any real eigenvalues at all. For example, if $R_\theta =$ (counterclockwise rotation about origin by θ radians) in $\mathbb{R}^2$, and if θ is not an integer multiple of π, then with respect to the standard $\mathbb{R}$-basis $\mathfrak{X} = \{e_1, e_2\}$ we have
$$[R_\theta]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$
whose complex eigenvalues are $e^{i\theta}$ and $e^{-i\theta}$; there are no real eigenvalues if $\theta \neq n\pi$, even though R_θ is a normal operator. (A rotation by $\theta \neq n\pi$ radians cannot send a vector

$v \neq 0$ to a scalar multiple of itself.)

The Group of Rigid Motions $\mathrm{M}(n)$**.** **Rigid motions** on $\mathbb{R}^n$ are the bijective maps $\rho : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that preserve distances between points,

$$\|\rho(x) - \rho(y)\| = \|x - y\| \quad \text{for all } x, y \ .$$

We DO NOT assume ρ is linear. The rigid motions form a group $\mathrm{M}(n)$ under composition; it includes two important subgroups

1. TRANSLATIONS: Operators $T = \{t_b : b \in \mathbb{R}^n\}$ where

   $$t_b(x) = x + b \qquad \text{for all } x \in \mathbb{R}^n \ \ (b \in \mathbb{R}^n \text{ fixed})$$

   Under the bijective map $\phi : \mathbb{R}^n \rightarrow T$ with $\phi(t) = t_b$ we have $\phi(s + t) = \phi(s) \circ \phi(t)$ and $\phi(0) = \mathrm{id}_{\mathbb{R}^n}$. Obviously translations are isometric mappings since

   $$\| \, t_b(x) - t_b(y)\| = \|(x + b) - (y - b)\| = \|x - y\| \qquad \text{for all } b \text{ and } x, y$$

   but they are NOT linear operators on $\mathbb{R}^n$ (unless $b = 0$) because the zero element does not remain fixed: $t_b(0) = b$.

2. LINEAR ISOMETRIES: Operators $H = \{L_A : A \in \mathrm{O}(n)\}$ where $L_A(x) = A{\cdot}x$ and A is any orthogonal real $n \times n$ matrix (so A is invertible with $A^{\mathrm{t}} = A^{-1}$).

Although rigid motions need not be linear operators, it is remarkable that they are nevertheless simple combinations of a linear isometry (an orthogonal linear mapping on $\mathbb{R}^n$) and a translation operator.

$$\rho(x) = \big(t_b \circ L_A\big)(x) = A{\cdot}x + b \qquad (b \in \mathbb{R}^n, A \in \mathrm{O}(n)) \tag{55}$$

for all $x \in \mathbb{R}^n$. In particular, any rigid motion $\rho : \mathbb{R}^n \rightarrow \mathbb{R}^n$ that leaves the origin fixed is *automatically linear.*

5.1 Proposition. *If $\rho : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a rigid motion that fixes the origin (so $\rho(\mathbf{0}) = \mathbf{0}$), then ρ is in fact a* LINEAR *operator on $\mathbb{R}^n$, $\rho = L_A$ for some $A \in \mathrm{O}(n)$. In general, every rigid motion is a composite of the form* (55).

Proof: The second statement is immediate from the first, for if ρ moves the origin to $b = \rho(\mathbf{0})$, the operation $t_{-b} \circ \rho$ is a rigid motion that fixes the origin, and $\rho = t_b \circ (t_{-b} \circ \rho)$.

To prove the first assertion, let $\{\mathbf{e}_j\}$ be the standard orthonormal basis in $\mathbb{R}^n$ and let $\mathbf{e}'_j = \rho(\mathbf{e}_j)$. Since $\rho(\mathbf{0}) = 0$ lengths are preserved because $\|\rho(\mathbf{x})\| = \|\rho(\mathbf{x}) - \rho(\mathbf{0}) = \|\mathbf{x}\|$, and then inner products are also preserved because

$$\begin{aligned} -2\left(\rho(\mathbf{x}), \rho(\mathbf{y})\right) &= \|\rho(\mathbf{x}) - \rho(\mathbf{y})\|^2 - \|\rho(\mathbf{x})\|^2 - \|\rho(\mathbf{y})\|^2 \\ &= \|\mathbf{x} - \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 \ = \ -2\,(\mathbf{x}, \mathbf{y}) \end{aligned}$$

Hence the images $\mathbf{e}'_i = \rho(\mathbf{e}_i)$ of the standard basis vectors are also an orthonormal basis.

Now let A be the matrix whose i^{th} column is $\mathbf{e}'_i$, so $L_A(\mathbf{e}_i) = \mathbf{e}'_i$. Then A is in $\mathrm{O}(n)$, L_A and $(L_A)^{-1} = L_{A^{-1}}$ are both *linear* orthogonal transformations on $\mathbb{R}^n$, and the product $L_A^{-1} \circ \rho$ as a rigid motion that fixes each $\mathbf{e}_i$ as well as the zero vector. But any such motion must be the identity map. In fact if $\mathbf{x} \in \mathbb{R}^n$ then $(\mathbf{x}, \mathbf{e}_i) = (\rho\mathbf{x}, \rho\mathbf{e}_i) = (\rho\mathbf{x}, \mathbf{e}'_i)$, and since $\mathbf{e}'_i = \mathbf{e}_i$ we get

$$x_i = (\mathbf{x}, \mathbf{e}_i) = (\rho\mathbf{x}, \mathbf{e}'_i) = (\rho\mathbf{x}, \mathbf{e}_i) = x'_i$$

for all i. Hence $\mathbf{x}' = \rho\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$, as claimed. $\square$

Every rigid motion on $\mathbb{R}^n$,

$$T(x) = A \cdot x + t_b = (t_b \circ L_A) \qquad \text{with } A \in \mathrm{O}(n) \text{ and } b \in \mathbb{R}^n$$

has two components, an orthogonal *linear* map L_A and a translation t_b. Rigid motions are of two types, **orientation preserving** and **orientation reversing**. Translations always preserve orientation of geometric figures, so the nature of a rigid motion T is determined by its linear component L_A, which preserves orientation if $\det(A) > 0$ and reverses it if $\det(A) < 0$. As a simple illustration, consider the matrices (with respect to the standard basis $\mathfrak{X}$ in $\mathbb{R}^2$) of a rotation about the origin R_θ (orientation preserving), and a reflection r_y across the y-axis (orientation reversing).

$$[R_\theta]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad\qquad [r_y]_{\mathfrak{X}\mathfrak{X}} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Rotation: R_θ, $\det[R_\theta] = +1$ Reflection: r_y, $\det[r_y] = -1$

We now examine rotations and reflections in more detail, describing them in terms of the inner product in $\mathbb{R}^n$.

5.2 Example (Reflections in Inner Product Spaces). *If V is an inner product space over $\mathbb{R}$, a* **hyperplane** *in V is any vector subspace M with $\dim(M) = n-1$ (so M has "codimension 1" in V). This determines a reflection of vectors across M.*

Discussion: Since $V = M \dot{\oplus} M^\perp$ (orthogonal direct sum) every vector v splits uniquely as $v = v_\| + v_\perp$ (with "parallel component" $v_\| \in M$, and $v_\perp \in M^\perp$). By definition, reflection r_M across M is the (linear) operator that reverses the "perpendicular component" $v_\perp$, so that

(56) $$r_M(v_\| + v_\perp) = v_\| - v_\perp = v - 2 \cdot v_\perp$$

as shown in Figure 6.9.

