Linear Algebra 2 - Notes

Reference: Linear Algebra Done Right - Sheldon Axler

Lim Zi Xiang

Contents

Chapter 6

Inner-product spaces

In this chapter we discuss the notion of length and angle, which are embedded in the concept of inner products. We let $\mathbb{F}$ denote $\mathbb{R}$ or $\mathbb{C}$, V be a finite-dimensional, nonzero vector space over $\mathbb{F}$.

6.1 Inner products

Definition 6.1.1: Norm and dot product

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ with $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots y_n)$. Then the **norm** of $\boldsymbol{x}$, $\|\boldsymbol{x}\|$ is given by

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + \ldots + x_n^2},$$

and the **dot product** of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{x} \cdot \boldsymbol{y}$ is given by

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + \ldots + x_n y_n.$$

Note:-

- $\boldsymbol{x} \cdot \boldsymbol{x} = \|\boldsymbol{x}\|^2$;
- $\boldsymbol{x} \cdot \boldsymbol{x} \geq 0 \, \forall \boldsymbol{x} \in \mathbb{R}^n$, with equality iif $\boldsymbol{x} = \boldsymbol{0}$;
- $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}; \phi(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{y}$ is linear;
- $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$.

Definition 6.1.2: Modulus, complex conjugate, complex norm

Let $\lambda = a + bi$ where $a, b \in \mathbb{R}$, and $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{C}^{\mathrm{n}}$, then

1. the **absolute value** of λ, $|\lambda| = \sqrt{a^2 + b^2}$;
2. the **complex conjugate** of λ, $\bar{\lambda} = a - bi$;
3. the **norm** of $\boldsymbol{z}$, $\|\boldsymbol{z}\| = \sqrt{|z_1|^2 + \ldots + |z_n|^2}$.

Using this definition, we have

$$\|\boldsymbol{z}\|^2 = z_1 \overline{z_1} + \ldots + z_n \overline{z_n},$$

which we want to think of as the inner product of $\boldsymbol{z}$ with itself like with $\mathbb{R}^n$. So perhaps the inner product of $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) \in \mathbb{C}^n$ with $\boldsymbol{z}$ should be

$$\begin{aligned}
\boldsymbol{w}_1\overline{\boldsymbol{z}_1} + \ldots + \boldsymbol{w}_n\overline{\boldsymbol{z}_n} &= \sum_{j=1}^{n}(a_j + b_j i)(c_j - d_j i) \\
&= \sum_{j=1}^{n}(a_j c_j + b_j d_j + i(b_j c_j - a_j d_j)) \\
&= \sum_{j=1}^{n}(a_j c_j + b_j d_j - i(a_j d_j - b_j c_j)) \\
&= \sum_{j=1}^{n}(a_j c_j + b_j d_j) - i\sum_{j=1}^{n}(a_j d_j - b_j c_j) \\
&= \overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle}
\end{aligned}$$

where each $\boldsymbol{w}_j = a_j + b_j i$ and $\boldsymbol{z}_j = c_j + d_j i$ with each $a_j, b_j, c_j, d_j \in \mathbb{R}$. We now formally define the inner product on V.

Definition 6.1.3: Inner product

An **inner product** on V is a function that takes each ordered pair $(\boldsymbol{u}, \boldsymbol{v})$ of elements of V toa number $\langle \boldsymbol{u}, \boldsymbol{v} \rangle \in \mathbb{F}$ and has the following properties:

1. Positivity: $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0 \, \forall \boldsymbol{v} \in V$;
2. Definiteness: $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0 \Leftrightarrow \boldsymbol{v} = 0$;
3. Additivity in first slot: $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle \, \forall \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$;
4. Homogeneity in first slot: $\langle a\boldsymbol{u}, \boldsymbol{w} \rangle = a\langle \boldsymbol{u}, \boldsymbol{w} \rangle \forall a \in \boldsymbol{F} \forall \boldsymbol{v}, \boldsymbol{w} \in V$;
5. Conjugate symmetry: $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \overline{\langle \boldsymbol{w}, \boldsymbol{v} \rangle} \, \forall \boldsymbol{v}, \boldsymbol{w} \in V$.

Note:-

The complex conjugate of every real number is itself. If $\boldsymbol{z} \in \mathbb{C}$, $\boldsymbol{z} \geq 0 \implies \boldsymbol{z}$ is real and nonnegative.

Definition 6.1.4: Inner-product space

An **inner-product space** is a vector space V along with an inner product on V.

Example 6.1.1 (The Euclidean inner product on $\mathbb{F}^n$)

We define the **Euclidean inner product** on $\mathbb{F}^n$ by

$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = \boldsymbol{w}_1\overline{z_1} + \boldsymbol{w}_n\overline{z_n}$$

We show that this definition satisfies our four conditions: Let $\boldsymbol{z} = (z_1, \ldots, z_n) \in \mathbb{F}^n$ where $z_1, \ldots, z_n \in \mathbb{F}$. Since $\mathbb{R} \subset \mathbb{C}$, $\boldsymbol{z} \in \mathbb{C}^n$, we have

1. $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = z_1\overline{z_1} + \ldots + z_n\overline{z_n} = |z_1|^2 + \ldots + |z_n|^2 \geq 0$;

2. $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = 0$ iif $\boldsymbol{z} = \boldsymbol{0}$ since each $|z_i|^2 \geq 0$ since $|z_i| \in \mathbb{R}$;

3. $\langle \boldsymbol{v}+\boldsymbol{w}, \boldsymbol{z} \rangle = (v_1+w_1)\overline{z_1}+\ldots+(v_n+w_n)\overline{z_n} = (v_1\overline{z_1}+\ldots+v_n\overline{z_n})+(w_1\overline{z_1}+\ldots+w_n\overline{z_n}) = \langle \boldsymbol{v}, \boldsymbol{z} \rangle+\langle \boldsymbol{w}, \boldsymbol{z} \rangle$;

4. $\langle k\boldsymbol{w}, \boldsymbol{z} \rangle = (kw_1)\overline{z_1} + \ldots + (kw_n)\overline{z_n} = k(w_1\overline{z_1} + \ldots + w_n\overline{z_n}) = k\langle \boldsymbol{w}, \boldsymbol{z} \rangle$

5. Shown previously;

where $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{F}^n$ with $\boldsymbol{v} = (v_1, \ldots, v_n)$ and $\boldsymbol{w} = (w_1, \ldots, w_n)$, $k \in \mathbb{F}$. We assume by default when $\mathbb{F}^n$ is referred to as an inner-product space, the inner product is the Euclidean inner product.

Example 6.1.2 (Other inner products)

For $c_1, \ldots, c_n \in \mathbb{R}_{>0}$,

$$\langle (w_1, \ldots, w_n), (z_1, \ldots, z_n) \rangle = c_1 w_1 \overline{z_1} + \ldots + c_n w_n \overline{z_n}$$

is an inner product on $\mathbb{F}^n$ which we can show satisfies the five conditions as well. Another example is an inner product on $\mathcal{P}_m(\mathbb{F})$, the vector space of all polynomials with coefficients in $\mathbb{F}$ and degree at most m:

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_0^1 \boldsymbol{p}(x)\overline{\boldsymbol{q}(x)}dx.$$

Note:-

Personal observations:

- $\langle \boldsymbol{w}, \boldsymbol{z} \rangle = 0$ is not an inner product on $\mathbb{F}^n$ because it violates the second condition.
- $\langle \boldsymbol{w}, \boldsymbol{z} \rangle = \boldsymbol{w} + \boldsymbol{z}$ is not an inner product on $\mathbb{F}^n$ because it only satisfies the second condition.
- $\langle (z_1, z_2, z_3), (w_1, w_2, w_3) \rangle = z_1 w_2 + z_2 w_3 + z_3 w_1$ is not an inner product on $\mathbb{F}^3$ because it violates the first, second, and fifth condition.

For now onwards, we let V be a finite-dimensional inner-product space over $\mathbb{F}$.

We can combine the conditions for additivity and homogeneity in the first slot into a requirement for linearity in the first slot, so for fixed $\boldsymbol{w} \in V$, the function mapping $\boldsymbol{v}$ to $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$ is a linear map from V to $\mathbb{F}$ and so we have

$$\langle 0, \boldsymbol{w} \rangle = 0, \qquad \langle \boldsymbol{w}, 0 \rangle = 0,$$

$\forall \boldsymbol{w} \in V$ with second equality due to the conjugate symmetry property.

Proposition 6.1.1

In an inner-product space, we have additivity and conjugate homogeneity in the second slot, i.e.

$$\langle \boldsymbol{u}, \boldsymbol{v} + \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{u}, \boldsymbol{w} \rangle, \qquad \langle \boldsymbol{u}, k\boldsymbol{v} \rangle = \overline{k}\langle \boldsymbol{u}, \boldsymbol{v} \rangle,$$

where $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $k \in \mathbb{F}$.

Proof: 1.

$$\begin{aligned}
\langle \boldsymbol{u}, \boldsymbol{v}+\boldsymbol{w} \rangle &= \overline{\langle \boldsymbol{v}+\boldsymbol{w}, \boldsymbol{u} \rangle} \\
&= \overline{\langle \boldsymbol{v}, \boldsymbol{u} \rangle + \langle \boldsymbol{w}, \boldsymbol{u} \rangle} \\
&= \overline{\langle \boldsymbol{v}, \boldsymbol{u} \rangle} + \overline{\langle \boldsymbol{w}, \boldsymbol{u} \rangle} \\
&= \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{u}, \boldsymbol{w} \rangle
\end{aligned}$$

2.

$$\begin{aligned}
\langle \boldsymbol{u}, k\boldsymbol{v} \rangle &= \overline{\langle k\boldsymbol{v}, \boldsymbol{u} \rangle} \\
&= \overline{k\langle \boldsymbol{v}, \boldsymbol{u} \rangle} \\
&= \overline{k\langle \boldsymbol{v}, \boldsymbol{u} \rangle} \\
&= \overline{k}\langle \boldsymbol{u}, \boldsymbol{v} \rangle
\end{aligned}$$

□

6.2 Norms

Definition 6.2.1: Norm

For $\boldsymbol{v} \in V$, the **norm** of $\boldsymbol{v}$, $\|\boldsymbol{v}\|$ is given by

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

Example 6.2.1 (Norms)

1. For $(z_1, \ldots, z_n) \in \mathbb{F}^n$ with the Euclidean inner product,

$$\begin{aligned}
\|(z_1, \ldots, z_n)\| &= \sqrt{z_1\overline{z_1} + \ldots + z_n\overline{z_n}} \\
&= \sqrt{|z_1|^2 + \ldots + |z_n|^2}
\end{aligned}$$

2. For $\boldsymbol{p} \in \mathcal{P}_m(\mathbb{F})$ with inner product given in Example 6.1.2,

$$\|\boldsymbol{p}\| = \sqrt{\int_0^1 |p(x)|^2 dx}$$

Proposition 6.2.1

Let $\boldsymbol{v} \in V$ and $a \in \mathbb{F}$, then

1. $\|\boldsymbol{v}\| = 0$ iif $\boldsymbol{v} = \boldsymbol{0}$;
2. $\|a\boldsymbol{v}\| = |a|\|\boldsymbol{v}\|$.

Proof: 1. Suppose that $\|\boldsymbol{v}\| = 0$, so by definition, $\sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle} = 0 \Rightarrow \langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0 \Leftrightarrow \boldsymbol{v} = 0$ by definiteness of inner products.

2.

$$
\begin{aligned}
\|a\boldsymbol{v}\|^2 &= \langle a\boldsymbol{v}, a\boldsymbol{v}\rangle \\
&= a\langle \boldsymbol{v}, a\boldsymbol{v}\rangle \text{ by homogeneity in first slot;} \\
&= a\bar{a}\langle \boldsymbol{v}, \boldsymbol{v}\rangle \text{ by conjugate homogeneity in second slot;} \\
&= |a|^2\|\boldsymbol{v}\|^2 \\
\|a\boldsymbol{v}\| &= |a|\|\boldsymbol{v}\|
\end{aligned}
$$

□

Note:-

It is usually easier to work with norms squared rather than norms directly.

Definition 6.2.2: Orthogonal

Let $\boldsymbol{u}, \boldsymbol{v} \in V$. $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be **orthogonal** if $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$.

Note that $\mathbf{0}$ is orthogonal to every vector and is the only vector orthogonal to itself. In the following theorem we look at the special case where $V = \mathbb{R}^2$:

Theorem 6.2.1 Pythagorean Theorem

Let $\boldsymbol{u}, \boldsymbol{v}$ be orthogonal vectors in $\mathbb{R}^2$, then

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2.$$

Proof:

$$
\begin{aligned}
\|\boldsymbol{u} + \boldsymbol{v}\|^2 &= \langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v}\rangle \\
&= \langle \boldsymbol{u}, \boldsymbol{u} + \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v}\rangle \text{ by additivity in first slot} \\
&= \langle \boldsymbol{u}, \boldsymbol{u}\rangle + \langle \boldsymbol{u}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u}\rangle + \langle \boldsymbol{v}, \boldsymbol{v}\rangle \text{ by additivity in second slot} \\
&= \|\boldsymbol{u}\|^2 + 0 + 0 + \|\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2
\end{aligned}
$$

□

Note:-

The converse of the Pythagorean theorem holds if $\langle \boldsymbol{u}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u}\rangle = 0$ which is true in real inner-product spaces.

6.2.1 Orthogonal decomposition

Suppose $\boldsymbol{u}, \boldsymbol{v} \in V$. We want to write $\boldsymbol{u}$ as the some of a scalar multiple of $\boldsymbol{v}$ and a vector $\boldsymbol{w} \in V$ orthogonal to $\boldsymbol{v}$, i.e.

$$\boldsymbol{u} = k\boldsymbol{v} + \boldsymbol{w} \Rightarrow \boldsymbol{w} = \boldsymbol{u} - k\boldsymbol{v}$$

where $k \in \mathbb{F}$. Since $\boldsymbol{w}$ is orthogonal to $\boldsymbol{v}$, we must have

$$\langle \boldsymbol{u} - k\boldsymbol{v}, \boldsymbol{v}\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle - k\|\boldsymbol{v}\|^2 = 0 \Rightarrow k = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|^2}.$$

Hence we can now write (assuming $\boldsymbol{v} \neq \mathbf{0}$)

$$\boldsymbol{u} = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|^2}\boldsymbol{v} + \left(\boldsymbol{u} - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|^2}\boldsymbol{v}\right).$$

6.2.2 Important results

Theorem 6.2.2 Cauchy-Schwarz inequality

If $\boldsymbol{u}, \boldsymbol{v} \in V$, then

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|,$$

with equality iif $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent.

Proof: Suppose $\boldsymbol{u}, \boldsymbol{v} \in V$, if $\boldsymbol{v} = \mathbf{0}$ then the inequality holds. If $\boldsymbol{v} \neq \mathbf{0}$, consider the orthogonal decomposition

$$\boldsymbol{u} = \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|^2} + \boldsymbol{w}.$$

By the Pythagorean theorem, we then have

$$\begin{aligned}
\|\boldsymbol{u}\|^2 &= \left\| \frac{\langle \boldsymbol{u}, \boldsymbol{v} \rangle}{\|\boldsymbol{v}\|^2} \boldsymbol{v} \right\|^2 + \|\boldsymbol{w}\|^2 \\
&= \frac{|\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2}{\|\boldsymbol{v}\|^4} \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \frac{|\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2}{\|\boldsymbol{v}\|^2} + \|\boldsymbol{w}\|^2 \\
\|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2 &= |\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2 + \|\boldsymbol{v}\|^2 \|\boldsymbol{w}\|^2 \geq |\langle \boldsymbol{u}, \boldsymbol{v} \rangle|^2 \\
\|\boldsymbol{u}\|\|\boldsymbol{v}\| &\geq |\langle \boldsymbol{u}, \boldsymbol{v} \rangle|;
\end{aligned}$$

with equality iif $\|\boldsymbol{w}\|^2 = 0$ iif $\boldsymbol{u}$ and $\boldsymbol{v}$ is linearly dependent. □

Theorem 6.2.3 Triangle inequality

If $\boldsymbol{u}, \boldsymbol{v} \in V$, then

$$\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|.$$

with equality iif $\boldsymbol{u}$ and $\boldsymbol{v}$ are nonnegative multiples of each other.

Proof: Let $\boldsymbol{u}, \boldsymbol{v} \in V$, then

$$\begin{aligned}
\|\boldsymbol{u} + \boldsymbol{v}\|^2 &= \langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v} \rangle \\
&= \langle \boldsymbol{u}, \boldsymbol{u} \rangle + \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{v}, \boldsymbol{u} \rangle + \langle \boldsymbol{v}, \boldsymbol{v} \rangle \\
&= \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \overline{\langle \boldsymbol{u}, \boldsymbol{v} \rangle} \\
&= \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2 \operatorname{Re}\langle \boldsymbol{u}, \boldsymbol{v} \rangle \\
&\leq \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \\
&\leq \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 + 2\|\boldsymbol{u}\|\|\boldsymbol{v}\| \text{ by Cauchy-Schwarz} \\
&= (\|\boldsymbol{u}\| + \|\boldsymbol{v}\|)^2 \\
\|\boldsymbol{u} + \boldsymbol{v}\| &\leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|
\end{aligned}$$

We have equality iif $\operatorname{Re}\langle \boldsymbol{u}, \boldsymbol{v} \rangle = |\langle \boldsymbol{u}, \boldsymbol{v} \rangle|$ and if we have equality in the Cauchy-Schwarz inequality, both of which implies that $\boldsymbol{u}$ and $\boldsymbol{v}$ are scalar multiples of each other. Let $\boldsymbol{v} = k\boldsymbol{u}$ for some scalar $k \in \mathbb{F}$, then

$$\begin{aligned}
\operatorname{Re}\langle \boldsymbol{u}, k\boldsymbol{u} \rangle &= \|\boldsymbol{u}\|\|k\boldsymbol{u}\| \\
\operatorname{Re}(k\|\boldsymbol{u}\|^2) &= \bar{k}\|\boldsymbol{u}\|^2 \\
k\|\boldsymbol{u}\|^2 &= |k|\|\boldsymbol{u}\|^2 \\
k &= |k|;
\end{aligned}$$

so k is nonnegative assuming $\boldsymbol{u} \neq \mathbf{0}$. □

Theorem 6.2.4 Parallelogram equality

If $\boldsymbol{u}, \boldsymbol{v} \in V$, then

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 + \|\boldsymbol{u}-\boldsymbol{v}\|^2 = 2(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2).$$

Proof: Let $\boldsymbol{u}, \boldsymbol{v} \in V$, then

$$\begin{aligned}
\|\boldsymbol{u}+\boldsymbol{v}\|^2 + \|\boldsymbol{u}-\boldsymbol{v}\|^2 &= \langle \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{u}+\boldsymbol{v}\rangle + \langle \boldsymbol{u}-\boldsymbol{v}, \boldsymbol{u}-\boldsymbol{v}\rangle \\
&= \|\boldsymbol{u}\|^2 + \langle \boldsymbol{u}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{u}\rangle + \|\boldsymbol{v}\|^2 + \|\boldsymbol{u}\|^2 - \langle \boldsymbol{u}, \boldsymbol{v}\rangle - \langle \boldsymbol{v}, \boldsymbol{u}\rangle + \|\boldsymbol{v}\|^2 \\
&= 2(\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2).
\end{aligned}$$

□

6.3 Orthonormal bases

Definition 6.3.1: Orthonormal list

A list of vectors is **orthonormal** if the vectors in the list are pairwise orthorgonal and have a norm of 1, i.e. a list $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ of vectors in V is orthonormal if $\langle \boldsymbol{e}_k, \boldsymbol{e}_k\rangle$ equals 0 for $j \neq k$ and equals 1 for $j = k$.

Proposition 6.3.1

If $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ is an orthonormal list of vectors in V, then

$$\|a_1\boldsymbol{e}_1 + \ldots + a_m\boldsymbol{e}_m\|^2 = |a_1|^2 + \ldots + |a_m|^2$$

$\forall a_1, \ldots, a_m \in \mathbb{F}$.

Proof: By the Pythagorean theorem,

$$\|a_1\boldsymbol{e}_1 + \ldots + a_m\boldsymbol{e}_m\|^2 = \|a_1\boldsymbol{e}_1\|^2 + \ldots + \|a_m\boldsymbol{e}_m\|^2 = |a_1|^2 + \ldots + |a_m|^2.$$

□

Corollary 6.3.1

Every orthonormal list of vectors is linearly independent.

Proof: Suppose $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ is an orthonormal list of vectors in V and $a_1, \ldots, a_m \in \mathbb{F}$ such that

$$a_1\boldsymbol{e}_1 + \ldots + a_m\boldsymbol{e}_m = 0.$$

Then by taking the squared norm of both sides, by Proposition 6.3.1, we have $|a_1|^2 + \ldots + |a_m|^2 = 0 \Rightarrow a_1 = \ldots = a_m = 0$ since the modulus of each a_j is nonnegative. □

Definition 6.3.2: Orthonormal basis

An **orthonormal basis** of V is an orthonormal list of vectors in V which is also a basis of V.

One example of such a basis is the standard basis of $\mathbb{F}^n$. By the previous corollary, any orthonormal list of vectors in V with length $\dim V$ must be an orthonormal basis of V. The importance of orthonormal bases stems mainly from the following theorem.

Theorem 6.3.1

Suppose $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is an orthonormal basis of V, then $\forall \boldsymbol{v} \in V$,

$$\boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \ldots + \langle \boldsymbol{v}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n,$$

and

$$\|\boldsymbol{v}\|^2 = |\langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle|^2 + \ldots + |\langle \boldsymbol{v}, \boldsymbol{e}_n \rangle|^2.$$

Proof: Let $\boldsymbol{v} \in V$. Because $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is a basis of V, $\exists a_1, \ldots, a_n$ such that

$$\boldsymbol{v} = a_1 \boldsymbol{e}_1 + \ldots + a_n \boldsymbol{e}_n.$$

By taking the inner product of both sides with $\boldsymbol{e}_j$, we obtain $\langle \boldsymbol{v}, \boldsymbol{e}_j \rangle = a_j$ since $\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = 0$ for $i \neq j$ and $\langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = 1$ for $i = j$. The second equality holds from applying Proposition 6.3.1. □

So instead of doing tedious matrix reductions everytime we want to express a vector in terms of a basis, if we have an orthonormal basis we simply have to compute some inner products to obtain the coefficients. Now to construct an orthonormal basis, we use the following algorithm:

Theorem 6.3.2 Gram-Schmidt orthorgonalisation

If $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m)$ is a linearly independent list of vectors in V, then there exists an orthonormal list $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ of vectors in V such that

$$\operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j) = \operatorname{span}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_j)$$

for $j = 1, \ldots, m$.

Corollary 6.3.2

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

Proof: Choose a basis of V. Apply the Gram-Schmidt procedure until we have an orthonormal list. This list is linearly independent by Corollary 6.3.1 and spans V, so it is an orthonormal basis of V. □

Corollary 6.3.3

Every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Proof: Suppose $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ is an orthonormal list of vectors in V. Then $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ is linearly independent by Corollary 6.3.1 so it can be extended to a basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of V by a previously proven result in Y1 Spring. Then we apply the Gram-Schmidt procedure to the list to obtain an orthonormal list which is linearly independent by Corollary 6.3.1 and spans V. □

The next result is about linear operators on V (recall the chapter on eigenvalues and eigenvectors).

Corollary 6.3.4

Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof: Suppose that T has an upper-triangular matrix with respect to some basis $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of V. Thus $\operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j)$ is invariant under T for each $j = 1, \ldots, n$.

Applying the Gram-Schmidt procedure, we obtain an orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of V. Since $\operatorname{span}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_j) = \operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j)$ for $j = 1, \ldots, n$, we conclude that $\operatorname{span}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_j)$ is invariant under T, so T has an upper-triangular matrix with respect to the orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$. □

Corollary 6.3.5 Schur's theorem

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. T has an upper-triangular matrix with respect to some orthonormal basis of V.

6.4 Orthogonal projections and minimisation problems

Definition 6.4.1: Orthogonal complement

If $U \subset V$, then the **orthogonal complement** of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{\boldsymbol{v} \in V : \langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0 \forall \boldsymbol{u} \in U\}.$$

Note:-

Note that:

- $U^\perp$ is always a subspace of V;
- $V^\perp = \{\mathbf{0}\}$;
- $\{\mathbf{0}\}^\perp = V$;
- $U_1 \subset U_2 \Rightarrow U_2^\perp \subset U_1^\perp$.

The following theorem shows that every subspace of an inner-product space leads to a natural direct sum decomposition of the whole space.

Theorem 6.4.1

If U is a subspace of V, then

$$V = U \oplus U^\perp.$$

Proof: Suppose that U is a subspace of V.

- First we show that $V = U + U^\perp$. Let $\boldsymbol{v} \in V$ and $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ be an orthonormal basis of U. Then we write

  $$\boldsymbol{v} = (\langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \ldots + \langle \boldsymbol{v}, \boldsymbol{e}_m \rangle \boldsymbol{e}_m) + (\boldsymbol{v} - \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 - \ldots - \langle \boldsymbol{v}, \boldsymbol{e}_m \rangle \boldsymbol{e}_m)$$

  Let $\boldsymbol{u} = \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \ldots + \langle \boldsymbol{v}, \boldsymbol{e}_m \rangle \boldsymbol{e}_m$ and $\boldsymbol{w} = (\boldsymbol{v} - \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 - \ldots - \langle \boldsymbol{v}, \boldsymbol{e}_m \rangle \boldsymbol{e}_m$. Clearly $\boldsymbol{u} \in U$. By taking the inner product between $\boldsymbol{w}$ and each $\boldsymbol{e}_j$, we have

  $$\langle \boldsymbol{w}, \boldsymbol{e}_j \rangle = \langle \boldsymbol{v}, \boldsymbol{e} \rangle - 0 - \ldots - 0 - \langle \boldsymbol{v}, \boldsymbol{e}_j \rangle = 0.$$

  Hence, $\boldsymbol{w}$ is orthogonal to every vector in U, so $\boldsymbol{w} \in U^\perp$. So we have written $\boldsymbol{v} \in V$ as a sum of a vector in U and $U^\perp$, so $V = U + U^\perp$.

- Suppose that $\boldsymbol{v} \in U \cap U^{\perp}$. So $\boldsymbol{v}$ is orthogonal to itself, implying that $\boldsymbol{v} = \mathbf{0}$, so $U \cap U^{\perp} = \{\mathbf{0}\}$. Hence $V = U \oplus U^{\perp}$.

□

Corollary 6.4.1

If U is a subspace of V, then

$$U = (U^{\perp})^{\perp}.$$

Proof: Suppose that U is a subspace of V.

- Suppose that $\boldsymbol{u} \in U$, then $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0 \, \forall \boldsymbol{v} \in U^{\perp}$. So $\boldsymbol{u}$ is orthogonal to every vector in $U^{\perp}$, hence $\boldsymbol{u} \in (U^{\perp})^{\perp} \Rightarrow U \subset (U^{\perp})^{\perp}$.
- Suppose that $\boldsymbol{u} \in (U^{\perp})^{\perp}$, then for some reason I don't understand, we can write $\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w}$ where $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in U^{\perp}$, even though the previous result requires $\boldsymbol{v} \in V$, idk anymore tbh. Then $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{v} \in (U^{\perp})^{\perp}$ and $\boldsymbol{w} \in U^{\perp}$ implies that $\boldsymbol{w} = \mathbf{0}$, so $\boldsymbol{u} = \boldsymbol{v}$, so $\boldsymbol{v} \in U$. Hence $(U^{\perp})^{\perp} \subset U$.

Combining both parts, $U = (U^{\perp})^{\perp}$. □

From Theorem 6.4.1, we know that each vector $\boldsymbol{v} \in V$ can be written uniquely as

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{w},$$

with $\boldsymbol{u} \in U$ and $\boldsymbol{w} \in U^{\perp}$. We use this decomposition to define the following operator on V.

Definition 6.4.2: Orthogonal projection

Let U be a subspace of V. The **orthogonal projection** of V onto U, P_U is an operator on V. For $\boldsymbol{v} \in V$, $P_U \boldsymbol{v}$ is such that

$$\boldsymbol{v} = P_U \boldsymbol{v} + \boldsymbol{w},$$

where $\boldsymbol{w} \in U^{\perp}$.

Note:-

- $\operatorname{Im} P_U = U$;
- $\ker P_U = U^{\perp}$;
- $\boldsymbol{v} - P_U \boldsymbol{v} \in U^{\perp} \, \forall \boldsymbol{v} \in V$;
- $P_U^2 = P_U$;
- $\|P_U \boldsymbol{v}\| \leq \|\boldsymbol{v}\| \, \forall \boldsymbol{v} \in V$.

Proposition 6.4.1

Suppose U is a subspace of V and $\boldsymbol{v} \in V$. Then

$$\|\boldsymbol{v} - P_U \boldsymbol{v}\| \leq \|\boldsymbol{v} - \boldsymbol{u}\|$$

$\forall \boldsymbol{u} \in U$. Furthermore, if $\boldsymbol{u} \in U$ and the inequality is an equality, then $\boldsymbol{u} = P_U \boldsymbol{v}$.

Proof: Suppose $\boldsymbol{u} \in U$, then

$$\begin{aligned}
\|\boldsymbol{v} - P_U\boldsymbol{v}\|^2 &\leq \|\boldsymbol{v} - P_U\boldsymbol{v}\|^2 + \|P_U\boldsymbol{v} - \boldsymbol{u}\|^2 \\
&= \|(\boldsymbol{v} - P_U\boldsymbol{v}) + (P_U\boldsymbol{v} - \boldsymbol{u})\|^2 \text{ by Pythagorean theorem} \\
&= \|\boldsymbol{v} - \boldsymbol{u}\|^2 \\
\|\boldsymbol{v} - P_U\boldsymbol{v}\|^2 &\leq \|\boldsymbol{v} - \boldsymbol{u}\|
\end{aligned}$$

We have equality iif $\|P_U\boldsymbol{v} - \boldsymbol{u}\| = 0 \Leftrightarrow P_U\boldsymbol{v} = \boldsymbol{u}$. □

6.5 Linear functionals and adjoints

Definition 6.5.1: Linear functional

A **linear functional** on V is a linear map from V to the scalars F.

Theorem 6.5.1

Suppose ϕ is a linear functional on V. Then there is a unique vector $\boldsymbol{v} \in V$ such that

$$\phi(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle$$

$\forall \boldsymbol{u} \in V$.