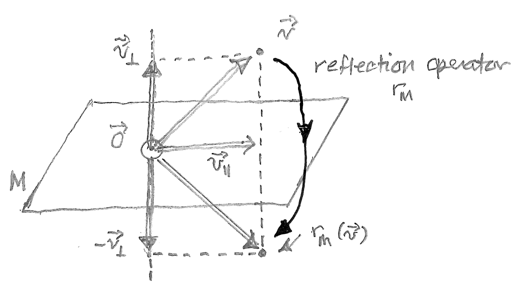

Figure 6.9. Geometric meaning of reflection r_M across an $(n-1)$-dimensional hyperplane in an n-dimensional inner product space over $\mathbb{R}$

Now, let $\{e_1, ..., e_{n-1}\}$ be an orthonormal basis in the subspace M and let e_n be $v_\perp$ renormalized to make $\|e_n\| = 1$, so $M^\perp = \mathbb{R}e_n$. We have seen that

$$v_\| = \sum_{k=1}^{n-1} (v, e_k)\, e_k \ ,$$

so $v_\perp = v - v_\| = c \cdot e_n$ for some $c \in \mathbb{R}$. But in fact $c = (v, e_n)$ because

$$c = (ce_n, e_n) = (v - v_\|, e_n) = (v, e_n) + 0$$

This yields an important formula involving only the inner product.

$$r_M = v_{\|} - v_{\perp} = \big(v_{\|} + v_{\perp}\big) - 2\cdot v_{\perp} = v - 2(v, e_n)\cdot e_n \tag{57}$$

Note: we need $\|e_n\| = 1$ to make this work. $\square$

5.3 Exercise. Show that (57) implies the following properties for any reflection.

(a) $r_M \circ r_M = id_V$, so r_M is its own inverse;

(b) $\det(r_M) = -1$, so all reflections are orientation-reversing.

(c) M is the set of fixed points $\text{Fix}(r_M) = \{x : r_M(x) = x\}$. $\square$

5.4 Exercise. Prove that every reflection r_M on an inner product space preserves distances,

$$\|r_M(x) - r_M(y)\| = \|x - y\|$$

for all $x, y \in V$.

5.5 Exercise. If M is a hyperplane in a real inner product space V and $b \notin M$, the translate $b + M$ (a coset in V/M) is an $n-1$ dimensional hyperplane parallel to M (but is not a vector subspace). Explain why the operation that reflects vectors across $M' = b + M$ must be the rigid motion $T = t_b \circ r_M \circ t_{-b}$.
Hint: Check that $T^2 = T$ and that the set of fixed points $\text{Fix}(T) = \{v \in V : T(v) = v\}$ is precisely M'.

In another direction, we have Euler's famous geometric characterization of orientation preserving orthogonal transformations $L_A : \mathbb{R}^3 \to \mathbb{R}^3$ with $A^{\mathrm{t}}A = I = AA^{\mathrm{t}}$ in $\mathrm{M}(3, \mathbb{R})$ and $\det(A) > 0$. In fact, $\det(A) = +1$ since $A^{\mathrm{t}}A = I$ implies $\big(\det(A)\big)^2 = 1$, so $\det(A) = \pm 1$ for $A \in \mathrm{O}(n)$.

5.6 Theorem (Euler). *Let $A \in \mathrm{SO}(3) = \{A \in \mathrm{M}(3, \mathbb{R}) : A^{\mathrm{t}}A = I$ and $\det(A) = 1\}$. If $A \neq I$ then $\lambda = 1$ is an eigenvalue such that $\dim_{\mathbb{R}}\big(E_{\lambda=1}\big) = 1$. If $v_0 \neq 0$ in $E_{\lambda=1}$ and $\ell = \mathbb{R}\, v_0$ there is some angle $\theta \notin 2\pi\mathbb{Z}$ such that*

$$L_A = R_{\ell,\theta} = (\textit{rotation by } \theta \textit{ radians about the line } \ell \textit{ through the origin}).$$

(*Rotations by a positive angle are determined by the usual "right hand rule," with your thumb pointing in the direction of v_0*).

Proof: The characteristic polynomial $p_T(x)$ for $T = L_A$ has real coefficients. Regarded as a polynomial $p_T \in \mathbb{R}[x] \subseteq \mathbb{C}[x]$, its complex roots are either real or occur in conjugate pairs $z = x + iy$, $\overline{z} = x - iy$ with $y \neq 0$. Since $\text{degree}(p_T) = 3$ there must be at least one real root λ. But because $T = L_A$ is unitary its complex eigenvalues have $|\lambda| = 1$, because if $v \neq 0$ in E_λ,

$$\|v\|^2 = \big(T(v), T(v)\big) = (\lambda v, \lambda v) = |\lambda|^2\, \|v\|^2 \;\Rightarrow\; |\lambda|^2 = 1\ .$$

If λ is real the only possibilities are $\lambda = \pm 1$. The real roots cannot all be -1, for then $\det(T) = (-1)^3 = -1$ and we require $\det(T) = +1$. Thus $\lambda = 1$ is an eigenvalue, and we will see below that $\dim_{\mathbb{R}}\big(E_{\lambda=1}\big) = 1$.

If $v_0 \neq 0$ in $E_{\lambda=1}$, let $M = \mathbb{R}v_0$. Then $M^\perp$ is 2-dimensional and is invariant under both T and $T^* = T^{-1}$. Furthermore (see Exercise 5.7) the restriction $T|_{M^\perp}$ is a unitary (= orthogonal) transformation on the 2-dimensional space $M^\perp$ equipped with the inner product it inherits from $\mathbb{R}^3$. If we fix an orthonormal basis $\{f_1, f_2\}$ in $M^\perp$ and let $f_0 = v_0/\|v_0\|$, we obtain an orthonormal basis for $\mathbb{R}^3$. The matrix A of $T|_{M^\perp}$ with respect to $\mathfrak{X}_0 = \{f_1, f_2\}$ is in

$$\mathrm{SO}(2) = \{A \in \mathrm{M}(2, \mathbb{R}) : A^{\mathrm{t}}A = I \text{ and } \det(A) = 1\}$$

because the matrix of T with respect to the orthonormal basis $\mathfrak{X} = \{f_0, f_1, f_2\}$ is

$$[T]_{\mathfrak{X}} = \left[\begin{array}{cc} 1 & 0 \\ 0 & A \end{array}\right] ,$$

so $A \in \mathrm{SO}(2)$ because $1 = \det\big([T]_{\mathfrak{X}}\big) = 1 \cdot \det(A)$. As noted below in Exercise VI-5.8, if $A \in \mathrm{SO}(2)$ the rows form an orthonormal basis for $\mathbb{R}^2$ and so do the columns, hence there exist $a, b \in \mathbb{R}$ such that

$$a^2 + b^2 = 1 \qquad \text{and} \qquad A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

It follows easily that there is some $\theta \in \mathbb{R}$ such that

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

This is the matrix $A = [R_\theta]_{\mathfrak{X}_0}$ of a rotation by θ radians about the origin in $M^\perp$, so $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a rotation $R_{\ell,\theta}$ by θ radians about the axis $\ell = \mathbb{R}v_0$. $\square$

We cannot have $\theta \in 2\pi\mathbb{Z}$ because then $T = \mathrm{id}$ is not really a rotation about any well-defined axis); that's why we required $A \neq I$ in the theorem.

5.7 Exercise. Let $T : V \to V$ be a linear operator on a finite dimensional inner product space, and M a subspace that is invariant under both T and T^*. Prove that the restriction $(T|M) : M \to M$ is unitary with respect to the inner product M inherits from V.
Hint: Recall Exercise 4.10.

5.8 Exercise. If $A = [a, b; c, d] \in \mathrm{M}(2, \mathbb{R})$ verify that $A^{\mathrm{t}}A = I \Leftrightarrow$ the rows of A are an orthonormal basis in $\mathbb{R}^2$, so that

$$a^2 + b^2 = 1 \qquad c^2 + d^2 = 1 \qquad ac + bd = 0$$

If, in addition we have

$$\det(A) = ad - bc = +1$$

prove that $c = -b$, $d = a$ and $a^2 + b^2 = 1$, and then explain why there is some $\theta \in \mathbb{R}$ such that $a = \cos(\theta)$ and $b = -\sin(\theta)$.
Note: Thus $L_A : \mathbb{R}^2 \to \mathbb{R}^2$ is a counterclockwise rotation about the origin by θ radians.
Hint: For the last step, think of $a^2+b^2 = 1$ in terms of a right triangle whose hypoteneuse has length $= 1$.