Proof: Let $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ be an orthonormal basis of V, then

$$\begin{aligned}
\phi(\boldsymbol{u}) &= \phi(\langle \boldsymbol{u}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \ldots + \langle \boldsymbol{u}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n) \\
&= \langle \boldsymbol{u}, \boldsymbol{e}_1 \rangle \phi(\boldsymbol{e}_1) + \ldots + \langle \boldsymbol{u}, \boldsymbol{e}_n \rangle \phi(\boldsymbol{e}_n) \text{ by linearity;} \\
&= \langle \boldsymbol{u}, \overline{\phi(\boldsymbol{e}_1)}\boldsymbol{e}_1 + \ldots + \overline{\phi(\boldsymbol{e}_n)}\boldsymbol{e}_n \rangle \text{ by conjugate homogeneity in second slot,}
\end{aligned}$$

$\forall \boldsymbol{u} \in V$. By setting $\boldsymbol{v} = \overline{\phi(\boldsymbol{e}_1)}\boldsymbol{e}_1 + \ldots + \overline{\phi(\boldsymbol{e}_n)}\boldsymbol{e}_n$, we have $\phi(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v} \rangle \, \forall \boldsymbol{u} \in V$.

To show the uniqueness of $\boldsymbol{v}$, suppose we have $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$ such that

$$\phi(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v}_1 \rangle = \langle \boldsymbol{u}, \boldsymbol{v}_2 \rangle$$

$\forall \boldsymbol{u} \in V$, then

$$0 = \langle \boldsymbol{u}, \boldsymbol{v}_1 \rangle - \langle \boldsymbol{u}, \boldsymbol{v}_2 \rangle = \langle \boldsymbol{u}, \boldsymbol{v}_1 - \boldsymbol{v}_2 \rangle$$

by additivity in second slot. Taking $\boldsymbol{u} = \boldsymbol{v}_1 - \boldsymbol{v}_2$ gives $\langle \boldsymbol{v}_1 - \boldsymbol{v}_2, \boldsymbol{v}_1 - \boldsymbol{v}_2 \rangle = 0 \Longrightarrow \boldsymbol{v}_1 = \boldsymbol{v}_2$. □

For here onwards, we let W be a finite-dimensional, nonzero, inner-product space over $\mathbb{F}$.

Definition 6.5.2: Adjoint

Let $T \in \mathcal{L}(V, W)$. The **adjoint** of T, T^*, is the function from W to V such that for $\boldsymbol{w} \in W$, $\boldsymbol{v} \in V$

$$\langle T\boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{v}, T^*\boldsymbol{w} \rangle.$$

Theorem 6.5.1 guarantees the existence and uniqueness of such a vector.

Example 6.5.1 (Computing adjoints)

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

Then T^* will be a function from $\mathbb{R}^2$ to $\mathbb{R}^3$. To compute T^*, fix a point $(y_1, y_2) \in \mathbb{R}^2$, then

$$\begin{aligned}\langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle = \langle T(x_1, x_2, x_3), (y_1, y_2)\rangle \\ &= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle \\ &= x_2 y_1 + 3x_3 y_1 + 2x_1 y_2 \\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle\end{aligned}$$

$\forall (x_1, x_2, x_3) \in \mathbb{R}^3 \Rightarrow T^*(y_1, y_2) = (2y_2, y_1, 3y_1)$. Note that T^* turns out to be a linear map: this is true in general.

Proposition 6.5.1

$T \in \mathcal{L}(V, W) \Rightarrow T^* \in \mathcal{L}(W, V)$.

Proof: Suppose $T \in \mathcal{L}(V, W)$, $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$, $k \in \mathbb{F}$, then

- $$\begin{aligned}\langle T\boldsymbol{v}, \boldsymbol{w}_1 + \boldsymbol{w}_2\rangle &= \langle T\boldsymbol{v}, \boldsymbol{w}_1\rangle + \langle T\boldsymbol{v}, \boldsymbol{w}_2\rangle \text{ by additivity on second slot} \\ &= \langle \boldsymbol{v}, T^*\boldsymbol{w}_1\rangle + \langle \boldsymbol{v}, T^*\boldsymbol{w}_2\rangle \\ \langle \boldsymbol{v}, T^*(\boldsymbol{w}_1 + \boldsymbol{w}_2)\rangle &= \langle \boldsymbol{v}, T^*\boldsymbol{w}_1 + T^*\boldsymbol{w}_2\rangle\end{aligned}$$

- $$\begin{aligned}\langle T\boldsymbol{v}, k\boldsymbol{w}_1\rangle &= \bar{k}\langle T\boldsymbol{v}, \boldsymbol{w}_1\rangle \text{ by conjugate homogeneity;} \\ &= \bar{k}\langle \boldsymbol{v}, T^*\boldsymbol{w}_1\rangle \\ \langle \boldsymbol{v}, T^*(k\boldsymbol{w}_1)\rangle &= \langle \boldsymbol{v}, kT^*\boldsymbol{w}_1\rangle\end{aligned}$$

Hence T^* is a linear map. □

Note:-

The function $T \mapsto T^*$ has the following properties (to be verified):

1. Additivity: $(S + T)^* = S^* + T^* \ \forall S, T \in \mathcal{L}(V, W)$;
2. Conjugate homogeneity: $(aT)^* = \bar{a}T^* \ \forall a \in \mathbb{F}, T \in \mathcal{L}(V, W)$;
3. Adjoint of adjoint: $(T^*)^* = T \ \forall T \in \mathcal{L}(V, W)$;
4. Identity: $I^* = I$, where I is the identity operator on V;
5. Products: $(ST)^* = T^*S^* \ \forall T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$, where U is an inner-product space over $\mathbb{F}$.

Verification:

1. Let $S, T \in \mathcal{L}(V, W)$. $\forall \boldsymbol{v} \in V, \boldsymbol{w} \in W$,
$$\begin{aligned}\langle \boldsymbol{v}, (S+T)^* \boldsymbol{w}\rangle &= \langle (S+T)\boldsymbol{v}, \boldsymbol{w}\rangle \\ &= \langle S\boldsymbol{v}, \boldsymbol{w}\rangle + \langle T\boldsymbol{v}, \boldsymbol{w}\rangle \\ &= \langle \boldsymbol{v}, S^*\boldsymbol{w}\rangle + \langle \boldsymbol{v}, T^*\boldsymbol{w}\rangle.\end{aligned}$$

2. Let $a \in \mathbb{F}, T \in \mathcal{L}(V, W)$, then $\forall \boldsymbol{v} \in V, \boldsymbol{w} \in W$,
$$\begin{aligned}\langle \boldsymbol{v}, (aT)^* \boldsymbol{w}\rangle &= \langle aT\boldsymbol{v}, \boldsymbol{w}\rangle \\ &= a\langle T\boldsymbol{v}, \boldsymbol{w}\rangle \text{ by homogeneity in first slot;} \\ &= a\langle \boldsymbol{v}, T^*\boldsymbol{w}\rangle \\ &= \langle \boldsymbol{v}, \bar{a}T^*\boldsymbol{w}\rangle \text{ by conjugate homogeneity in second slot.}\end{aligned}$$

3. Let $T \in \mathcal{L}(V, W)$, then $\forall \boldsymbol{v} \in V, \boldsymbol{w} \in W$,
$$\begin{aligned}\langle \boldsymbol{v}, (T^*)^* \boldsymbol{w}\rangle &= \langle T^*\boldsymbol{v}, \boldsymbol{w}\rangle \\ &= \overline{\langle \boldsymbol{w}, T^*\boldsymbol{v}\rangle} \text{ by conjugate symmetry;} \\ &= \overline{\langle T\boldsymbol{w}, \boldsymbol{v}\rangle} \\ &= \langle \boldsymbol{v}, T\boldsymbol{w}\rangle.\end{aligned}$$

   Hence $T = (T^*)^*$.

4. $\forall \boldsymbol{v} \in V$, $\langle I\boldsymbol{v}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, I^*\boldsymbol{v}\rangle = \langle \boldsymbol{v}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, I\boldsymbol{v}\rangle$. Hence $I = I^*$.

5. Let $S \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(W, V)$, then $\forall \boldsymbol{v} \in V, \boldsymbol{w} \in W$, we have
$$\begin{aligned}\langle \boldsymbol{v}, (ST)^* \boldsymbol{w}\rangle &= \langle ST\boldsymbol{v}, \boldsymbol{w}\rangle \\ &= \langle T\boldsymbol{v}, S^*\boldsymbol{w}\rangle \\ &= \langle \boldsymbol{v}, T^*S^*\boldsymbol{w}\rangle.\end{aligned}$$

   Hence $(ST)^* = T^*S^*$.

Proposition 6.5.2

Suppose $T \in \mathcal{L}(V, W)$, then

1. $\ker T^* = (\operatorname{Im} T)^\perp$;
2. $\operatorname{Im} T* = (\ker T)^\perp$;
3. $\ker T = (\operatorname{Im} T^*)^\perp$;
4. $\operatorname{Im} T = (\ker T^*)^\perp$.

Proof: 1. Let $\boldsymbol{w} \in W$, then
$$\begin{aligned}\boldsymbol{w} \in \ker T^* &\Leftrightarrow T^*\boldsymbol{w} = \boldsymbol{0} \\ &\Leftrightarrow \langle \boldsymbol{v}, T^*\boldsymbol{w}\rangle = 0\,\forall \boldsymbol{v} \in V \\ &\Leftrightarrow \langle T\boldsymbol{v}, \boldsymbol{w}\rangle = 0\,\forall \boldsymbol{v} \in V \\ &\Leftrightarrow \boldsymbol{w} \in (\operatorname{Im} T)^\perp\end{aligned}$$

   So $\ker T^* = (\operatorname{Im} T)^\perp$.

3. Let $\boldsymbol{v} \in V$, then

$$\begin{aligned} \boldsymbol{v} \in \ker T &\Leftrightarrow T\boldsymbol{v} = \mathbf{0} \\ &\Leftrightarrow \langle T\boldsymbol{v}, \boldsymbol{w} \rangle = 0 \,\forall \boldsymbol{w} \in W \\ &\Leftrightarrow \langle \boldsymbol{v}, T^*\boldsymbol{w} \rangle = 0 \,\forall \boldsymbol{w} \in W \\ &\Leftrightarrow \boldsymbol{v} \in (\operatorname{Im} T)^{\perp}. \end{aligned}$$

   So $\ker T = (\operatorname{Im} T^*)^{\perp}$.

We get (4) and (2) by taking the orthogonal complement of both sides of (1) and (3). □

Definition 6.5.3: Conjugate transpose

The **conjugate tranpose** of an $A \in M_{mn}$ is the $n \times m$ matrix obtained by taking the transpose of A and the complex conjugate of each entry.

Proposition 6.5.3

Suppose $T \in \mathcal{L}(V, W)$. If $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is an orthonormal basis of V and $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m)$ is an orthonormal basis of W, then

$$\mathcal{M}(T^*, (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m), (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n))$$

is the conjugate transpose of

$$\mathcal{M}(T, (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n), (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m)).$$

Proof: Suppose that $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is an orthonormal basis of V and $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m)$ is an orthonormal basis of W. We abbreviate the two matrices as $\mathcal{M}(T^*)$ and $\mathcal{M}(T)$ omitting the bases.

We can obtain the kth column of $\mathcal{M}(T)$ be writing $T\boldsymbol{e}_k$ as a linear combination of $\boldsymbol{f}_i$. Since $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_m)$ is an orthonormal basis, by Theorem 6.3.1, we can write $T\boldsymbol{e}_k$ as

$$T\boldsymbol{e}_k = \langle T\boldsymbol{e}_k, \boldsymbol{f}_1 \rangle \boldsymbol{f}_1 + \ldots + \langle T\boldsymbol{e}_k, \boldsymbol{f}_m \rangle \boldsymbol{f}_m.$$

So the jkth entry of $\mathcal{M}(T)$ is $\langle T\boldsymbol{e}_k, \boldsymbol{f}_j \rangle$ (with respect to basis of W). Similarly for $\mathcal{M}(T^*)$, the jkth entry is

$$\langle T^*\boldsymbol{f}_k, \boldsymbol{e}_j \rangle = \langle \boldsymbol{f}_k, T\boldsymbol{e}_j \rangle = \overline{\langle T\boldsymbol{e}_j, \boldsymbol{f}_k \rangle}$$

by conjugate symmetry of inner products, which is the complex conjugate of the entry in row k, column j of $\mathcal{M}(T)$. Hence $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$. □

Chapter 7

Operators on inner-product spaces

7.1 Self-adjoint and normal operators

Definition 7.1.1: Self-adjoint

An operator $T \in \mathcal{L}(V)$ is **self-adjoint** if $T = T^*$.

The sum of two self-adjoint operators is self adjoint since by additivity of adjoints, for self-adjoint operators $S, T \in \mathcal{L}(V)$, $(S+T)* = S * + T* = S + T$. Also, the product of a real scalar and a self-adjoint operator is self-adjoint, since by conjugate homogeneity of adjoints, for $k \in \mathbb{R}$, $(kT)* = \bar{k}T^* = kT$.

Proposition 7.1.1

Every eigenvalue of a self-adjoint operator is real.

Proof: Suppose T is a self-adjoint operator on V. Let λ be an eigenvalue of T and $\boldsymbol{v}$ be a nonzero vector in V such that $T\boldsymbol{v} = \lambda\boldsymbol{v}$, then

$$\begin{aligned}
\lambda\|\boldsymbol{v}\|^2 &= \langle \lambda\boldsymbol{v}, \boldsymbol{v}\rangle \text{ by homogeneity in the first slot;}\\
&= \langle T\boldsymbol{v}, \boldsymbol{v}\rangle\\
&= \langle \boldsymbol{v}, T\boldsymbol{v}\rangle \text{ since } T \text{ is self-adjoint;}\\
&= \langle \boldsymbol{v}, \lambda\boldsymbol{v}\rangle\\
&= \bar{\lambda}\|\boldsymbol{v}\|^2\\
\lambda = \bar{\lambda} &\Rightarrow \lambda \in \mathbb{R} \text{ since } \boldsymbol{v} \neq \mathbf{0}.
\end{aligned}$$

□

Proposition 7.1.2

If V is a complex inner-product space and T is an operator on V such that

$$\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$$

$\forall \boldsymbol{v} \in V$, then $T = 0$.

Proof: Suppose V is a complex inner-product space and $T \in \mathcal{L}(V)$, then $\forall \boldsymbol{u}, \boldsymbol{w} \in V$, we have

$$\begin{aligned}
\langle T\boldsymbol{u}, \boldsymbol{w}\rangle =& \frac{\langle T(\boldsymbol{u}+\boldsymbol{w}), \boldsymbol{u}+\boldsymbol{w}\rangle - \langle T(\boldsymbol{u}-\boldsymbol{w}), \boldsymbol{u}-\boldsymbol{w}\rangle}{4} + \frac{\langle T(\boldsymbol{u}+i\boldsymbol{w}), \boldsymbol{u}+i\boldsymbol{w}\rangle - \langle T(\boldsymbol{u}-i\boldsymbol{w}), \boldsymbol{u}-i\boldsymbol{w}\rangle}{4} i \\
=& \frac{1}{4}(\langle T\boldsymbol{u}, \boldsymbol{u}\rangle + \langle T\boldsymbol{w}, \boldsymbol{u}\rangle + \langle T\boldsymbol{u}, \boldsymbol{w}\rangle + \langle T\boldsymbol{w}, \boldsymbol{w}\rangle - \langle T\boldsymbol{u}, \boldsymbol{u}\rangle + \langle T\boldsymbol{w}, \boldsymbol{u}\rangle + \langle T\boldsymbol{u}, \boldsymbol{w}\rangle - \langle T\boldsymbol{w}, \boldsymbol{w}\rangle) \\
&+ \frac{1}{4} i(\langle T\boldsymbol{u}, \boldsymbol{u}\rangle + i\langle T\boldsymbol{w}, \boldsymbol{u}\rangle - i\langle T\boldsymbol{u}, \boldsymbol{w}\rangle - i^2\langle T\boldsymbol{w}, \boldsymbol{w}\rangle - (\langle T\boldsymbol{u}, \boldsymbol{u}\rangle - i\langle T\boldsymbol{w}, \boldsymbol{u}\rangle + i\langle T\boldsymbol{u}, \boldsymbol{w}\rangle - i^2\langle T\boldsymbol{w}, \boldsymbol{w}\rangle)) \\
=& \frac{1}{4}(2\langle T\boldsymbol{u}, \boldsymbol{w}\rangle + 2\langle T\boldsymbol{w}, \boldsymbol{u}\rangle) + \frac{1}{4} i(2i\langle T\boldsymbol{w}, \boldsymbol{u}\rangle - 2i\langle T\boldsymbol{u}, \boldsymbol{w}\rangle) \\
=& \frac{1}{4}(2\langle T\boldsymbol{u}, \boldsymbol{w}\rangle + 2\langle T\boldsymbol{w}, \boldsymbol{u}\rangle - 2\langle T\boldsymbol{w}, \boldsymbol{u}\rangle + 2\langle T\boldsymbol{u}, \boldsymbol{v}\rangle) \\
=& \frac{1}{4}(4\langle T\boldsymbol{u}, \boldsymbol{w}\rangle) \\
=& \langle T\boldsymbol{u}, \boldsymbol{w}\rangle
\end{aligned}$$

as verified. Since each inner product on the RHS is of the form $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle$, if we have $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$, then $\langle T\boldsymbol{u}, \boldsymbol{w}\rangle = 0$ for all $\boldsymbol{u}, \boldsymbol{w} \in V$, implying that $T = 0$ by taking $\boldsymbol{w} = T\boldsymbol{u}$. □

Corollary 7.1.1

Let V be a complex inner-product space and let $T \in \mathcal{L}(V)$. T is self-adjoint iif

$$\langle T\boldsymbol{v}, \boldsymbol{v}\rangle \in \mathbb{R}$$

$\forall \boldsymbol{v} \in V$.

Proof: Let $\boldsymbol{v} \in V$, then

$$\begin{aligned}
\langle T\boldsymbol{v}, \boldsymbol{v}\rangle - \overline{\langle T\boldsymbol{v}, \boldsymbol{v}} &= \langle T\boldsymbol{v}, \boldsymbol{v}\rangle - \langle \boldsymbol{v}, T\boldsymbol{v}\rangle \\
2\operatorname{Im}\langle T\boldsymbol{v}, \boldsymbol{v}\rangle &= \langle T\boldsymbol{v}, \boldsymbol{v}\rangle - \langle T^*\boldsymbol{v}, \boldsymbol{v}\rangle \\
&= \langle (T - T^*)\boldsymbol{v}, \boldsymbol{v}\rangle
\end{aligned}$$

- ($\Rightarrow$) Suppose $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle \in \mathbb{R}$, then $\forall \boldsymbol{v} \in V$, $\langle (T - T^*)\boldsymbol{v}, \boldsymbol{v}\rangle = 0 \Rightarrow T - T^* = 0 \Rightarrow T = T^*$ i.e. T is self-adjoint by previous result.
- ($\Leftarrow$) Suppose that T is self-adjoint, then $2\operatorname{Im}\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = \langle 0, \boldsymbol{v}\rangle = 0$, so $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle \in \mathbb{R}\, \forall \boldsymbol{v} \in V$.

□

For Proposition 7.1.1 to hold for real inner-product spaces, we require T to be a self-adjoint operator.

Proposition 7.1.3

T is a self-adjoint operator on V such that

$$\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$$

$\forall \boldsymbol{v} \in V$ iif $T = 0$.

Proof: Suppose that T is a self-adjoint operator on V such that $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle = 0\ \forall \boldsymbol{v} \in V$. For $\boldsymbol{u}, \boldsymbol{w} \in V$, we can write $\langle T\boldsymbol{u}, \boldsymbol{w}\rangle$ as

$$
\begin{aligned}
&\frac{1}{4}(\langle T(\boldsymbol{u}+\boldsymbol{w}),\boldsymbol{u}+\boldsymbol{w}\rangle - T\langle(\boldsymbol{u}-\boldsymbol{w}),\boldsymbol{u}-\boldsymbol{w}\rangle)\\
&=\frac{1}{4}(2\langle T\boldsymbol{u},\boldsymbol{w}\rangle + 2\langle T\boldsymbol{w},\boldsymbol{u}\rangle)\\
&=\frac{1}{4}(2\langle T\boldsymbol{u},\boldsymbol{w}\rangle + 2\langle \boldsymbol{w},T*\boldsymbol{u}\rangle)\\
&=\frac{1}{4}(2\langle T\boldsymbol{u},\boldsymbol{w}\rangle + 2\langle \boldsymbol{w},T\boldsymbol{u}\rangle)\\
&=\frac{1}{4}(2\langle T\boldsymbol{u},\boldsymbol{w}\rangle + 2\langle T\boldsymbol{u},\boldsymbol{w}\rangle) \text{ by conjugate symmetry of inner products}\\
&=\langle T\boldsymbol{u},\boldsymbol{w}\rangle.
\end{aligned}
$$

Since we have $\langle T\boldsymbol{v},\boldsymbol{v}\rangle = 0$, $\langle T\boldsymbol{u},\boldsymbol{w}\rangle = 0$. By taking $\boldsymbol{w} = T\boldsymbol{u}$, we have $T = 0$.
For the converse statement, suppose $T = 0_V$, then $\langle T\boldsymbol{v},\boldsymbol{v}\rangle = \langle 0_V,\boldsymbol{v}\rangle = 0$. Clearly T is self-adjoint since the conjugate transpose of a zero matrix is still the zero matrix. Hence the converse statement holds. □

Definition 7.1.2: Normal operators

An operator on an inner-product space is **normal** if it commutes with its adjoint, i.e.

$$TT^* = T^*T.$$

Example 7.1.1 (Normal operators)

Every self-adjoint operator is normal, but not all normal operators are self-adjoint. For example, consider the matrix

$$A = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} \in M_{22}.$$

Computing the adjoint of the operator on $\mathbb{F}^2$ represented by A with respect to the standard basis, that is

$$T : \mathbb{F}^2 \longrightarrow \mathbb{F}^2; (x, y) \mapsto (2x + 3y, -3x + 2y).$$

We fix $\boldsymbol{w} = (w_1, w_2) \in \mathbb{F}^2$, then

$$
\begin{aligned}
\langle (x,y), T^*\boldsymbol{w}\rangle &= \langle T(x,y),\boldsymbol{w}\rangle\\
&= \langle (2x+3y, -3x+2y), (w_1,w_2)\rangle\\
&= 2xw_1 + 3yw_1 - 3xw_2 + 2yw_2\\
&= (2w_1 - 3w_2)x + (3w_1 + 2w_2)y\\
&= \langle (x,y), (2w_1 - 3w_2, 3w_1 + 2w_2)\rangle
\end{aligned}
$$

So $T^*(w_1, w_2) = (2w_1 - 3w_2, 3w_1 + 2w_2)$ so T is not self-adjoint since $T \neq T^*$. But T is normal since

$$
\begin{aligned}
TT^*(x,y) &= T(2x - 3y, 3x + 2y)\\
&= (2(2x-3y) + 3(3x+2y), -3(2x-3y) + 2(3x+2y))\\
&= (4x - 6y + 9x + 6y, -6x + 9y + 6x + 4y)\\
&= (13x, 13y)\\
T^*T(x,y) &= T^*(2x + 3y, -3x + 2y)\\
&= (2(2x-3y) + 3(3x+2y), -3(2x-3y) + 2(3x+2y))\\
&= (4x - 6y + 9x + 6y, -6x + 9y + 6x + 4y)\\
&= (13x, 13y) = TT^*(x,y)
\end{aligned}
$$

$\forall x, y \in \mathbb{F}$.

Proposition 7.1.4

An operator $T \in \mathcal{L}(V)$ is normal iif
$$\|T\boldsymbol{v}\| = \|T^*\boldsymbol{v}\|$$
$\forall \boldsymbol{v} \in V$.

Proof: Let $T \in \mathcal{L}(V)$. T is normal iif
$$\begin{aligned} T^*T - TT^* = 0 &\Leftrightarrow \langle (T^*T - TT^*)\boldsymbol{v}, \boldsymbol{v}\rangle = 0 \,\forall \boldsymbol{v} \in V \text{ by Proposition 7.1.3} \\ &\Leftrightarrow T^*T\boldsymbol{v}, \boldsymbol{v}\rangle = \langle TT^*\boldsymbol{v}, \boldsymbol{v}\rangle \,\forall \boldsymbol{v} \in V \\ &\Leftrightarrow \langle T\boldsymbol{v}, T\boldsymbol{v}\rangle = \langle T^*\boldsymbol{v}, T^*\boldsymbol{v}\rangle \,\forall \boldsymbol{v} \in V \\ &\Leftrightarrow \|T\boldsymbol{v}\|^2 = \|T^*\boldsymbol{v}\|^2 \,\forall \boldsymbol{v} \in V \\ &\Leftrightarrow \|T\boldsymbol{v}\| = \|T^*\boldsymbol{v}\| \,\forall \boldsymbol{v} \in V. \end{aligned}$$

□

Corollary 7.1.2

Suppose $T \in \mathcal{L}(V)$ is normal. If $\boldsymbol{v} \in V$ is an eigenvector of T with eigenvalue $\lambda \in \mathbb{F}$, then $\boldsymbol{v}$ is also an eigenvector of T^* with eigenvalue $\bar{\lambda}$.

Proof: Suppose $\boldsymbol{v} \in V$ is an eigenvector of T with eigenvalue λ. By definition of eigenvectors we have
$$(T - \lambda I)\boldsymbol{v} = \mathbf{0}.$$
Since T is normal, $TT^* = T^*T$. Then we consider
$$\begin{aligned} (T - \lambda I)(T - \lambda I)^* &= (T - \lambda I)(T^* - \bar{\lambda} I^*) \\ &= TT^* - T\bar{\lambda}I - \lambda IT^* + \lambda\bar{\lambda}II \\ &= T^*T - \bar{\lambda}T - \lambda T^* + \lambda\bar{\lambda}I \\ &= (T^* - \bar{\lambda}I^*)(T - \lambda I) \\ &= (T - \lambda I)^*(T - \lambda I) \end{aligned}$$
so $(T - \lambda I)$ is also normal. Then by Proposition 7.1.4, we have
$$\|(T - \lambda I)\boldsymbol{v}\| = \|(T - \lambda I)^*\boldsymbol{v}\| = \|(T^* - \bar{\lambda}I)\boldsymbol{v}\| = 0$$
so $\boldsymbol{v}$ is an eigenvector of T^* corresponding to eigenvalue $\bar{\lambda}$. □

Corollary 7.1.3

If $T \in \mathcal{L}(V)$ is normal, then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof: Suppose $T \in \mathcal{L}(V)$ is normal and α, β are distinct eigenvalues of T, with corresponding eigenvectors $\boldsymbol{u}, \boldsymbol{v} \in V$. So we have
$$T\boldsymbol{u} = \alpha \qquad \text{and} \qquad T\boldsymbol{v} = \beta\boldsymbol{v}.$$
To determine if the eigenvectors are orthogonal, we consider
$$\begin{aligned} (\alpha - \beta)\langle \boldsymbol{u}, \boldsymbol{v}\rangle &= \alpha\langle \boldsymbol{u}, \boldsymbol{v}\rangle - \beta\langle \boldsymbol{u}, \boldsymbol{v}\rangle \\ &= \langle \alpha\boldsymbol{u}, \boldsymbol{v}\rangle - \langle \boldsymbol{u}, \bar{\beta}\boldsymbol{v}\rangle \\ &= \langle T\boldsymbol{u}, \boldsymbol{v}\rangle - \langle \boldsymbol{u}, T^*\boldsymbol{v}\rangle \\ &= 0. \end{aligned}$$
Since $\alpha \neq \beta$, we have $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$ implying that $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal. □

7.2 Diagonal matrices

Definition 7.2.1

An operator $T \in \mathcal{L}(V)$ has a diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \in M_{nn}$$

with respect to a basis $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ of V iif we have

$$\begin{aligned} T\boldsymbol{v}_1 &= \lambda_1 \boldsymbol{v}_1 \\ &\vdots \\ T\boldsymbol{v}_n &= \lambda_n \boldsymbol{v}_n; \end{aligned}$$

Note:-

Recall that we say that M is a matrix of an operator T with respect to a basis of V when the output vector given when applied to a vector (with respect to a basis of V) is with respect to the basis. To obtain such a matrix, we perform change of basis:

$$A_B = P^{-1}AP$$

where P is the matrix with columns containing the basis vectors from B and A_B is the matrix A with respect to basis B. For example, consider the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in M_{22}$$

with respect to the standard basis. Suppose we want to change it so that it is with respect to basis $B = \{(1,1)^T, (1,0)^T\}$, by performing change of basis:

$$A_B = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 7 & 3 \\ -4 & -2 \end{pmatrix}.$$

Now if we apply A and A_B both to $(a,b)^T$, we would obtain $(7a+3b, -4a-2b)^T$ and $(3a+b, 7a+3b)^T$ respectively. Note that when applying A_B, it takes in $(a,b)^T$ with respect to B, so in terms of the standard basis, $(a,b)^T$ really is $(a+b,a)^T$, to which when A is applied, gives $(3a+b, 7a+3b)^T$. If we express B in terms of the standard basis, we get

$$(7a+3b)(1,1) + (-4a-2b)(1,0) = (7a+3b, 7a+3b) + (-4a, -2b, 0) = (3a+b, 7a+3b).$$

Note that after we left-multiply A to P, the resulting matrix now treats its **input vectors** in terms of basis B since when AP is applied to $\boldsymbol{v}$, P is multiplied first, then applying A gives the **output vector** with respect to the standard basis, so we apply P^{-1} so that this vector is with respect to B.

Returning to our definition, this makes sense because if the diagonal matrix is with respect to a basis $B = (\boldsymbol{v}_1, \dots, \boldsymbol{v}_n)$ of V, then applying T to any $\boldsymbol{v}_i$ would be equivalent to applying the diagonal matrix to a column vector $\boldsymbol{u}_i \in V$ with entries $u_{ij} = 1$ for $i = j$ and zero otherwise, resulting in the definition given.

Now we happen to notice that those equations imply that $\boldsymbol{v}_i$ are eigenvectors corresponding to eigenvalues λ_i. Hence we can conclude further that T **has a diagonal matrix** D **with respect to basis** B **iif** B **consists of eigenvectors of** T **corresponding to entries in** D.

Proposition 7.2.1

If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T has a diagonal matrix with respect to some basis of V.

Proof: Suppose that $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues $\lambda_1, \ldots, \lambda_{\dim V}$. For each j, let $\boldsymbol{v}_j \in V$ be a nonzero eigenvector corresponding to distinct eigenvalues λ_j, so $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{\dim V})$ is linearly independent. Since this is a list of $\dim V$ linear independent vectors, it is a basis of V. By definition, T has a diagonal matrix with diagonal entries equal to eigenvalues corresponding to each $\boldsymbol{v}_j$. □

Proposition 7.2.2

Suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, then the following are equivalent:

1. T has a diagonal matrix with respect to some basis of V;
2. V has a basis consisting of eigenvectors of T;
3. there exist one-dimensional subspaces $U_1, \ldots, U_n$ of V, each invariant under T, such that
$$V = U_1 \oplus \ldots \oplus U_n;$$
4. $V = \ker(T - \lambda_1 I) \oplus \ldots \oplus \ker(T - \lambda_m T)$;
5. $\dim V = \operatorname{nullity}(T - \lambda_1 I) + \ldots + \operatorname{nullity}(T - \lambda_m I)$.