5.9 Exercise. Consider the linear map $L_A : \mathbb{R}^2 \to \mathbb{R}^2$ for the matrix

$$A = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \qquad \text{in } \mathrm{O}(2)$$

What is the geometric action of L_A? If a rotation, find the angle θ; if not, show that the set of fixed points for L_A is a line through the origin L, and $L_A =$ (reflection across L).

5.10 Exercise. If $A = \left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$ is in $\mathrm{O}(2)$ and has $\det(A) = -1$,

1. Prove that $L_A : \mathbb{R}^2 \to \mathbb{R}^2$ is reflection across some line ℓ through the origin.
2. Explain why

   $$a^2 + b^2 = 1 \qquad c^2 + d^2 = 1 \qquad ac + bd = 0 \qquad \det(A) = ad - bc = -1$$

   and then show there is some θ such that $A = \left[\begin{array}{cc} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{array}\right]$

Note: The preceding matrix is *not* a rotation matrix since $\det(A) = -1$. The angle θ determined here is related to the angle between the line of reflection ℓ and the $+x$-axis. **Hints:** The map $L_A : \mathbb{C}^2 \to \mathbb{C}^2$ is unitary, and in particular is orthogonally diagonalizable. What are the possible patterns of complex eigenvalues (counted according to multiplicity), and how do they relate to the requirement that $\det(A) = -1$?

VI.6. Spectral Theorem for Vector and Inner Product Spaces.

If V is a vector space over a field $\mathbb{K}$ (not necessarily an inner product space), and if $T : V \to V$ is diagonalizable over $\mathbb{K}$, then $V = \bigoplus_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$ (an ordinary direct sum) – see Proposition II-3.9. This decomposition determines projection operators $P_\lambda = P_\lambda^2$ of V onto $E_\lambda(T)$ along the complementary subspaces $\bigoplus_{\mu \neq \lambda} E_\mu(T)$. The projections $P_\lambda = P_\lambda(T)$ have the following easily verified properties:

1. $P_\lambda^2 = P_\lambda$
2. $P_\lambda P_\mu = P_\mu P_\lambda = 0$ if $\lambda \neq \mu$ in $\mathrm{sp}(T)$;
3. $I = \sum_\lambda P_\lambda$;

Condition (1.) simply reflects the fact that P_λ is a projection operator. Each $v \in V$ has a unique decomposition $v = \sum_\lambda v_\lambda$ with $v_\lambda \in E_\lambda(T)$, and (by definition) $P_\lambda(v) = v_\lambda$. Property (3.) follows from this. For (2.) write $v = \sum_\lambda v_\lambda$ and consider distinct $\alpha \neq \beta$ in $\mathrm{sp}(T)$. Then

$$P_\alpha P_\beta(v) = P_\alpha P_\beta\Big(\sum_\lambda v_\lambda\Big) = P_\alpha(v_\beta) = 0 \qquad \text{(since } \alpha \neq \beta\text{)}$$

and similarly for $P_\beta P_\alpha$. The operators $\{P_\lambda : \lambda \in \mathrm{sp}_{\mathbb{K}}(T)\}$ are the **spectral projections** associated with the diagonalizable operator T.

Now let V be an inner product space. If T is orthogonally diagonalizable we have additional information regarding the spectral projections $P_\lambda(T)$:

4. The eigenspaces $E_\lambda(T)$ are orthogonal, $E_\lambda \perp E_\mu$ if $\lambda \neq \mu$, and $V = \dot{\bigoplus}_\lambda E_\lambda(T)$ is an orthogonal direct sum decomposition.
5. The P_λ are *orthogonal* projections, hence they are *self-adjoint* in addition to having the preceeding properties, so that $P_\lambda^2 = P_\lambda = P_\lambda^*$.

In this setting we can prove useful facts relating diagonalizability and eigenspaces of an operator $T : V \to V$ and its adjoint T^*. These follow by recalling that there is a natural isomorphism between any finite dimensional inner product space V and its dual space V^*, as explained in Lemma VI-3.2. Therefore given any basis $\mathfrak{X} = \{e_1, \ldots, e_n\}$ in V there exists *within* V a matching basis $\mathfrak{X}' = \{f_1, \ldots, f_n\}$ that is "dual to" $\mathfrak{X}$ in the sense that

$$(e_i\,,\, f_j) = \delta_{ij} \qquad \text{(Kronecker delta)}$$

These paired bases can be extremely useful in comparing properties of T with those of its adjoint T^* .

6.1 Exercise. Let $\mathfrak{X} = \{e_1, \ldots, e_n\}$ be an arbitrary basis (not necessarily orthonormal) in a finite dimensional inner product space V.

(a) Use induction on n to prove that there exist vectors $\mathfrak{Y} = \{f_1, \ldots, f_n\}$ such that $(e_i, f_j) = \delta_{ij}$.

(b) Explain why the f_j are uniquely determined and a basis for V.

Note: If the initial basis $\mathfrak{X}$ is orthonormal then $f_i = e_i$ and the result trivial; we are interested in *arbitrary* bases in an inner product space.

6.1A Exercise. Let V be an inner product space and T a linear operator that is diagonalizable in the ordinary sense, but not necessarily orthogonally diagonalizable. Prove that

(a) The adjoint operator T^* is diagonalizable. What can you say about its eigenvalues and eigenspaces?

(b) T is *orthogonally* diagonalizable so is T^*.

Hint: If $\{e_i\}$ diagonalizes T what does the "dual basis" $\{f_j\}$ of Exercise 6.1 do for T^*?

6.1B Exercise. If V is a finite dimensional inner product space and $T : V \rightarrow V$ is diagonalizable in the ordinary sense, prove that the spectral projections for T^* are the adjoints of those for T:

$$P_{\overline{\lambda}}(T^*) = \big(P_\lambda(T)\big)^* \qquad \text{for all } \lambda \in \text{sp}(T)$$

Hint: Use VI-6.1A and dual diagonalizing bases; we already know $\text{sp}(T^*) = \overline{\text{sp}(T)}$.
Note: $\big(P_\lambda(T)\big)^*$ might differ from $P_\lambda(T)$.

We now procede to prove the spectral theorem and examine its many applications.

6.2 Theorem (The Spectral Theorem). *If a linear operator $T : V \rightarrow V$ is diagonalizable on a finite dimensional vector space V over a field $\mathbb{K}$, and if $\{P_\lambda : \lambda \in \text{sp}_{\mathbb{K}}(T)\}$ are the spectral projections, then T has the following description in terms of those projections*

$$T = \sum_{\lambda \in \text{sp}(T)} \lambda \cdot P_\lambda \tag{58}$$

If $f(x) = \sum_{k=0} c_k x^k \in \mathbb{K}[x]$ is any polynomial the operator $f(T) = \sum_{k=0} c_k T^k$ takes the form

$$f(T) = \sum_{\lambda \in \text{sp}(T)} f(\lambda) \cdot P_\lambda \tag{59}$$

In particular, the powers T^k are diagonalizable, with $T^k = \sum_{\lambda \in \text{sp}(T)} \lambda^k \cdot P_\lambda$.

If we define the map $\Phi : \mathbb{K}[x] \rightarrow \text{Hom}_{\mathbb{K}}(V, V)$ from polynomials to linear operators on V, letting $\Phi(1) = I$ and

$$\Phi(f) = \sum_{k=0} c_k T^k \qquad \textit{for} \qquad f(x) = \sum_{k=0} c_k x^k \quad ,$$

then Φ is linear and a homomorphism of associative algebras over $\mathbb{K}$, so that

$$\Phi(fg) = \Phi(f) \circ \Phi(g) \qquad \textit{for } f, g \in \mathbb{K}[x] \tag{60}$$

Finally, $\Phi(f) = 0$ (the zero operator on V) if and only if $f(\lambda) = 0$ for each $\lambda \in \text{sp}_{\mathbb{K}}(T)$. Thus $\Phi(f) = \Phi(g)$ if and only if f and g take same values on the spectrum $\text{sp}(T)$, so many polynomials $f \in \mathbb{K}[x]$ can yield the same operator $f(T)$.