Proof: We prove this in 5 directions:

($1 \Leftrightarrow 2$) Shown previously.

($2 \Rightarrow 3$) Suppose that (2) holds. For each j, let $U_j = \operatorname{span}(\boldsymbol{v}_j)$. Each U_j is a one-dimensional subspace of V that is invariant under T. Since $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is a basis of V, each vector in V can be written uniquely as a linear combination of $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$, which can then be written as a sum $\boldsymbol{u}_1 + \ldots + \boldsymbol{u}_n$ which each $\boldsymbol{u}_j \in U_j$. So $V = U_1 \oplus \ldots \oplus U_n$ since $U_1 \cap \ldots \cap U_n = \{\mathbf{0}\}$ as $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is linearly independent.

($3 \Rightarrow 2$) Suppose that (3) holds. For each j, let $\boldsymbol{v}_j$ be a nonzero vector in U_j which are invariant under T, so each $\boldsymbol{v}_j$ is an eigenvector of T. Each vector in V can be written uniquely as $\boldsymbol{u}_1 + \ldots + \boldsymbol{u}_n$ where each $\boldsymbol{u}_j \in U_j$ by our assumption, which is a linear combination of $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ since each $\boldsymbol{u}_j$ is a scalar multiple of $\boldsymbol{v}_j$. Hence this list of eigenvectors spans V and is linearly independent, so it is a basis of V.

($2 \Rightarrow 4$) Suppose that (2) holds, then every vector in V is a linear combination of eigenvectors of T, so we have
$$V = \ker(T - \lambda_1 I) + \ldots + \ker(T - \lambda_m I),$$
which makes sense if you consider the definition of sums of subspaces and eigenvectors. Also, by considering
$$\boldsymbol{u}_1 + \ldots + \boldsymbol{u}_m = \mathbf{0}$$

for $\boldsymbol{u}_j \in \ker(T - \lambda_j I)$. Since nonzero eigenvectors corresponding to distinct eigenvalues are linearly independent, each $\boldsymbol{u}_j = \mathbf{0}$, so

$$\ker(T - \lambda_1 I) \cap \ldots \cap \ker(T - \lambda_m I) = \{\mathbf{0}\}$$

since containing any more vectors would imply that every eigenvector is linearly dependent to each other, and so there is only one distinct eigenvalue which is a contradiction. Hence the sum is a direct sum.

(4 ⇒ 5) Suppose that (4) holds, then (5) holds immediately from Proposition 5.3.2 from Spring semester notes.

(5 ⇒ 2) Suppose that (5) holds. We choose a basis of each $\ker(T - \lambda_j I)$ and put all these bases together to form a list $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of eigenvectors of T, where $n = \dim V$. To show that this list is linearly independent, suppose

$$a_1 \boldsymbol{v}_1 + \ldots + a_n \boldsymbol{v}_n = \mathbf{0}$$

where $a_1, \ldots, a_n \in \mathbb{F}$. For each $j = 1, \ldots, m$, let $\boldsymbol{u}_j$ denote the sum of all the terms $a_k \boldsymbol{v}_k$ such that $\boldsymbol{v}_k \in \ker(T - \lambda_j I)$. Thus each $\boldsymbol{u}_j$ is an eigenvector of T corresponding to eigenvalue λ_j and we have

$$\boldsymbol{u}_1 + \ldots + \boldsymbol{u}_m = \mathbf{0}.$$

Since nonzero eigenvectors corresponding to distinct eigenvalues are linearly independent, this implies that each $\boldsymbol{u}_j = \mathbf{0}$, and since each u_j is a linear combination of $\boldsymbol{v}_k$, each $a_k = 0$, hence $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ is linearly independent and is a basis of V.

□

7.3 The spectral theorem

Theorem 7.3.1 Complex spectral theorem

Suppose that V is a complex inner-product space and $T \in \mathcal{L}(V)$. V has an orthonormal basis consisting of eigenvectors of T iif T is normal.

Proof: We prove this in two directions:

- (⇒) Suppose that V has an orthonormal basis consisting of eigenvectors of T. By Proposition 7.2.2, with respect to this basis, T has a diagonal matrix $\mathcal{M}(T)$.

  Our goal is to show that T is normal, i.e. $TT^* = T^*T$, hence we now consider T^*. The matrix of T^* with respect to the same basis can be obtained by taking the conjugate transpose of $\mathcal{M}(T)$, which is also a diagonal matrix. Since any two diagonal matrices commute, T commutes with T^*, hence T is normal.

- (⇐) Suppose that T is normal. There is an orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of V such that T has an upper-triangular matrix with respect to this basis by Corollary 6.4.5. Let

$$\mathcal{M}(T, (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)) = \begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}.$$

We now want to show that this matrix is actually a diagonal matrix, i.e. $a_{ij} = 0$ for $i \neq j$, which by Proposition 7.2.2 implies that $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ consists of eigenvectors of T.

We see from the matrix above that

$$\|T\boldsymbol{e}_1\|^2 = |a_{1,1}|^2 \qquad \text{and} \qquad \|T^*\boldsymbol{e}_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \ldots + |a_{1,n}|^2.$$

by considering the matrix $\mathcal{M}(T^*, (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n))$ constructed by taking the conjugate transpose, noting that the modulus of a complex number is equal to its conjugate. Now, since T is normal by our assumption, we must have $\|T\boldsymbol{e}_1\| = \|T^*\boldsymbol{e}\|$, giving

$$|a_{1,1}|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \ldots + |a_{1,n}|^2 \quad \Rightarrow \quad a_{1,2} = \ldots = a_{1,n} = 0.$$

By considering $\|T\boldsymbol{e}_j\| = \|T^*\boldsymbol{e}_j\|$ for $j = 1, \ldots, n$, we finally obtain each $a_{i,j} = 0$ for $i \neq j$, hence the matrix is a diagonal matrix with respect to some basis of V, so by Proposition 7.2.2, V has an orthonormal basis consisting of eigenvectors of T.

□

Lemma 7.3.1

Suppose that $T \in \mathcal{L}(V)$ is self-adjoint. If $\alpha, \beta \in \mathbb{R}$ such that $\alpha^2 < 4\beta$, then

$$T^2 + \alpha T + \beta I$$

is invertible. (A linear map is invertible if its inverse exists.)

Proof: Suppose $\alpha, \beta \in \mathbb{R}$ are such that $\alpha^2 < 4\beta$. Let $\boldsymbol{v}$ be a nonzero vector, then

$$\begin{aligned}
\langle (T^2 + \alpha T + \beta I)\boldsymbol{v}, \boldsymbol{v}\rangle &= \langle T^2\boldsymbol{v}, \boldsymbol{v}\rangle + \alpha\langle T\boldsymbol{v}, \boldsymbol{v}\rangle + \beta\langle \boldsymbol{v}, \boldsymbol{v}\rangle \text{ by additivity in first slot;}\\
&= \langle T\boldsymbol{v}, T^*\boldsymbol{v}\rangle + \alpha\langle T\boldsymbol{v}, \boldsymbol{v}\rangle + \beta\|\boldsymbol{v}\|^2\\
&= \langle T\boldsymbol{v}, T\boldsymbol{v}\rangle + \alpha\langle T\boldsymbol{v}, \boldsymbol{v}\rangle + \beta\|\boldsymbol{v}\|^2\\
&= \|T\boldsymbol{v}\|^2 + \alpha\langle T\boldsymbol{v}, \boldsymbol{v}\rangle + \beta\|\boldsymbol{v}\|^2\\
&\geq \|T\boldsymbol{v}\|^2 - |\alpha\langle T\boldsymbol{v}, \boldsymbol{v}\rangle| + \beta\|\boldsymbol{v}\|^2 \text{ with equality if } \alpha\langle T\boldsymbol{v}, \boldsymbol{v}\rangle < 0\\
&\geq \|T\boldsymbol{v}\|^2 - |\alpha|\|T\boldsymbol{v}\|\|\boldsymbol{v}\| + \beta\|\boldsymbol{v}\|^2 \text{ by the Cauchy-Schwarz inequality;}\\
&= \left(\|T\boldsymbol{v}\| - \frac{|\alpha|\|\boldsymbol{v}\|}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right)\|\boldsymbol{v}\|^2 \text{ by completing the square}\\
&> 0 \text{ since } \beta - \frac{\alpha^2}{4} > 0 \text{ for } \alpha^2 < 4\beta.
\end{aligned}$$

Hence $\langle (T^2 + \alpha T + \beta I)\boldsymbol{v}, \boldsymbol{v}\rangle \neq 0 \ \forall \boldsymbol{v} \in V$, so $\ker(T^2 + \alpha T + \beta I) = \{\boldsymbol{0}\}$, so $T^2 + \alpha T + \beta I$ is injective, implying that it is invertible. □

We now see how this lemma is useful for our purpose by visiting some results from a previous chapter.

Proposition 7.3.1

Suppose $p \in \mathcal{P}(\mathbb{F})$ is a polynomial with degree $m \geq 1$. Let $\lambda \in \mathbb{F}$. Then λ is a root of p iif there is a polynomial $q \in \mathcal{P}(\mathbb{F})$ with degree $m - 1$ such that

$$p(z) = (z - \lambda)q(z)$$

$\forall z \in \mathbb{F}$.

Corollary 7.3.1

Suppose $p \in \mathcal{P}(\mathbb{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct roots in $\mathbb{F}$.

Corollary 7.3.2

Suppose $a_0, \ldots, a_m \in \mathbb{F}$. If

$$a_0 + a_1 z + a_2 z^2 + \ldots + a_m z^m = 0$$

$\forall z \in \mathbb{F}$, then $a_0 = \ldots = a_m = 0$.

Theorem 7.3.2

If $p \in \mathcal{P}(\mathbb{R})$ is a nonconstant polynomial, then p has a unique factorisation (except for the order of the factors) of the form

$$p(x) = c(x - \lambda_1) \ldots (x - \lambda_m)(x^2 + \alpha_1 x + \beta_1) \ldots (x^2 + \alpha_M x + \beta_M),$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbb{R}$ and $(\beta_1, \beta_1), \ldots, (\alpha_M, \beta_M) \in \mathbb{R}^2$ with $\alpha_j^2 < 4\beta_j$ for each j.

Proof: Let $p \in \mathcal{P}(\mathbb{R})$ be a nonconstant polynomial. $p \in \mathcal{P}(\mathbb{C})$ since $\mathbb{R} \subset \mathbb{C}$. If the factorisation of p as an element of $\mathcal{P}(\mathbb{C})$ includes terms of the form $(x - \lambda)$ with λ being a nonreal complex number, then $(x - \bar{\lambda})$ is also a term in the factorisation, so we obtain the $(x^2 + \alpha_j x + \beta_j)$ by multiplying these pair of terms together.

However we need to be careful of the fact that each factor in the pair has no reason to appear the same number of times. To show that this cannot be the case, we prove that after factorising $p(x) = (x^2 + \alpha_j x + \beta_j)q(x)$, $q(x)$ has real coefficients, so we can apply the same result inductively on the degree of p. We need to solve for

$$q(x) = \frac{p(x)}{x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2}$$

$\forall x \in \mathbb{R}$. This implies that $q(x) \in \mathbb{R}$ since $p \in \mathcal{P}(\mathbb{R})$ and $x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2$, so $\operatorname{Im} q(x) = 0$. Let $q(x) = a_0 + a_1 x + \ldots + a_{n-2}x^{n-2}$, then

$$0 = (\operatorname{Im} a_0) + (\operatorname{Im} a_1)x + \ldots + (\operatorname{Im} a_{n-2}x^{n-2})$$

$\forall x \in \mathbb{R}$. □

Lemma 7.3.2

Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Then T has an eigenvalue.

Proof: If V is a complex inner-product space, then T has an eigenvalue regardless of whether T is self-adjoint.

Assume that V is a real inner-product space. Let $n = \dim V$ and choose $\boldsymbol{v} \in V$ with $\boldsymbol{v} \neq \boldsymbol{0}$. Then

$$(\boldsymbol{v}, T\boldsymbol{v}, T^2\boldsymbol{v}, \ldots, T^n\boldsymbol{v})$$

cannot be linearly independent the number of vectors in this list is greater than the dimension of V. So there exists $a_0, \ldots, a_n \in \mathbb{R}$ such that

$$a_0\boldsymbol{v} + a_1T\boldsymbol{v} + \ldots + a_nT^n\boldsymbol{v} = \mathbf{0}$$

where not all $a_j = 0$. We then consider the polynomial $a_0 + a_1x + \ldots + a_nx^n$ which by Theorem 7.3.2 can be factored as

$$c(x^2 + \alpha_1 x + \beta_1)\ldots(x^2 + \alpha_M x + \beta_M)(x - \lambda_1)\ldots(x - \lambda_m),$$

where $c \neq 0$, each $\alpha_j, \beta_j, \lambda_j \in \mathbb{R}$, each $\alpha_j^2 < 4\beta_j$, $m + M \geq 1$, and the equation holds for all real x. So by applying this factorisation to the context of linear maps, we have

$$\begin{aligned}
\mathbf{0} &= a_0\boldsymbol{v} + a_1T\boldsymbol{v} + \ldots + a_nT^n\boldsymbol{v} \\
&= (a_0I + a_1T + \ldots + a_nT^n)\boldsymbol{v} \\
&= c(T^2 + \alpha_1T + \beta_1I)\ldots(T^2 + \alpha_MT + \beta_MI)(T - \lambda_1I)\ldots(T - \lambda_mI)\boldsymbol{v}.
\end{aligned}$$

Since each $\alpha_j^2 < 4\beta_j$ and T is self-adjoint, by Lemma 7.3.1, each $T^2 + \alpha_jT + \beta_jI$ is invertible. Since $c \neq 0$, the equation implies that

$$(T - \lambda_1I\ldots(T - \lambda_mI)\boldsymbol{v}) = \mathbf{0}.$$

Since $\boldsymbol{v} \neq \mathbf{0}$, $\boldsymbol{v}$ belongs to the null space of at least one $(T - \lambda_jI)$, so T has at least one eigenvalue. □

Theorem 7.3.3 Real spectral theorem

Suppose that V is a real inner-product space and $T \in \mathcal{L}(V)$. V has an orthonormal basis consisting of eigenvectors of T iif T is self-adjoint.

Proof:

- ($\Rightarrow$) Suppose that V has an orthonormal basis consisting of eigenvectors of T. With respect to this basis, T has a diagonal matrix. Diagonal matrices is equal to their conjugate transpose, so $T = T^*$ i.e. T is self-adjoint.
- ($\Leftarrow$) Suppose T is self-adjoint. If $\dim V$, then every basis of V is an orthonormal basis and every vector is an eigenvector of T, so the converse holds trivially.

  Let λ be any eigenvalue of T, which exists by Lemma 7.3.2 since T is self-adjoint. Let $\boldsymbol{u} \in V$ denote a corresponding eigenvector with $\|\boldsymbol{u}\| = 1$, U denote the one-dimensional subspace of V consisting of all scalar multiples of $\boldsymbol{u}$. (We can visualise this as the line where all the eigenvectors corresponding to that eigenvalue lies on in the space of V), Then a vector $\boldsymbol{v} \in V$ is in $U^\perp$ iif $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$.

  Suppose that $\boldsymbol{v} \in U^\perp$. Then because T is self-adjoint, we have

$$\langle \boldsymbol{u}, T\boldsymbol{v}\rangle = \langle T\boldsymbol{u}, \boldsymbol{v}\rangle = \langle \lambda\boldsymbol{u}, \boldsymbol{v}\rangle = \lambda\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0,$$

  so $T\boldsymbol{v} \in U^\perp \forall \boldsymbol{v} \in U^\perp$, i.e. $U^\perp$ is invariant under T. So we can define an operator $S \in \mathcal{L}(U^\perp)$ by $S = T|_{U^\perp}$. If $\boldsymbol{v}, \boldsymbol{w} \in U^\perp$, then

$$\langle S\boldsymbol{v}, \boldsymbol{w}\rangle = \langle T\boldsymbol{v}, \boldsymbol{w}\rangle = \langle \boldsymbol{v}, T\boldsymbol{w}\rangle = \langle \boldsymbol{v}, S\boldsymbol{w}\rangle,$$

  which shows that S is self-adjoint. Thus by our inductive hypothesis, there is an orthonormal basis of $U^\perp$ consisting of eigenvectors of S. Every eigenvector of S is an eigenvector of T, so adjoining $\boldsymbol{u}$ to an orthonormal basis of $U^\perp$ consisting of eigenvectors of S gives an orthonormal basis of V consisting of eigenvectors of T. □

Corollary 7.3.3

Suppose that $T \in \mathcal{L}(V)$ is self-adjoint (or that $\mathbb{F} = \mathbb{C}$ and that $T \in \mathcal{L}(V)$ is normal.) Let $\lambda_1, \ldots, \lambda_m$ denote distinct eigenvalues of T. Then

$$V = \ker(T - \lambda_1 I)$$

Proof: If T is self-adjoint, by the real spectral theorem, V has an orthonormal basis consisting of eigenvectors of T. The same conclusion is reached if we assume T is normal for $\mathbb{F} = \mathbb{C}$ by the complex spectral theorem. Then by Proposition 7.2.2, this decomposition holds.

If T is self-adjoint, then it is also normal. Hence by Corollary 7.1.3, each vector in each $\ker(T - \lambda_j I)$ is orthogonal to all vectors in the other subspaces of this decomposition since they corresponding to distinct eigenvalues as defined. □

7.4 Normal operators on real inner-product spaces

The complex spectral theorem describes completely normal operators on complex inner-product spaces. We will do the same for real inner-product spaces in this section.

Lemma 7.4.1

Suppose V is a **two-dimensional** real inner-product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:

1. T is normal but not self-adjoint;
2. the matrix of T with respect to every orthonormal basis of V has the form
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$
with $b \neq 0$;
3. the matrix of T with respect to some orthonormal basis of V has the form
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$
with $b > 0$;

Proof: We prove this in three directions:

(1⟹ 2) Suppose that (1) holds. Let (e_1, e_2) be an orthonormal basis of V. Suppose we have

$$A = \mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

This matrix is with respect to (e_1, e_2), so we have

$$\|Te_1\|^2 = a^2 + b^2 \qquad \text{and} \qquad \|T^* e_1\|^2 = a^2 + c^2$$

recalling that $A^* = \mathcal{M}(T^*, (e_1, e_2))$ is the conjugate transpose of $\mathcal{M}(T, (e_1, e_2))$. Since T is normal, by Proposition 7.1.4, we also have $\|Te_1\| = \|T^* e_1\| \implies b^2 = c^2 \implies c = \pm b$. If $c = b$ then T is self adjoint since

$A = A^*$ (we are dealing with real inner-product spaces), so $c = -b$.

Since T is normal, we must have

$$\begin{aligned} AA^* &= A^*A \\ \begin{pmatrix} a & -b \\ b & d \end{pmatrix}\begin{pmatrix} a & b \\ -b & d \end{pmatrix} &= \begin{pmatrix} a & b \\ -b & d \end{pmatrix}\begin{pmatrix} a & -b \\ b & d \end{pmatrix} \\ = \begin{pmatrix} a^2+b^2 & ab-bd \\ ab-bd & b^2+d^2 \end{pmatrix} &= \begin{pmatrix} a^2+b^2 & -ab+bd \\ -ab+bd & b^2+d^2 \end{pmatrix} \\ ab-bd &= -ab+bd \\ 2ab-2bd &= 0 \\ 2b(a-d) &= 0 \\ a &= d \because b \neq 0 \end{aligned}$$

($2 \implies 3$) Suppose that (2) holds. We choose an orthonormal basis $(\boldsymbol{e}_1, \boldsymbol{e}_2)$ of V. The matrix of T with respect to this basis has the form given, with $b \neq 0$. If $b > 0$, then (3) holds. If $b < 0$, then we can choose $(\boldsymbol{e}_1, -\boldsymbol{e}_2)$ to be the basis, then we obtain a matrix in the same form with $-b > 0$, so (3) holds.

($3 \implies 1$) Suppose that (3) holds. The matrix of T is clearly not equal to its transpose so T is not self-adjoint. By computing matrices of TT^* and T^*T, we obtain the same matrices (in the first section we constructed that matrix by forcing the two to be equal), so T is normal and (1) holds.

□

Proposition 7.4.1

Suppose $T \in \mathcal{L}(V)$ is normal and U is a subspace of V that is invariant under T, then

1. $U^\perp$ is invariant under T;
2. U is invariant under T^*;
3. $(T|_U)^* = (T^*)|_U$;
4. $T|_U$ is a normal operator on U;
5. $T|_{U^\perp}$ is a normal operator on $U^\perp$;

where $T|_U$ denotes the restriction of the operator T to the subspace U.

Proof: 1. Let $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$ be an orthonormal basis of U. Extend to an orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$ of V, which we know we can do by Corollary 6.3.3. Since U is invariant under T, each $T\boldsymbol{e}_j \in U$ can be expressed as a linear combination of $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m)$. We write the matrix of T with respect to the basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m, \boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$ as

$$\mathcal{M}(T) = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix};$$

where $A \in M_{mm}$, $B \in M_{mn}$, $C \in M_{nn}$, and $0 \in M_{nm}$ and is a zero matrix because U is invariant under T. For each $j \in \{1, \ldots, m\}$, $\|T\boldsymbol{e}_j\|^2$ is the sum of squares of the absolute values of the entries in the j^{th} column of A, so

$$\sum_{j=1}^{m} \|T\boldsymbol{e}_j\|^2$$

is the sum of squares of the absolute values of the entries of A (Noting that we also take the squares of the zero matrix, but it adds no). Similar for T^*,

$$\sum_{j_1}^{m} \|T^*\boldsymbol{e}_j\|^2$$

is the sum of the squares of the absolute values of the entries of A and B, noting that the matrix represented by T^* is the conjugate transpose of T. C is not included because we are only applying T^* to each $\boldsymbol{e}_j$ and not $\boldsymbol{e}_f$.

Now, since T is normal, by Proposition 7.1.4, we have

$$\sum_{j=1}^{m} \|T\boldsymbol{e}_j\|^2 = \sum_{j_1}^{m} \|T^*\boldsymbol{e}_j\|^2$$

which implies that B must be a zero matrix. So we have

$$\mathcal{M}(T) = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix},$$

which shows that $T\boldsymbol{f}_k \in \operatorname{span}(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$ for each k. Since $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$ is a basis of $U^\perp$ $(U \oplus U^\perp = V)$, so $T\boldsymbol{v} \in U^\perp$ when $\boldsymbol{v} \in U^\perp$, i.e. $U^\perp$ is invariant under T.

2. Note that we now have
$$\mathcal{M}(T) = \begin{pmatrix} C & 0 \\ 0 & A \end{pmatrix},$$
which shows that $T^*\boldsymbol{e}_j \in \operatorname{span}(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m) \Rightarrow T\boldsymbol{u} \in U$ for $\boldsymbol{u} \in U$, U is invariant under T^*.

3. Let $S = T|_U$. Fix $\boldsymbol{v} \in U$, then
$$\langle S\boldsymbol{u}, \boldsymbol{v}\rangle = \langle T\boldsymbol{u}, \boldsymbol{v}\rangle$$
$$\langle \boldsymbol{u}, S^*\boldsymbol{v}\rangle = \langle \boldsymbol{u}, T^*\boldsymbol{v}\rangle$$
$\forall \boldsymbol{u} \in U$. Then since $T^*\boldsymbol{v}$ by (2), we have $S^*\boldsymbol{v} = T^*\boldsymbol{v}$, so $(T|_U)^* = (T^*)|_U$.

4. Since T is normal, $TT^* = T^*T$. By (3), we have $(T|_U)^* = (T^*)|_U$. Then,
$$(T|_U)(T|_U)^* = (T|_U)(T^*)|_U = (TT^*)|_U = (T^*T)|_U = (T^*)|_U(T|_U) = (T|_U)^*(T|_U).$$

5. In (4) we showed that the restriction of T to any invariant subspace is normal. Since U^T is invariant under T, $T|_{U^\perp}$ is normal.

□

Definition 7.4.1: Block diagonal matrix

A **block diagonal matrix** is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where $A_1, \ldots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0.

If A and B are block diagonal matrices of the form

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}, \quad B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix},$$

where A_j and B_j both has the same size for $j = 1, \ldots, m$, then AB is a block diagonal matrix with

$$AB = \begin{pmatrix} A_1 B_1 & & 0 \\ & \ddots & \\ 0 & & A_m B_m \end{pmatrix}.$$

We can view diagonal matrices as a special case of block diagonal matrices by letting each matrix have size 1×1, or some other funky configuration like taking the entire matrix to be one block. Hence, saying that an operator has a block diagonal matrix with respect to some basis doesn't tell us much unless we know the size of the blocks.

Theorem 7.4.1 Invariant subspaces on real vector spaces

Every operator on a finite-dimensional, nonzero, real vector space has an invariant subspace of dimension 1 or 2.

Theorem 7.4.2

Suppose that V has a real inner-product space and $T \in \mathcal{L}(V)$. Then T is normal iif there is an orthonormal basis of V with respect to which T has a block diagonal matrix where each block is a 1×1 or 2×2 matrix with the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b > 0$.

Proof: We prove the statement from both directions:

- ($\Leftarrow$) Suppose that there is an orthonormal basis of V such that the matrix of T is a block diagonal matrix where each block is a 1×1 matrix or a 2×2 matrix with the given form. With respect to this basis, the matrix of T commutes with the matrix of T^* (the transpose), so T commutes with T^*, so T is normal.

- ($\Rightarrow$) Suppose that T is normal. We will prove the converse statement inductively on the dimension of V. The base case holds if $\dim V = 1$ trivially or if $\dim V = 2$ (if T is self-adjoint, use the real spectral theorem, else use Lemma 7.4.1.)

Assume that $\dim V > 2$ and that the result holds on vector space with smaller dimension. Let U be a subspace of V of dimension 1 that is invariant under T is such a subspace exists (i.e. if T has a nonzero eigenvector, let U be the span of this eigenvector).

1. Let U be a subspace of V of dimension 1 that is invariant under T is such a subspace exists (i.e. if T has a nonzero eigenvector, let U be the span of this eigenvector), then choose a vector in U with norm 1 to be an orthonormal basis of U, and the matrix of $T|_U$ is a 1×1 matrix.
2. If no such subspace exists, let U be a subspace of V of dimension 2 that is invariant under T (an invariant subspace of dimension 1 or 2 always exists by Theorem 7.5.1), then $T|_U$ is normal by Proposition 7.4.1 but not self-adjoint otherwise $T|_U$ and subsequently T will havee a nonzero eigenvector by Lemma 7.3.2. Then, by Lemma 7.4.1 we can choose an orthonormal basis of U with respect to which the matrix of $T|_U$ has the form

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

for all $b > 0$.

By Proposition 7.4.1, $U^\perp$ is also invariant under T and $T|_{U^\perp}$, and is a normal operator on $U^\perp$. Hence by out induction hypothesis, there is an orthonormal basis of $U^\perp$ with respect to which the matrix of $T|_{U^\perp}$ has the desired form. Adjoin this basis to the basis of U, we obtain an orthonormal basis of V with respect to which the matrix of T has the desired form.

□

7.5 Positive operators

Definition 7.5.1: Positive operators

An operator $T \in \mathcal{L}(V)$ is **positive** if T is self-adjoint and

$$\langle T\boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$$

$\forall \boldsymbol{v} \in V$. If V is a complex vector space, then the condition that T be self-adjoint can be dropped.

Note:-

Every orthogonal projection is positive (TBV).

Definition 7.5.2: Square root operators

An operator S is a **square root** of an operator if $S^2 = T$.

Example 7.5.1 (Square root operators)

Let $T \in \mathcal{L}(\mathbb{F}^3)$ such that $T(z_1, z_2, z_3) = (z_3, 0, 0)$, then the operator $S \in \mathcal{L}(\mathbb{F}^3)$ such that $S(z_1, z_2, z_3) = (z_2, z_3, 0)$ is a square root of T, which can verify:

$$S^2(z_1, z_2, z_3) = S(z_2, z_3, 0) = (z_3, 0, 0) = T(z_1, z_2, z_3).$$

Theorem 7.5.1

Let $T \in \mathcal{L}(V)$. The following are equivalent:

1. T is positive;
2. T is self-adjoint and all the eigenvalues of T are nonnegative;
3. T has a positive square root;
4. T has a self-adjoint square root;
5. there exists an operator $S \in \mathcal{L}(V)$ such that $T = S^*S$.

Proof: We will prove (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (1).

(1 $\Rightarrow$ 2) Suppose that (1) holds, so by definition $\langle T\boldsymbol{v}, \boldsymbol{v}\rangle \geq 0$ for all $\boldsymbol{v} \in V$ and T is self-adjoint. All that remains is to show that all eigenvalues of T are nonnegative. Suppose λ is an eigenvalue of T and $\boldsymbol{u} \in V$ be a nonzero corresponding eigenvector of T. Then,

$$\begin{aligned}\langle T\boldsymbol{u}, \boldsymbol{u}\rangle &\geq 0\\ \langle \lambda\boldsymbol{u}, \boldsymbol{u}\rangle &\geq 0\\ \lambda\langle \boldsymbol{u}, \boldsymbol{u}\rangle &\geq 0\\ \lambda\|\boldsymbol{u}\|^2 &\geq 0\end{aligned}$$

So $\lambda \geq 0$ since $\|\boldsymbol{u}\|^2 \geq 0$.

(2 $\Rightarrow$ 3) Suppose that (2) holds, then by the complex and real spectral theorem, there is an orthogonal basis $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$ of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of T corresponding to $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n$, such that each λ_j is a nonnegative number. Set $S \in \mathcal{L}(V)$ as

$$S\boldsymbol{e}_j = \sqrt{\lambda_j}\boldsymbol{e}_j,$$

for $j = 1, \dots, n$. We observe that S is a positive operator:

$$\begin{aligned}\langle S\boldsymbol{e}_j, \boldsymbol{e}_j\rangle &= \langle \sqrt{\lambda_j}\boldsymbol{e}_j, \boldsymbol{e}_j\rangle\\ \langle \boldsymbol{e}_j, S^*\boldsymbol{e}_j\rangle &= \sqrt{\lambda_j}\langle \boldsymbol{e}_j, \boldsymbol{e}_j\rangle\\ &= \langle \boldsymbol{e}_j, \sqrt{\lambda_j}\boldsymbol{e}_j\rangle = \langle \boldsymbol{e}_j, S\boldsymbol{e}_j\rangle\end{aligned}$$

So $S = S^*$, S is adjoint, and $\langle S\boldsymbol{e}_j, \boldsymbol{e}_j\rangle \geq 0$ since $\sqrt{\lambda_j} \geq 0$. Furthermore, $S^2\boldsymbol{e}_j = \lambda_j\boldsymbol{e}_j = T\boldsymbol{e}_j$ for each j, implying that $S^2 = T$, so S is a positive square root of T, so (3) holds.