Note: This is all remains true for orthogonally diagonalizable operators on an inner product space, but in this case we have the additional property

$$\Phi(\overline{f}\,) = \Phi(f)^* \qquad \text{(adjoint operator)} \tag{61}$$

where $\overline{f}(x) = \sum_{k=0} \overline{c_k} x^k$ and $\bar{c}$ is the complex conjugate of c. $\square$

Proof of (6.2): If $v \in V$ decomposes as $v = \sum_\lambda v_\lambda \in \bigoplus_{\lambda \in \mathrm{sp}(T)} E_\lambda(T)$, then

$$\begin{aligned} T(v) = T\Big(\sum_\lambda v_\lambda\Big) &= \sum_\lambda \lambda \cdot v_\lambda = \sum_\lambda \lambda \cdot P_\lambda(v) \\ &= \Big(\sum_{\lambda \in \mathrm{sp}(T)} \lambda \cdot P_\lambda\Big)\, v \end{aligned}$$

for all $v \in V$, proving (58). Then $T^k = \sum_\lambda \lambda^k P_\lambda$ becomes

$$T^k(v) = T^k\Big(\sum_\lambda v_\lambda\Big) = \sum_\lambda T^k v_\lambda$$

But $T(v_\lambda) = \lambda v_\lambda \Rightarrow$

$$T^2(v_\lambda) = T\big(\lambda \cdot v_\lambda\big) = \lambda^2 v_\lambda, \quad T^3(v_\lambda) = \lambda^3 v_\lambda, \quad \text{etc}$$

so if $v = \sum_\lambda v_\lambda$ we get

$$T^k(v) = \sum_\lambda \lambda^k v_\lambda = \sum_\lambda \lambda^k P_\lambda(v) = \Big(\sum_\lambda \lambda^k P_\lambda\Big)\, v$$

for all $v \in V$. Noting that the powers T^k and the sum $f(T)$ are linear operators, (59) follows: For any $f(x) = \sum_k c_k x^k$ we have

$$\begin{aligned} f(T)(v) &= f(T)\Big(\sum_\lambda v_\lambda\Big) = \sum_\lambda f(T)(v_\lambda) \\ &= \sum_\lambda \Big(\sum_k c_k T^k\Big)(v_\lambda) = \sum_\lambda \sum_k c_k T^k(v_\lambda) \\ &= \sum_\lambda \sum_k c_k \lambda^k v_\lambda = \sum_\lambda \Big(\sum_k c_k \lambda^k\Big)\, v_\lambda \\ &= \sum_\lambda f(\lambda)\, v_\lambda = \sum_\lambda f(\lambda) P_\lambda(v) \\ &= \Big(\sum_\lambda f(\lambda) P_\lambda\Big)\, v \qquad \text{for all } v \in V \end{aligned}$$

Thus $f(T) = \sum_\lambda f(\lambda) P_\lambda$ as operators on V.

When $f(x)$ is the constant polynomial $f(x) = 1$ we get

$$\sum_{\lambda \in \mathrm{sp}(T)} f(\lambda) P_\lambda = \sum_\lambda P_\lambda = I$$

as expected. Linearity of Φ is easily checked by applying the operators on either side to a typical vector. As for the multiplicative property, let $f = \sum_{k=0} a_k x^k$ and $g = \sum_{\ell \geq 0} b_\ell x^\ell$, so $fg = \sum_{k,\ell=0} a_k b_\ell\, x^{k+\ell}$. First notice that the multiplicative property holds for monomials $f = x^k$, $g = x^\ell$ because

$$\begin{aligned} \Phi(x^k)\Phi(x^\ell) &= \Big(\sum_{\lambda \in \mathrm{sp}(T)} \lambda^k P_\lambda\Big) \cdot \Big(\sum_{\mu \in \mathrm{sp}(T)} \mu^\ell P_\mu\Big) \\ &= \sum_{\lambda,\mu} \lambda^k \mu^\ell\, P_\lambda P_\mu = \sum_\lambda \lambda^{k+\ell} P_\lambda \\ &= \Phi(x^{k+\ell}) \end{aligned}$$

$(P_\lambda P_\mu = 0$ if $\lambda \neq \mu$, and $P_\lambda^2 = P_\lambda)$. Then use linearity of Φ to get

$$\begin{aligned}
\Phi(fg) &= \Phi\Big(\sum_{k,\ell=0} a_k b_\ell\, x^{k+\ell}\Big) = \sum_{k,\ell=0} a_k b_\ell \Phi(x^{k+\ell}) \\
&= \sum_{k,\ell} a_k b_\ell \Phi(x^k)\Phi(x^\ell) = \Big(\sum_{k=0} a_k \Phi(x^k)\Big)\cdot\Big(\sum_{\ell=0} b_\ell \Phi(x^\ell)\Big) \\
&= \Phi(f)\circ\Phi(g)
\end{aligned}$$

That completes the proof of Theorem 6.2. $\square$

Although the operator $\Phi(f) = \sum_\lambda f(\lambda)P_\lambda$ was defined for polynomials in $\mathbb{K}[x]$, this sum involves only the values of f on the finite subset $\mathrm{sp}(T) \subseteq \mathbb{K}$, so it makes sense for *all* functions $h : \mathrm{sp}(T) \to \mathbb{K}$ whether or not they are defined off of the spectrum, or related in any way to polynomials. Thus the spectral decomposition of T determines a linear map

(62) $$\Phi : \mathcal{E} \to \mathrm{Hom}_{\mathbb{K}}(V,V) \qquad \Phi(h) = \sum_{\lambda\in\mathrm{sp}(T)} h(\lambda)P_\lambda$$

defined on the larger algebra $\mathcal{E} \supseteq \mathbb{F}[x]$ whose elements are arbitrary functions h from $\mathrm{sp}_{\mathbb{K}}(T) \to \mathbb{K}$. The same argument used for polynomials shows that the extended version of Φ is again a homomorphism between associative algebras, as in (60). Incidentally, the Lagrange Interpolation formula tells us that any $h(x)$ in $\mathcal{E}$ is the restriction of some (nonunique) polynomial $f(x)$, so that

$$\Phi(h) = \Phi\big(f\big|_{\mathrm{sp}(T)}\big) = \Phi(f)$$

All this applies to matrices as well as operators since a matrix is diagonalizable $\Leftrightarrow$ the left multiplication operator $L_A : \mathbb{K}^m \to \mathbb{K}^m$ on coordinate space is diagonalizable.

We can now define "functions $h(T)$ of an operator" for a much broader class of functions than polynomials, as in the next examples.