(3 $\Rightarrow$ 4) Suppose that (3) holds, so T has a positive square root, which must be self-adjoint since all positive operators are self-adjoint by definition.

(4 $\Rightarrow$ 5) Suppose that (4) holds, let $S \in \mathcal{L}(V)$ be the self-adjoint square root. So $T = S^2 = SS = SS^*$, so (5) holds.

(5 $\Rightarrow$ 1) Suppose that (5) holds, then

$$\begin{aligned}\langle T\boldsymbol{v}, \boldsymbol{v}\rangle &= \langle S^*S\boldsymbol{v}, \boldsymbol{v}\rangle\\ &= \langle S\boldsymbol{v}, S\boldsymbol{v}\rangle = \langle \boldsymbol{v}, S^*S\boldsymbol{v}\rangle = \langle \boldsymbol{v}, T\boldsymbol{v}\rangle\\ &= \|S\boldsymbol{v}\|^2 \geq 0\end{aligned}$$

□

The following proposition is anologous to how every nonnegative number has a unique nonnegative square root, which allows us extend the notation of square roots to operators.

Proposition 7.5.1

Every positive operator on V has a unique positive square root.

Proof: By Theorem 7.5.1, we already have that every positive operator on V has a positive square root. All that remains is to show uniqueness.

Suppose $T \in \mathcal{L}(V)$ is positive. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, by Theorem 7.5.1, all these numbers are nonnegative. Since T is positive, T is self-adjoint, so we have

$$V = \ker(T - \lambda_1 I) \oplus \ldots \oplus \ker(T - \lambda_m I);$$

by Corollary 7.3.1. Now suppose $S \in \mathcal{L}(V)$ is a positive square root of T. Suppose α is an eigenvalue fo S. If $\boldsymbol{v} \in \ker(S - \alpha I)$, then $S\boldsymbol{v} = \alpha\boldsymbol{v}$, which implies that

$$T\boldsymbol{v} = S^2\boldsymbol{v} = \alpha^2\boldsymbol{v}, \quad \ldots (1)$$

so $\boldsymbol{v} \in \ker(T - \alpha^2 I)$. So α^2 is an eigenvalue of T, which means that α^2 must equal some λ_j, i.e. $\alpha = \sqrt{\lambda_j}$ for some j. Furthermore, equation (1) implies that

$$\ker(S - \sqrt{\lambda_j} I) \subset \ker(T - \lambda_j I).$$

Also, we showed that the only possible eigenvalues for S are $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_m}$. Since S is self-adjoint, which again implies

$$V = \ker(S - \sqrt{\lambda_1} I) \oplus \ldots \oplus \ker(S - \sqrt{\lambda_m} I);$$

which implies that

$$\ker(S - \sqrt{\lambda_j} I) = \ker(T - \lambda_j I)$$

for each j. Hence on $\ker(T - \lambda_j I)$, operator S effectively just multiplies by $\sqrt{\lambda_j}$, so S is uniquely determined by T. □

7.6 Isometries

Definition 7.6.1: Isometry

An operator $S \in \mathcal{L}(V)$ is an **isometry** if

$$\|S\boldsymbol{v}\| = \|\boldsymbol{v}\|$$

$\forall \boldsymbol{v} \in V$, i.e. S preserves the norm of $\boldsymbol{v}$.

Example 7.6.1 ()

Generally, suppose $\lambda_1, \ldots, \lambda_n$ are scalars with absolute value 1 and $S \in \mathcal{L}(V)$ satisfies $S(\boldsymbol{e}_j) = \lambda_j \boldsymbol{e}_j$ for some orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of V. Suppose $\boldsymbol{v} \in V$, then we have

$$\boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \ldots + \langle \boldsymbol{v}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n$$

and

$$\|\boldsymbol{v}\|^2 = |\langle \boldsymbol{v}, \boldsymbol{w}_1 \rangle|^2 + \ldots + |\langle \boldsymbol{v}, \boldsymbol{e}_n \rangle|^2. \tag{7.1}$$

by Theorem 6.3.1. Applying S to both sides, we have

$$\begin{aligned} S\boldsymbol{v} &= \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle S\boldsymbol{e}_1 + \ldots + \langle \boldsymbol{v}, \boldsymbol{e}_n \rangle S\boldsymbol{e}_n \\ &= \lambda_1 \langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \ldots + \lambda_n \langle \boldsymbol{v}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n. \end{aligned}$$

Since each $|\lambda_k| = 1$, we have by the Pythagorean theorem

$$\|S\boldsymbol{v}\|^2 = |\langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle|^2 + \ldots + |\langle \boldsymbol{v}, \boldsymbol{e}_n \rangle|^2. \tag{7.2}$$

By comparing (7.1) and (7.2), we have $\|\boldsymbol{v}\| = \|S\boldsymbol{v}\|$, so S is an isometry.

Theorem 7.6.1

Suppose $S \in \mathcal{L}(V)$. Then the following are equivalent:

1. S is an isometry;
2. $\langle S\boldsymbol{u}, S\boldsymbol{u} \rangle = \langle \boldsymbol{u}, \boldsymbol{v} \rangle \ \forall \boldsymbol{u}, \boldsymbol{v} \in V$;
3. $S^*S = I$;
4. $(S\boldsymbol{e}_1, \ldots, S\boldsymbol{e}_n)$ is orthonormal whenever $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is an orthonormal list of vectors in V;
5. $\exists$ an orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of V such that $(S\boldsymbol{e}_1, \ldots, S\boldsymbol{e}_n)$ is orthonormal;
6. S^* is an isometry;
7. $\langle S^*\boldsymbol{u}, S^*\boldsymbol{v} \rangle = \langle \boldsymbol{u}, \boldsymbol{v} \rangle \ \forall \boldsymbol{u}, \boldsymbol{v} \in V$;
8. $SS^* = I$;
9. $(S^*\boldsymbol{e}_1, \ldots, S^*\boldsymbol{e}_n)$ is orthonormal whenever $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is an orthonormal list of vectors in V;
10. $\exists$ an orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of V such that $(S^*\boldsymbol{e}_1, \ldots, S^*\boldsymbol{e}_n)$ is orthonormal.

Proof: We prove that (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (6):

(1 $\Rightarrow$ 2) Suppose that (1) holds. If V is a real inner-product space, then $\forall \boldsymbol{u}, \boldsymbol{v} \in V$ we have

$$\begin{aligned} \langle S\boldsymbol{u}, S\boldsymbol{v} \rangle &= \frac{1}{4}(\|S\boldsymbol{u} + S\boldsymbol{v}\|^2 - \|S\boldsymbol{u} - S\boldsymbol{v}\|^2) \\ &= \frac{1}{4}(\|SS(\boldsymbol{u} + \boldsymbol{v})\|^2 - \|S(\boldsymbol{u} - \boldsymbol{v})\|^2) \\ &= \frac{1}{4}(\|\boldsymbol{u} + \boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|^2) \text{ since } S \text{ is an isometry;} \\ &= \langle \boldsymbol{u}, \boldsymbol{v} \rangle. \end{aligned}$$

(2 $\Rightarrow$ 3) Suppose that (2) holds, then

$$\langle (S^*S - I)\boldsymbol{u}, \boldsymbol{v} \rangle = \langle S^*S\boldsymbol{u}, \boldsymbol{v} \rangle - \langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle S\boldsymbol{u}, S\boldsymbol{v} \rangle - \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$$

for every $\boldsymbol{u}, \boldsymbol{v} \in V$. Taking $\boldsymbol{v} = (S^*S - I)\boldsymbol{u}$, we see that $S^*S - I = 0$. Hence $S^*S = I$.

(3 ⟹ 4) Suppose that (3) holds. Suppose that $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ is an orthonormal list of vectors in V, then

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = \langle S^*S\boldsymbol{e}_j, \boldsymbol{e}_k \rangle = \langle S\boldsymbol{e}_j, S\boldsymbol{e}_k \rangle.$$

Hence $\langle S\boldsymbol{e}_j, S\boldsymbol{e}_k \rangle$ is orthogonal for any $j, k = 1, \ldots, n$, $j \neq k$. If $j = k$, then the expression is equal to 1, so $(S\boldsymbol{e}_1, \ldots, S\boldsymbol{e}_n)$ is orthonormal.

(4 ⟹ 5) Suppose that (4) holds. Let the standard basis of V be $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$, which is orthonormal, so $(S\boldsymbol{e}_1, \ldots, S\boldsymbol{e}_n)$ is orthonormal.

(5 ⟹ 6) Suppose that (5) holds. Let $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ be an orthonormal basis of V such that $(S\boldsymbol{e}_1, \ldots, S\boldsymbol{e}_n)$ is orthonormal. If $\boldsymbol{v} \in V$, then we have

$$\begin{aligned}
\|S\boldsymbol{v}\|^2 &= \|S(\langle \boldsymbol{e}, \boldsymbol{e}_1 \rangle \boldsymbol{e}_1 + \ldots + \langle \boldsymbol{v}, \boldsymbol{e}_n \rangle \boldsymbol{e}_n)\|^2 \\
&= \|\langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle S\boldsymbol{e}_1 + \ldots + \langle \boldsymbol{v}, \boldsymbol{e}_n \rangle S\boldsymbol{e}_n\|^2 \\
&= |\langle \boldsymbol{v}, \boldsymbol{e}_1 \rangle|^2 + \ldots + |\langle \boldsymbol{v}, \boldsymbol{e}_n \rangle|^2 \\
&= \|\boldsymbol{v}\|^2 \\
\|S\boldsymbol{v}\| &= \|\boldsymbol{v}\|,
\end{aligned}$$

with first and third equality by Theorem 6.3.1 and second equality by the Pythagorean theorem (since $\|S\boldsymbol{e}_j\| = 1$).

Replacing S with S^* gives (6) to (10). To complete the proof, we show that one of the conditions in (1-5) is equivalent to (6-10). We show that (3) is equivalent to (8):

Suppose that $S^*S = I$, then for all $\boldsymbol{u}, \boldsymbol{v} \in V$,

$$\begin{aligned}
\langle S^*S\boldsymbol{u}, \boldsymbol{v} \rangle &= \langle S\boldsymbol{u}, S\boldsymbol{v} \rangle \\
\langle \boldsymbol{u}, (S^*S)^*\boldsymbol{v} \rangle &= \langle \boldsymbol{u}, S^*S\boldsymbol{v} \rangle.
\end{aligned}$$

So S^*S is self-adjoint, so it is normal, so $S^*S = SS^* = I$. The same result follows for the converse statement. □

From this theorem, we see that every isometry is normal, so we can use the characterizations of normal operators to give descriptions of isometries.

Theorem 7.6.2

Suppose V is a complex inner-product space and $S \in \mathcal{L}(V)$. Then S is an isometry iif there is an orthonormal basis of V consisting of eigenvectors of S all of whose corresponding eigenvalues have absolute value 1.

Proof: (⟸) This is shown in Example 7.6.1.

(⟹) Suppose that S is an isometry, so it is normal. By the complex spectral theorem, there is an orthonormal basis $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ of V consisting of eigenvectors of S. All that remains is to show that all the corresponding eigenvalues have absolute value 1. For $j = \{1, \ldots, n\}$, let λ_j be the eigenvalue corresponding to $\boldsymbol{e}_j$, then

$$|\lambda_j| = \|\lambda_j \boldsymbol{e}_j\| = \|S\boldsymbol{e}_j\| = \|\boldsymbol{e}_j\| = 1.$$

□

Theorem 7.6.3

Suppose that V is a real inner-product space and $S \in \mathcal{L}(V)$. S is an isometry iif there is an orthonormal basis of V with respect to which S has a block diagonal matrix where each block on the diagonal is either a 1×1 matrix containing 1 or -1 or

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in M_{22},$$

with $\theta \in (0, \pi)$.

Proof: We prove the statement from both directions:

- ($\Rightarrow$) Suppose that S is an isometry, so S is normal. By Theorem 7.4.2, there is an orthonormal basis B of V such that with respect to this basis, S has a block diagonal matrix, where each block is a 1×1 matrix or a 2×2 matrix with the form

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

  with $b > 0$. We now need to show that the 1×1 matrix contains either 1 or -1 and that the 2×2 matrix is of the form given.

  If λ is an entry in a 1×1 along the diagonal of the matrix of S with respect to basis B, then there is a basis vector $\boldsymbol{e}_j$ with $S\boldsymbol{e}_j = \lambda \boldsymbol{e}_j$. Since S is an isometry, $|\lambda| = 1$ for $\|S\boldsymbol{e}_j\| = \|\lambda \boldsymbol{e}_j\|$, so $\lambda = \pm 1$.

  For the 2×2 case, there are basis vectors $\boldsymbol{e}_j, \boldsymbol{e}_{j+1}$ such that

  $$S\boldsymbol{e}_j = a\boldsymbol{e}_j + b\boldsymbol{e}_{j+1}.$$

  So, we have $1 = \|\boldsymbol{e}_j\|^2 = \|S\boldsymbol{e}_j\|^2 = \|a\boldsymbol{e}_j + b\boldsymbol{e}_{j+1}\| = a^2 + b^2$ by the Pythagorean theorem. Since we also have $b > 0$, there exists a number $\theta \in (0, \pi)$ such that $a = \cos\theta$ and $b = \sin\theta$.

- ($\Leftarrow$) Suppose that there is an orthonormal basis of V with respect to which the matrix of S has the form given. Hence there is a direct sum decomposition

  $$V = U_1 \oplus \ldots \oplus U_m,$$

  where each U_j is a subspace of V of dimension 1 or 2. Furthermore, any two vectors in distinct U's are orthogonal, and each $S|_{U_j}$ is an isometry mapping U_j into U_j. If $\boldsymbol{v} \in V$, we can write

  $$\boldsymbol{v} = \boldsymbol{u}_1 + \ldots + \boldsymbol{u}_m,$$

  where each $\boldsymbol{u}_j \in U_j$. Applying S to the equation above and then taking the norms gives

  $$\begin{aligned} \|S\boldsymbol{v}\|^2 &= \|S\boldsymbol{u}_1 + \ldots + S\boldsymbol{u}_m\|^2 \\ &= \|S\boldsymbol{u}_1\|^2 + \ldots + \|S\boldsymbol{u}_m\|^2 \\ &= \|\boldsymbol{u}_1\|^2 + \ldots + \|\boldsymbol{u}_m\|^2 \\ &= \|\boldsymbol{v}\|^2. \end{aligned}$$

  Hence S is an isometry.

□

7.7 Polar and singular-value decompositions.

Note:-

There is an analogy between $\mathbb{C}$ and $\mathcal{L}(V)$, where complex numbers corresponds to linear operators, complex conjugates correspond to adjoint operators, real numbers correspond to self-adjoint operators (the conjugate of a real number is equal to itself), and the non-negative numbers correspond to positive operators.

Consider the unit circle in the complex plane, which consists of $z \in \mathbb{C}$ such that $|z| = \bar{z}z = 1$. In the context of linear operators, this corresponds to the condition $T^*T = I$, then the unit circle in $\mathbb{C}$ would correspond to the set of all isometries.