6.3 Example. If a diagonalizable linear operator $T : V \to V$ over $\mathbb{C}$ has spectral decomposition $T = \sum_\lambda \lambda\cdot E_\lambda$, we can define such operators $h(T)$ as

1. $|T| = \sum_\lambda |\lambda| = h(T)$ taking $h(z) = |z|$.
2. $e^T = \sum_\lambda e^\lambda = h(T)$ taking $h(z) = e^z = \sum_{n=0}^\infty z^n/n!$
3. $\sqrt{T} = \sum_\lambda \lambda^{1/2} P_\lambda$ assigning any (complex) determination of $h(z) = \sqrt{z}$ at each point in the spectrum. Thus there are r^2 possible operator square roots if T has r distinct eigenvalues that are all nonzero. As in Exercise 6.4 below, every such "square root" has the property $h(T)^2 = T$.
4. The **indicator function** of a finite subset $E \subseteq \mathbb{C}$ is

   $$1_E(z) = \begin{cases} 1 & \text{if } z \in E \\ 0 & \text{otherwise} \end{cases}$$

   Then by (60), $1_E(T)$ is a projection operator with $1_E(T)^2 = 1_E(T)$. In particular, if $E = \{\lambda_1, \ldots, \lambda_s\} \subseteq \mathrm{sp}(T)$ we have

   $$1_E(T) = \sum_{\lambda\in E} P_\lambda = \bigoplus_{i=1}^s P_{\lambda_i} \ \Big(\text{projection onto } \bigoplus_{i=1}^s E_{\lambda_i}(T)\Big) \ .$$

   We get $1_E(T) = I$ if $E = \mathrm{sp}(T)$, and if $E = \{\lambda_0\}$ is a single eigenvalue we recover the individual spectral projections: $1_E(T) = P_{\lambda_0}$. $\square$

6.4 Exercise. Let $T : V \to V$ be a diagonalizable linear operator over any ground field $\mathbb{K}$. If T is invertible ($\lambda = 0$ not in the spectrum), explain why

$$h(T) = \sum_{\lambda \in \mathrm{sp}(T)} \frac{1}{\lambda} \cdot P_\lambda \qquad \big(h(x) = \frac{1}{x} \text{ for } x \neq 0\big)$$

is the usual inverse T^{-1} of T.
Hint: Show $T \circ h(T) = h(T) \circ T = I$

Similarly we have

$$T^{-k} = \left(T^{-1}\right)^k = \sum_\lambda \frac{1}{\lambda^k} P_\lambda$$

for $k = 0, 1, 2 \ldots$, with $T^0 = I$.

6.5 Exercise. Prove (61) when V is an inner product space over $\mathbb{C}$. (There is nothing to prove when $\mathbb{K} = \mathbb{R}$.)

6.6 Exercise. Prove that a normal operator $T : V \to V$ on a finite dimensional inner product space over $\mathbb{C}$ is self adjoint if and only if its spectrum is real: $\mathrm{sp}_\mathbb{C}(T) \subseteq \mathbb{R} + i0$.
Note: We already explained ($\Rightarrow$); you do ($\Leftarrow$).

6.7 Exercise. If T is diagonalizable over $\mathbb{R}$ or $\mathbb{C}$, prove that

$$e^T = \sum_{\lambda \in \mathrm{sp}(T)} e^\lambda P_\lambda$$

is the same as the linear operator given by the exponential series

$$e^T = \sum_{k=0}^{\infty} \frac{1}{k!} T^k$$

. **Note:** If T has spectral decomposition $T = \sum_\lambda \lambda \cdot P_\lambda$ then $T^k = \sum_\lambda \lambda^k P_\lambda$. To discuss convergence of the operator-valued exponential series in Exercise VI-6.7, fix a basis $\mathfrak{X} \subseteq V$. Then a sequence of operators converges, with $T_n \to T$ as $n \to \infty$, if and only if the corresponding matrices converge entry-by-entry, $[T_n]_{\mathfrak{X}\mathfrak{X}} \to [T]_{\mathfrak{X}\mathfrak{X}}$ as $n \to \infty$ in matrix space, as described in Chapter II, Section 5.3. The partial sums of a series converge to a limit

$$S_n = I + T + \frac{1}{2!} T^2 + \ldots + \frac{1}{n!} T^n \;\to\; S_0 \ ,$$

$\Leftrightarrow$ $(S_n)_{ij} \to (S_0)_{ij}$ in $\mathbb{C}$ for all $1 \leq i, j \leq N$. $\square$

6.8 Exercise. Let $S \in \mathrm{M}(2, \mathbb{C})$ be a *symmetric* matrix, so $A^{\mathrm{t}} = A$

(a) Is $L_A : \mathbb{C}^2 \to \mathbb{C}^2$ diagonalizable in the ordinary sense?

(b) Is $L_A : \mathbb{C}^2 \to \mathbb{C}^2$ orthogonally diagonalizable when $\mathbb{C}^2$ is given the usual inner product?

Prove or provide a counterexample.
Note: If we take $\mathbb{R}$ instead of $\mathbb{C}$ the answer is "yes" for both (a) and (b) because $A^* = A^{\mathrm{t}}$ when $\mathbb{K} = \mathbb{R}$. Recall that $\left(L_A\right)^* = L_{A^*}$ for the standard inner product on $\mathbb{C}^2$ – see Exercise VI-3.9. Self-adjoint matrices are diagonalizable over both $\mathbb{R}$ and $\mathbb{C}$, but we are not assuming $A = A^*$ here, only $A = A^{\mathrm{t}}$.

6.9 Exercise. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be the operator $T = L_A$ for

$$A = \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$$

Explain why T is self-adjoint with respect to the standard inner product $(z, w) = z_1\overline{w_1} + z_2\overline{w_2}$ on $\mathbb{C}^2$. Then determine

(a) The spectrum $\mathrm{sp}_{\mathbb{C}}(T) = \{\lambda_1, \lambda_2\}$;

(b) The eigenspaces $E_\lambda(T)$ and find an orthonormal basis $\{f_1, f_2\}$ in $\mathbb{C}^2$ that diagonalize T. Then

(c) Find a unitary matrix $U^*U = I$ such that

$$UAU^* = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

where $\mathrm{sp}(T) = \{\lambda_1, \lambda_2\}$.

6.10 Exercise (Uniquess of Spectral Decompositions). Suppose $T : V \to V$ is diagonalizable on an arbitrary vector space (not necesarily an inner product space), so $T = \sum_{i=1}^{r} \lambda_i P_{\lambda_i}$ where $\mathrm{sp}(T) = \{\lambda_1, \ldots, \lambda_r\}$ and P_{λ_i} is the projection onto the λ_i-eigenspace. Now suppose $T = \sum_{j=1}^{s} \mu_j Q_j$ is some other decomposition such that

$$Q_j^2 = Q_j \neq 0 \qquad Q_jQ_k = Q_kQ_j = 0 \quad \text{if } j \neq k \qquad \sum_{j=1}^{s} Q_j = I$$

and $\{\mu_1, \ldots, \mu_r\}$ are distinct. Prove that

(a) $r = s$ and if the μ_j are suitably relabeled we have $\mu_i = \lambda_i$ for $1 \leq i \leq r$.

(b) $Q_i = P_{\lambda_i}$ for $1 \leq i \leq r$.

Hint: First show $\{\mu_1, \ldots, \mu_s\} \subseteq \{\lambda_1, \ldots, \lambda_r\} = \mathrm{sp}(T)$; then relabel.

Here is another useful observation about spectra of diagonalizable operators.

6.11 Lemma (Spectral Mapping Theorem). If $T : V \to V$ is a diagonalizable operator on a finite dimensional vector space, and $f(x)$ is any function $f : \mathrm{sp}(T) \to \mathbb{C}$, then

$$\mathrm{sp}(f(T)) = f\big(\mathrm{sp}(T)\big) = \{f(\lambda) : \lambda \in \mathrm{sp}(T)\} \ .$$

Proof: We have shown that $T = \sum_{\lambda \in \mathrm{sp}(T)} \lambda P_\lambda$ where the P_λ are the spectral projections determined by the direct sum decomposition $V = \bigoplus_\lambda E_\lambda(T)$. Then $f(T) = \sum_\lambda f(\lambda) P_\lambda$, from which it is obvious that $f(T)v = f(\lambda)v$ for $v \in E_\lambda(T)$; hence $f(T)$ is diagonalizable. The eigenvalues are the values $f(\lambda)$ for $\lambda \in \mathrm{sp}(T)$, but notice that we might have $f(\lambda) = f(\mu)$ for different eigenvalues of T. To get the eigenspace $E_\alpha(f(T))$ we must add together all these spaces

$$E_\alpha(f(T)) = \bigoplus\nolimits_{\{\lambda : f(\lambda) = \alpha\}} E_\lambda(T) \qquad \text{for every } \alpha \in f(\mathrm{sp}(T)) \ .$$

The identity is now clear. $\square$

As an extreme illustration, if $f(z) \equiv 1$ then $f(T) = I$ and $\mathrm{sp}(T) = \{1\}$.

VI.7. Positive Operators and the Polar Decomposition.

If $T : V \to W$ but $V \neq W$ one cannot speak of "diagonalizing T." (What would "eigenvector" and "eigenvalue" mean in that context?) But we can still seek other decompositions of T as a product of particularly simple, easily understood operators. Even when $V = W$ one might profitably explore such options if T fails to be diagonalizable – diagonalization is not the only useful decomposition of a linear operator.

When $V = W$ and V is an inner product space over $\mathbb{R}$ or $\mathbb{C}$, all self-adjoint (or normal) operators are orthogonally diagonalizable, and among them the *positive operators* are particularly simple.

7.1 Definition. *A linear operator $T : V \to V$ on an inner product space is* **positive** *if*

$$\text{(i)} \quad T^* = T \qquad \textit{and} \qquad \text{(ii)} \quad (Tv, v) \geq 0 \textit{ for all } v \in V.$$

It is **positive definite** *if $(Tv, v) = 0$ only when $v = 0$. We write $T \geq 0$ or $T > 0$, respectively, to indicate these possibilities. A matrix $A \in \mathrm{M}(n, \mathbb{C})$ is said to be positive (or positive definite) if the multiplication operator $L_A : \mathbb{C}^n \to \mathbb{C}^n$ is positive (positive definite) with respect to the usual inner product, so that $(Av, v) \geq 0$ for all v.*

Note that self-adjoint projections $P^2 = P^* = P$ are examples of positive operators, and sums of positive operators are again positive (but not linear combinations unless the coefficients are posiive).

If T is diagonalizable, $\mathrm{sp}(T) = \{\lambda_1, \ldots, \lambda_r\}$, and if $T = \sum_i \lambda_i P_{\lambda_i}$ is the spectral decomposition, a self-adjoint operator is positive $\Leftrightarrow \lambda_i \geq 0$ for all i, so $\mathrm{sp}(T) \subseteq [0, +\infty) + i0$. [In fact if $T \geq 0$ and $v_i \in E_{\lambda_i}$, we have $(Tv_i, v_i) = \lambda_i \|v_i\|^2 \geq 0$. Conversely if all $\lambda_i \geq 0$ and $v = \sum_{i=1}^r v_i$ we get $(Tv, v) = \sum_{i,j} (Tv_i, v_j)$, and since $E_{\lambda_i} \perp E_{\lambda_j}$ for $i \neq j$ this reduces to $\sum_i (Tv_i, v_i) = \sum_i \lambda_i \|v_i\|^2 \geq 0$.]

If T is positive *definite* then $\lambda_i = 0$ cannot occur in $\mathrm{sp}(T)$ and T is invertible, with

$$T^{-1} = \sum_i \frac{1}{\lambda_i} P_{\lambda_i} \qquad \text{(also a positive definite operator)} .$$

Positive Square Roots. If $T \geq 0$ there is a positive square root (a positive operator $S \geq 0$ such that $S^2 = T$), namely

$$\text{(63)} \qquad \sqrt{T} = \sum_i \sqrt{\lambda_i}\, P_{\lambda_i} \qquad \left(\sqrt{\lambda_i} = \text{the nonnegative square root of } \lambda_i \geq 0\right) ,$$

which is also denoted by $T^{1/2}$. This is a square root because

$$S^2 = \sum_{i,j} \sqrt{\lambda_i}\sqrt{\lambda_j}\, P_{\lambda_i} P_{\lambda_j} = \sum_i \lambda_i P_{\lambda_i} = T$$

where $P_{\lambda_i} P_{\lambda_j} = \delta_{ij} \cdot P_{\lambda_i}$. Notice that the spectral decompositions of T and $\sqrt{T}$ involve the same spectral projections P_{λ_i}; obviously the eigenspaces match up too, because $E_{\lambda_i}(T) = E_{\sqrt{\lambda_i}}(\sqrt{T})$ for all i.

Subject to the requirement that $S \geq 0$, this square root is *unique*, as a consequence of uniqueness of the spectral decomposition on any vector space (see Exercise VI-6.10)

7.2 Exercise. Use uniqueness of spectral decompositions to show that the positive square root operator $\sqrt{T} = \sum_i \sqrt{\lambda_i} P_{\lambda_i}$ defined above is *unique* – i.e. if $A \geq 0$ and $B \geq 0$ and $A^2 = B^2 = T$ for some $T \geq 0$, then $A = B$.

Positivity of $T : V \to V$ has an interesting connection with the exponential map on matrices $\mathrm{Exp} : \mathrm{M}(n, \mathbb{C}) \to \mathrm{M}(n, \mathbb{C})$,

$$\mathrm{Exp}(A) = e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n$$

We indicated in Section V.3 that commuting matrices A, B satisfy the Exponent Law $e^{A+B} = e^A \cdot e^B$, with $e^0 = I$. In particular all matrices in the range of *Exp* are invertible, with $\left(e^A\right)^{-1} = e^{-A}$.

7.3 Exercise. Let $\mathcal{P}$ be the set of *positive definite* matrices A in $\mathrm{M}(n,\mathbb{C})$, which are all self-adjoint by definition of $A > 0$. Let $\mathcal{H}$ be the set of all self-adjoint matrices in $\mathrm{M}(n,\mathbb{C})$, which is a vector subspace *over* $\mathbb{R}$ but not over $\mathbb{C}$ since iA is skew-adjoint if A is self-adjoint. Prove that

(a) The exponential matrix e^H is positive and invertible for self-adjoint matrices H.

(b) The exponential map $\mathrm{Exp} : \mathcal{H} \to \mathcal{P}$ is a *bijection.*

Hint: Explain why $\left(e^A\right)^* = e^{A^*}$ and then use the Exponent Law applied to matrices e^{tA}, $t \in \mathbb{R}$ (you could also invoke the spectral theorem).

It follows that every positive definite matrix $A > 0$ has a unique self-adjoint *logarithm* $\mathrm{Log}(A)$ such that

$$\begin{aligned} \mathrm{Exp}\big(\mathrm{Log}(A)\big) &= A \qquad && \text{for } A \in \mathcal{P} \\ \mathrm{Log}\big(e^H\big) &= H \qquad && \text{for } H \in \mathcal{H}\ \ , \end{aligned}$$

namely the inverse of the bijection $\mathrm{Exp} : \mathcal{H} \to \mathcal{P}$. In terms of spectral decompositions,

$$\mathrm{Log}(T) \text{ of a positive definite } T \text{ is } \mathrm{Log}(T) = \sum_i \mathrm{Log}(\lambda_i) P_{\lambda_i} \text{ if } T = \sum_i \lambda_i P_{\lambda_i}$$
$$\mathrm{Exp}(H) \text{ of a self-adjoint matrix } H \text{ is } e^H = \sum_i e^{\mu_i} Q_{\lambda_i} \text{ if } H = \sum_i \mu_i Q_{\mu_i}$$

When $V = W$ the *unitary* operators $U : V \to V$ are another well-understood family of (diagonalizable) operators on an inner product space. They are particularly interesting and easy to understand because they correspond to the possible choices of orthonormal bases in V. Every unitary U is obtained by specifying a pair of orthonormal bases $\mathfrak{X} = \{e_i\}$ and $\mathfrak{Y} = \{f_j\}$ and defining U to be the unique linear map such that

$$U\Big(\sum_{i=1}^n c_i e_i\Big) = \sum_{j=1}^n c_j f_j \qquad (\text{arbitrary } c_i \in \mathbb{C})$$

Polar Decompositions. The positive operators $P \geq 0$ and unitary operators U on an inner product space provide a natural *polar decomposition* $T = U \cdot P$ of any linear operator $T : V \to V$. In its simplest form (when T is invertible) it asserts that any *invertible* map T has a *unique* factorization

$$T = U \cdot P \qquad \begin{cases} U : V \to V & \text{unitary (a bijective isometry of } V) \\ P : V \to V & \text{positive definite, invertible} = e^H \text{ with } H \text{ self-adjoint} \end{cases}$$

Both factors are orthogonally diagonalizable (U because it is normal and P because it is self-adjoint), but the original operator T need not itself be diagonalizable over $\mathbb{C}$, let alone orthogonally diagonalizable.

We will develop the polar decomposition first for an invertible operator $T : V \to V$ since that proof is particularly transparent. We then address the general result (often referred to as the *singular value decomposition* when it is stated for matrices). This involves operators that are not necessarily invertible, and may be maps $T : V \to W$ between quite different inner product spaces. The positive component $P : V \to V$ is still unique but the unitary component U may be nonunique (in a harmless sort of way). The "*singular values*" of T are the eigenvalues $\lambda_i \geq 0$ of the positive component P.

7.4 Theorem (Polar Decomposition I). *Let V be a finite dimensional inner product* *space over $\mathbb{C}$. Every invertible operator $T : V \to V$ has a unique decomposition $T = U{\cdot}P$ where*

$$U \in \mathrm{U}(n) = (\textit{the group of unitary operators } U^* = U^{-1})$$

$$P \in \mathcal{P} = (\textit{invertible positive definite operators } P > 0)$$

By Exercise 7.3 we can also write $T = U \cdot e^H$ for a unique self-adjoint operator $A \in \mathcal{H}$.

This is the linear operator (or matrix) analog of the polar decomposition

$$z = |z|e^{i\theta} = r \cdot e^{i\theta} \qquad \text{with } r > 0 \text{ and } \theta \text{ real (so } |e^{i\theta}| = 1)$$

for nonzero complex numbers. If we think of "positive definite" = "positive," "self-adjoint" as "real," and "unitary" = "absolute value 1," the analogy with the polar decomposition $z = re^{i\theta}$ of a nonzero complex number z is clear.

Some Preliminary Remarks. If $T : V \to W$ is a linear map between two inner product spaces, its *absolute value* $|T|$ is the linear map from $V \to V$ determined in the following way.

The product T^*T maps $V \to V$ and is a positive operator because

$$(T^*T)^* = T^*T^{**} = T^*T \qquad (T^* : W \to V \text{ and } T^{**} = T \text{ on } V)$$

$$\big(T^*Tv, v\big) = (Tv, Tv) = \|Tv\|^2 \geq 0 \quad \text{ for } v \in V$$

Thus T^*T is self-adjoint and has a spectral decomposition $T^*T = \sum_i \lambda_i P_{\lambda_i}$, with eigenvalues $\lambda_i \geq 0$ and self-adjoint projections $P_{\lambda_i} : V \to E_{\lambda_i}(T^*T)$ onto orthogonal subspaces. The **absolute value** $|T| : V \to V$ is then defined as the unique positive square root

$$|T| = \big(T^*T\big)^{1/2} = \sum_i \sqrt{\lambda_i}\, P_{\lambda_i} \ ,$$

whose spectral decomposition involves the same projections that appeared in T^*T. For any linear operator $T : V \to W$ we have $T^*T = |T|^2$ and hence

$$\|Tv\|_W^2 = (T^*Tv, v)_V = \big(|T|^2 v, v\big)_V = \big\|\, |T|(v)\, \big\|_V^2 \qquad \text{for all } v \in V \ . \tag{64}$$

Thus $|T|(v) \in V$ and $Tv \in W$ have the same norm for every $v \in V$. It follows from (64) that T, T^*T, and $|T|$ have the same kernel because

$$\begin{aligned} Tv = 0 \Rightarrow T^*T(v) = 0 \quad &\Rightarrow \quad (T^*Tv, v) = \big(|T|^2(v), v\big) = \big\|\, |T|(v)\, \big\|^2 = 0 \\ &\Rightarrow \quad |T|(v) = 0 \Rightarrow Tv = 0 \quad \text{(by (64))} \ , \end{aligned}$$

Thus the kernels coincide

$$K(T) = K(T^*T) = K(|T|) \tag{65}$$

even if the ranges differ, and one of these operators is invertible if and only if they all are. In particular $|T|$ is positive *definite* on V ($|T| > 0$) if and only if $T : V \to W$ is invertible. (Comparisons between T and $|T|$ do not follow from spectral theory because T itself need not be diagonalizable, even if $V = W$.)

Proof of VI-7.4: The proof in the invertible case is simple. For any linear operator $T : V \to V$ we have $T^*T = |T|^2$ and have seen in (64) that $|T|(v)$ and Tv always have the same norm. When T is invertible, so is $|T|$ and we have $R(T) = R(|T|) = V$. The identities (64) determine a bijective isometry $U : V \to V$ that sends $T(v) \mapsto |T|(v)$

for all v, as indicated in Figure 6.10. This map is also linear because $U = T \circ |T|^{-1}$ is a composite of linear operators on V. Thus when T is invertible the desired polar decomposition is

$$U \circ |T| = \left(T \circ |T|^{-1}\right) \circ |T| = T$$

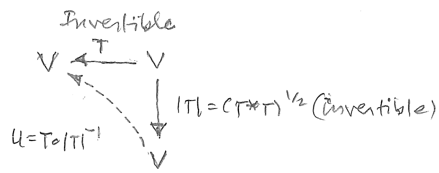

Figure 6.10. The maps involved in defining $|T| : V \to V$ for an invertible map $T : V \to W$ between two inner product spaces. In the discussion we show that the positive operator $|T| = (T^*T)^{1/2}$ is invertible and $R(T^*T) = R\big(|T|\big) = R(T) = V$ when T is invertible. The induced bijection $U = T \circ |T|^{-1} : V \to V$ is a bijective linear isometry (a unitary map of $V \to V$) and the polar decomposition of T is $U \cdot |T|$.

As for uniqueness (valid only in the invertible case), suppose $T = UP = U_0P_0$ with U, U_0 unitary and P, P_0 positive definite. Then $P^* = P, P_0^* = P_0$, and $T^* = P^*U^* = PU^* = P_0^*U_0^* = P_0U_0^*$ since the positive components are self-adjoint; hence

$$P^2 = PU^*UP = PU^*\big(PU^*\big)^* = P_0U_0^*U_0P_0 = P_0^2$$

Now $P^2 = P^*P$ is a positive operator which has a unique positive square root, namely P; likewise for P_0^2. By uniqueness we get $P_0 = P$, from which $U_0 = U$ folllows. $\square$

Computing U for Invertible $T : V \to V$. Determining the positive part $P = |T|$ is straightforward: $P^2 = T^*T$ is self-adjoint and its spectral decomposition can be computed in the usual way. If $\{e_i\}$ is an orthonormal basis of eigenvectors for T^*T, which are also eigenvectors for $P = |T|$, we have

$$T^*T(e_i) = \lambda_i e_i \qquad \text{and} \qquad |T|(e_i) = \sqrt{\lambda_i}\, e_i \tag{66}$$

(with all $\lambda_i > 0$ because $|T|$ is invertible $\Leftrightarrow$ T is invertible $\Leftrightarrow$ all $\lambda_i \neq 0$). From this we get

$$\begin{array}{rcl} |T|^{-1}(e_i) & = & |T|^{-1}\left(\dfrac{1}{\sqrt{\lambda_i}}\,|T|(e_i)\right) = \dfrac{1}{\sqrt{\lambda_i}}\, e_i \\ & \Downarrow & \\ U(e_i) & = & T\Big(|T|^{-1}e_i\Big) \;=\; \dfrac{1}{\sqrt{\lambda_i}}\, T(e_i) \end{array}$$

By its construction U is unitary on V so the vectors

$$f_i = \frac{1}{\sqrt{\lambda_i}}\, T(e_i)$$

are a new orthonormal basis in V. This completely determines U. $\square$

Note that

$$\|U(e_i)\| = \frac{1}{\sqrt{\lambda_i}}\, \|T(e_i)\| = \frac{1}{\sqrt{\lambda_i}}\, \big\|\, |T|(e_i)\, \big\| = \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_i}}\, \|e_i\| = 1$$

as expected.

The General Polar Decomposition. When $T : V \to V$ is not invertible the polar decomposition is somewhat more complicated. The positive component in $T = U{\cdot}P$ is still the unique positive square root $P = |T| = (T^*T)^{1/2}$. But the unitary part is based on a uniquely determined isometry $U_0 : R(|T|) \to R(T)$ between proper subspaces in V that can have various extensions to a unitary map $U : V \to V$. This ambiguity has no effect on the factorization $T = U \cdot P$; the behavior of U off of $R(|T|)$ is completely irrelevant.

7.5 Theorem (Polar Decomposition II). *Any linear operator $T : V \to V$ on a finite dimensional complex inner product space has a factorization $T = U\,|T|$ where*

1. *$|T|$ is the positive square root of T*
2. *U is a unitary operator on V.*

The positive factor is uniquely determined only on the range $R(|T|)$, which is all that matters in the decomposition $R = U\,|T|$, but it can be extended in various ways to a unitary map $V \to V$ when T is not invertible.

Proof: First note that

$$R(|T|) = K(|T|)^{\perp} = K(T)^{\perp} = K(T^*T)^{\perp} = R(T^*T)$$

$$R(|T|)^{\perp} = K(|T|) = K(T) = K(T^*T) = R(T^*T)^{\perp}$$

The subspaces in the first row are just the orthocomplements of those in the second. The first and last identities in Row 2 hold because $|T|$ and T^*T are self-adjoint (Proposition VI-4.2); the rest have been proved in (65). We now observe that equation (64)

$$\|Tv\|^2 = (T^*Tv, v) = \big(|T|^2 v, v\big) = \big\|\,|T|(v)\,\big\|^2 \qquad \text{for all } v \in V\ ,$$

implies that there is a norm-preserving bijection U_0 from $R(|T|) \to R(T)$, defined by letting

$$U_0\big(|T|(v)\big) = T(v)\ . \tag{67}$$

This makes sense despite its seeming ambiguity: If an element $y \in R(|T|)$ has realizations $y = |T|(v') = |T|(v)$ we get $|T|(v' - v) = 0$, and then

$$T(v' - v) = T(v') - T(v) = 0$$

because $|T|(v' - v)$ and $T(v' - v)$ have equal norms. Thus $T(v') = T(v)$ and the operator (67) is in fact a well-defined bijective map from $R(|T|)$ into $R(T)$. It is linear because

$$\begin{aligned} U_0\big(|T|v_1 + |T|v_2\big) \;&=\; U_0\big(|T|(v_1 + v_2)\big) \;=\; T(v_1 + v_2) \\ &=\; Tv_1 + Tv_2 \;=\; U_0(|T|v_1) + U_0(|T|v_2) \end{aligned}$$

It is then immediate that $\|U_0(y)\| = \|y\|$ for all $y \in R(|T|)$, and $R(U_0) \subseteq R(T)$. But $K(T) = K(|T|)$ so $\dim R(T) = \dim R(|T|)$; hence $\dim R(U_0) = \dim R(|T|)$ because U_0 is an isometry. We conclude that $R(U_0) = R(T)$ and $U_0 : R(|T|) \to R(T)$ is a bijective isometry between subspaces of equal dimension. By definition we get

$$T(v) = \big(U_0 \cdot |T|\big)(v) \qquad \text{for all } v \in V\ .$$

We can extend U to a globally defined unitary map $U : V \to V$ because $K(T) = K(|T|) \Rightarrow \dim R(T) = \dim R(|T|)$ and $\dim R(T)^{\perp} = \dim R(|T|)^{\perp}$; therefore there exist various isometries

$$U_1 : R(|T|)^{\perp} \to R(T)^{\perp}\ .$$

corresponding to orthonormal bases in these subspaces. Using the orthogonal decompositions

$$V = R(|T|) \dot{\oplus} R(|T|)^{\perp} = R(T) \dot{\oplus} R(T)^{\perp}$$

we obtain a bijective map

$$U(v, v') = \big(U_0(v), U_1(v')\big)$$

such that $U|T| = U_0|T| = T$ on all of V. □

There is a similar decomposition for operators $T : V \to W$ between different inner products spaces; we merely sketch the proof. Once again we define the positive component $|T| = (T^*T)^{1/2}$ as in (63). The identity

$$\big\| \, |T|(v) \, \big\|_V^2 = \| \, T(v) \, \|_W^2 \qquad \text{for all } v \in V$$

holds exactly as in (64), and this induces a linear isometry U_0 from $M = R(|T|) \subseteq V$ to $N = R(T) \subseteq W$ such that

$$T = U_0 \cdot |T| = \Big[T \cdot \big(|T|_M\big)^{-1}\Big] \cdot |T|$$

where $|T|_M =$ (restriction of $|T|$ to M).

The fact that U_0 is only defined on $R(|T|)$ is irrelevant, as it was in Theorem 7.5, but now U_0 cannot be extended unitary map (bijective isometry) from V to W unless $\dim(V) = \dim(W)$. On the other hand since $|T|$ is self-adjoint we have

$$R(|T|) = K(|T|)^{\perp} = K(T)^{\perp}$$

and can define $U \equiv 0$ on $K(T)$ to get a globally defined "*partial isometry*" $U : V \to W$

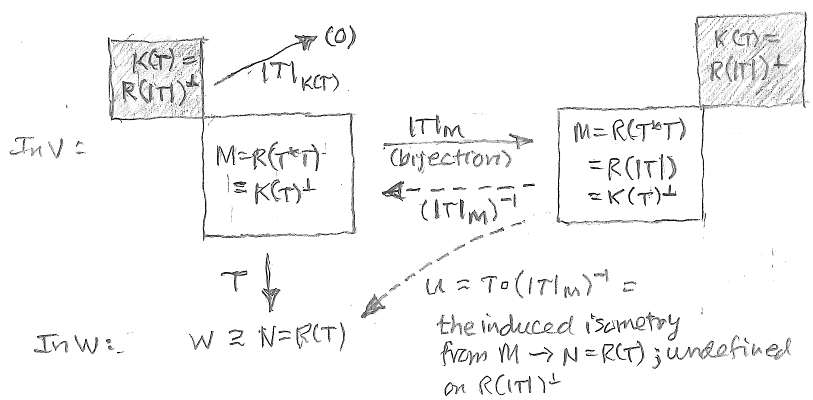

Figure 6.11. The maps involved in defining a polar decomposition $T = U_0 \cdot |T|$ for an arbitrary linear map $T : V \to W$ between different inner product spaces. Here we abbreviate $M = K(T)^{\perp} \subseteq V$ and $N = R(T) \subseteq W$; $U_0 : M \to N$ is an induced isometry such that $T = U_0 \cdot |T|$.

such that $K(U) = K(T)$, $R(U) = R(U_0) = R(T)$, and

$$U|_{K(T)} = 0 \qquad U|_{K(T)^{\perp}} = U|_{R(|T|)} = U_0$$

The players involved are shown in the commutative diagram Figure 6.11.

The **singular value decomposition** is a useful variant of Theorem 7.5.

7.6 Theorem (Singular Value Decomposition). *Let $T : V \to W$ be a linear operator between complex inner product spaces. There exist nonnegative scalars*

$$\lambda_1 \geq \ldots \geq \lambda_r \geq 0 \qquad (r = \mathrm{rank}(T))$$

and orthonormal bases $\{e_1, \ldots, e_r\}$ for $K(T)^\perp \subseteq V$ and $\{f_1, \ldots, f_r\}$ for $R(T) \subseteq W$ such that

$$T(e_i) = \lambda_i f_i \text{ for } 1 \leq i \leq r \qquad \text{and} \qquad T \equiv 0 \text{ on } K(T) = K(T)^{\perp\perp}$$

The λ_i are the eigenvalues of $|T| = (T^*T)^{1/2}$ counted according to their multiplicities.