We can write each complex number z except 0 as

$$z = \left(\frac{z}{|z|}\right)|z| = \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z},$$

where the first factor is an element of the unit circle. Following our analogy, turns out it is true that we can do something similar for linear operators:

Theorem 7.7.1

If $T \in \mathcal{L}(V)$, then $\exists$ an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Proof: (very awesome proof) □

This theorem states that each operator on V is the product of an isometry and a positive operator. Suppose $T = S\sqrt{T^*T}$ is the polar decomposition of $T \in \mathcal{L}(V)$, where S is an isometry, then there is an orthonormal basis of V with respect to which S has a diagonal matrix (if $\mathbb{F} = \mathbb{C}$) or a block diagonal matrix with blocks of size at most 2×2 (if $\mathbb{F} = \mathbb{R}$), and there is an orthonormal basis of V with respect to which $\sqrt{T^*T}$ has a diagonal matrix.

Note:-

There may not exist an orthonormal basis that simultaneously fulfills both these roles.

Definition 7.7.1: Singular values

Suppose $T \in \mathcal{L}(V)$. The **singular values** of T are the eigenvalues of $\sqrt{T^*T}$ with each eigenvalue λ repeated nullity$(\sqrt{T^*T} - \lambda I)$ times.

The singular values of T are nonnegative since they are the eigenvalues of the positive operator $\sqrt{T^*T}$.

Example 7.7.1

Define $T \in \mathcal{L}(\mathbb{F}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4),$$

then $T^*T(z_1, z_2, z_3, z_4) = T^*(0, 3z_1, 2z_4) = (9z_1, 4z_2, 0, 9z_4)$ and

$$\sqrt{T^*T}(z_1, z_2, z_3, z_4) = (3z_1, 2z_2, 0, 3z_4),$$

and we see that the eigenvalues of $\sqrt{T^*T}$ are 3,2,0. Clearly

$$\text{nullity}(\sqrt{T^*T} - 3I) = 2, \quad \text{nullity}(\sqrt{T^*T} - 2I) = 1, \quad \text{nullity}\,\sqrt{T^*T = 1},$$

noting that $\sqrt{T^*T} - 3I = (0, -z_2, -3z_3, 0)$. So the singular values of T are 3,3,2,0. -3 and 0 are the only eigenvalues of T by solving

$$\begin{aligned} T(z_1, z_2, z_3, z_4) &= \lambda(z_1, z_2, z_3, z_4) \\ (0, 3z_1, 2z_2, -3z_4) &= (\lambda z_1, \lambda z_2, \lambda z_3, \lambda z_4) \end{aligned}$$

Comparing the first entries, we have either $\lambda = 0$ or $z_1 = 0$. If $\lambda = 0$ then $z_1 = z_2 = z_4 = 0$, so the eigenspace is $\text{span}(0, 0, 1, 0)$. If $z_1 = 0$ then $z_2 = z_3 = 0$ and $\lambda = -3$ with eigenspace $\text{span}(0, 0, 0, 1)$

Theorem 7.7.2 Singular-value decomposition

Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \ldots, s_n$. Then there exist orthonormal bases $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)$ and $(\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)$ of V such that

$$T\boldsymbol{v} = s_1\langle \boldsymbol{v}, \boldsymbol{e}_1\rangle \boldsymbol{f}_1 + \ldots + s_n\langle \boldsymbol{v}, \boldsymbol{e}_n\rangle \boldsymbol{f}_n$$

for every $n \in V$.

Chapter 8

Matrix decompositions

This chapter focuses on the singular value decomposition from the more practical point of view, where instead of linear operators, we use matrices.

8.1 Matrix-matrix products

Usually we think of the product of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as the rows of $\boldsymbol{A}$ times the columns of $\boldsymbol{B}$ giving the corresponding entry in the product matrix. Here, for reasons that will become apparent later, we instead think of them as a sum of rank 1 matrices, which we obtain by multiplying the columns of $\boldsymbol{A}$ with the rows of $\boldsymbol{B}$.

Example 8.1.1

$$\begin{pmatrix} 1 & 1 & 2 \\ 2 & 3 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \sum_{i=1}^{3} \boldsymbol{a}_i \boldsymbol{b}_i^T = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 3 \end{pmatrix} \begin{pmatrix} 4 & 5 & 5 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \end{pmatrix} \begin{pmatrix} 7 & 8 & 9 \end{pmatrix}$$

8.2 Spectral/eigen decomposition

8.2.1 Eigenvalues and eigenvectors

Definition 8.2.1: Eigenvalue and eigenvector

Let $\boldsymbol{A} \in M_{nn}$. The vector $\boldsymbol{x} \in \mathbb{R}^n$ is an eigenvector of $\boldsymbol{A}$ corresponding to eigenvalue λ iif

$$A\boldsymbol{x} = \lambda \boldsymbol{x}.$$

Note that we only discuss the real spectral theorem here, which then restricts us to only discussing symmetric matrices. For the specifics, refer to the previous chapter.

Proposition 8.2.1

If $\boldsymbol{A}$ is symmetric ($\boldsymbol{A}^T = \boldsymbol{A}$), then the eigenvalues and eigenvectors of A are real.

Proposition 8.2.2

If $\boldsymbol{A}$ is an $n \times n$ symmetric matrix, then its determinant is the product of its eigenvalues, i.e.

$$\det(\boldsymbol{A}) = \lambda_1 \dots \lambda_n.$$

Hence, we have

$$\boldsymbol{A} \text{ is invertible } \Leftrightarrow \det(\boldsymbol{A}) \neq 0 \Leftrightarrow \lambda_i \neq 0 \, \forall i \Leftrightarrow \boldsymbol{A} \text{ is of full rank}$$

8.2.2 Spectral decomposition

Proposition 8.2.3 Spectral/eigen-decomposition

Any symmetric matrix $\boldsymbol{A} \in M_{nn}$ can be written as

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T = \sum_{i=1}^{n} \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^T,$$

where $\boldsymbol{\Lambda} = \operatorname{diag}\{\lambda_1, \dots, \lambda_n\} \in M_{nn}$ consists of the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{Q}$ is orthogonal matrix whose columns are unit eigenvectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ of $\boldsymbol{A}$.

Note:-

By convention, we arrange the eigenvalues so that λ_1 is the largest eigenvalue and λ_n is the smallest.

Corollary 8.2.1

The rank of a symmetric matrix is equal to the number of non-zero eigenvalues (counting accoding to their multiplicities).

Proof: If r is the number of non-zero eigenvalues of $\boldsymbol{A}$, then we have

$$\boldsymbol{A} = \sum_{i=1}^{r} \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^T.$$

Each $\boldsymbol{q}_i \boldsymbol{q}_i^T$ is a rank 1 matrix, with column space equal to the span of $\boldsymbol{q}_i$. As the $\boldsymbol{q}_i$ are orthogonal the column spaces of each $\boldsymbol{q}_i \boldsymbol{w}_1^T$ are orthogonal, and their union is a vector space of dimension r. Hence the rank of $\boldsymbol{A}$ is r. □

Definition 8.2.2: Positive definite matrices

A **positive definite matrix** is a symmetric matrix with all positive eigenvalues.

Note that this definition makes sense since all eigenvalues of symmetric matrices are real. Now, since eigenvalues are rather cumbersome to compute, we may want a simpler way to tell whether a matrix is positive definite: A matrix is positive definite if it is symmetric and **all its pivots are positive**.

Lemma 8.2.1

Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix with real eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$. Then $\boldsymbol{A}$ is positive definite iif $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0$ and positive semi-definite iif $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq 0$ for any non-zero vector $\boldsymbol{x}$.

Proof: We will prove for the positive definite case. For the positive semi-definite case, replace the signs in the initial assumption and everything follows similarly.

- ($\Rightarrow$) Suppose that $\boldsymbol{A}$ is positive definite. Since $\boldsymbol{A}$ is symmetric with real eigenvalues, we can write it as

  $$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T.$$

  Let $\boldsymbol{x}$ be a non-zero vector, then $\boldsymbol{y} = \boldsymbol{Q}^T\boldsymbol{x} \neq \boldsymbol{0}$ since $\dim\ker(\boldsymbol{Q}) = 0$ as Q^T has inverse Q. Thus, returning to our target expression:

  $$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y} = \sum_i^n \lambda_i y_i^2 > 0$$

  since each $\lambda_i > 0$ and $\boldsymbol{y} \neq \boldsymbol{0}$.

- ($\Leftarrow$) Suppose that we have $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ for any non-zero $\boldsymbol{x}$, let $\boldsymbol{x}$ be a unit eigenvector of $\boldsymbol{A}$ corresponding to λ_n, then

  $$0 < \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^T\lambda_n\boldsymbol{x} = \lambda_n\|\boldsymbol{x}\|^2 = \lambda_n.$$

□

8.2.3 Matrix square roots

By the (real) spectral theorem, we see that if $\boldsymbol{A}$ is a symmetric positive semi-definite matrix, then for any integer p, we have

$$\boldsymbol{A}^p = \boldsymbol{Q}\boldsymbol{\Lambda}^p\boldsymbol{Q}^T.$$

In addition, if $\boldsymbol{A}$ is positive definite, then

$$\boldsymbol{A}^{-1} = \boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T$$

since $\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T = \boldsymbol{I}$, where $\boldsymbol{\Lambda}^{-1} = \text{diag}\{\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}\}$. We will then abuse this notation and extend it to matrix square roots, and verify that it fulfills our intuition of how such a square root should behave like:

Definition 8.2.3: Matrix square root

Let $\boldsymbol{A}$ be positive semi-definite with nonnegative eigenvalues. $\boldsymbol{A}^{\frac{1}{2}}$ is a matrix square root of $\boldsymbol{A}$ with

$$\boldsymbol{A}^{\frac{1}{2}} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^T$$

where $\text{diag}\{\lambda_1^{\frac{1}{2}}, \dots, \lambda_n^{\frac{1}{2}}\}$.

Now consider

$$\begin{aligned}\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{A}^{\frac{1}{2}} &= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}(\boldsymbol{Q}^T\boldsymbol{Q})\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^T \\ &= \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}^T \\ &= \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T = \boldsymbol{A}.\end{aligned}$$

If now instead $\boldsymbol{A}$ is positive definite, then all of its eigenvalues are positive and non-zero and hence we can define

$$\boldsymbol{A}^{-\frac{1}{2}} = \boldsymbol{Q}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{Q}^T$$

with $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1^{-\frac{1}{2}}, \dots, \lambda_n^{-\frac{1}{2}}\}$. Then we have $\boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{A}^{-\frac{1}{2}} = \boldsymbol{A}^{-1}$ and $\boldsymbol{A}^{-\frac{1}{2}} = (\boldsymbol{A}^{\frac{1}{2}})^{-1}$.

8.3 Singular value decomposition (SVD)

We are already able to decompose any symmetric matrix using the spectral theorem. Now we are interested in extending this decomposition to rectangular matrices: instead of eigenvectors and eigenvalues, non-square matrices have **singular vectors** corresponding to **singular values**.

Definition 8.3.1: Singular vectors and singular values

Let $\boldsymbol{A}$ be a $n \times p$ matrix, then σ is a **singular value** with corresponding left and right singular vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ respectively if

$$\boldsymbol{A}\boldsymbol{v} = \sigma\boldsymbol{u} \quad \text{and} \quad \boldsymbol{A}^T\boldsymbol{u} = \sigma\boldsymbol{v}.$$

Note that if $\boldsymbol{A}$ is a symmetric matrix then $\boldsymbol{u} = \boldsymbol{v}$ is an eigenvector corresponding to eigenvalue σ.

Proposition 8.3.1 Singular value decomposition

Let $\boldsymbol{A}$ be an $n \times p$ matrix with rank r, where $1 \leq r \leq \min(n, p)$. Then there exists an $n \times r$ matrix $\boldsymbol{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_r\}$, an $p \times r$ matrix $\boldsymbol{V} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_r\}$, and an $r \times r$ matrix $\boldsymbol{\Sigma} = \operatorname{diag}\{\sigma_1, \dots, \sigma_r\}$, such that

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

where $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_r = \boldsymbol{V}^T\boldsymbol{V}$ and $\sigma_1 \geq \dots \geq \sigma_r > 0$.

Note that all singular vectors are necessarily unit vectors and the singular values are ordered from largest to smallest. The form given above is called the **compact SVD**. The non-compact form is given by

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

where $\boldsymbol{U}$ is an $n \times n$ orthogonal matrix, $\boldsymbol{V}$ is a $p \times p$ orthogonal matrix and $\boldsymbol{\Sigma}$ is a $n \times p$ diagonal matrix, leaving the remaining diagonals after the r^{th} entry zero.

Proposition 8.3.2

Let $\boldsymbol{A}$ be a $n \times p$ matrix with rank r, then

$$\operatorname{rank}(\boldsymbol{A}^T\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}) = r.$$

Proof: We show that $\boldsymbol{A}$ and $\boldsymbol{A}^T\boldsymbol{A}$ have the same null space and hence the same nullity.

- Let $\boldsymbol{x} \in \mathbb{R}^p$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Then, $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T(\boldsymbol{0}_n) = \boldsymbol{0}_r$. So nullity$(\boldsymbol{A}) \subseteq$ nullity$(\boldsymbol{A}^T\boldsymbol{A})$.
- Let $\boldsymbol{x} \in \mathbb{R}^p$ such that $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Left multiplying both sides by $\boldsymbol{x}^T$,

$$\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|^2 = \boldsymbol{0}.$$

  This implies that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, so nullity$(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}) \subseteq$ nullity$(\boldsymbol{A}\boldsymbol{x})$.

Then since the dimension of the domain of both matrices are the same (p), hence by the rank-nullity formula, they have the same rank. □

Proposition 8.3.3

Let A be a $n \times p$ matrix of rank r. Then

- the non-zero eigenvalues of both $\boldsymbol{AA}^T$ and $\boldsymbol{A}^T\boldsymbol{A}$ are $\sigma_1^2, \ldots, \sigma_r^2$;
- the corresponding unit eigenvectors of $\boldsymbol{AA}^T$ are given by the columns of $\boldsymbol{U}$;
- the corresponding unit eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ are given by the columns of $\boldsymbol{V}$.

Proof: $\boldsymbol{A}^T\boldsymbol{A}$ is a $p \times p$ symmetric matrix, so by the spectral theorem we can write it as

$$\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$$

where $\boldsymbol{V}$ is a $p \times r$ **semi-orthogonal** matrix containing the orthonormal eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ (and hence its columns $\boldsymbol{v}_i$ are orthonormal) and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_r)$ is a diagonal matrix of eigenvalues with $\lambda_1 \geq \ldots \geq \lambda_r > 0$. Note that we know that there are r eigenvalues by Proposition 8.3.2.

Now, we let each $\sigma_i = \sqrt{\lambda_i}$ and $\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i$ for $i = 1, \ldots, r$. We will then show that the vectors $\boldsymbol{u}_i$ are orthonormal and subsequently, $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are left and right singular vectors corresponding to singular values σ_i.

$$\boldsymbol{u}_i^T\boldsymbol{u}_j = \frac{1}{\sigma_i\sigma_j}\boldsymbol{v}_i^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_j = \frac{1}{\sigma_i\sigma_j}\boldsymbol{v}_i^T(\lambda_j\boldsymbol{v}_j) = \frac{\sigma_j^2}{\sigma_i\sigma_j}\boldsymbol{v}_i^T\boldsymbol{v}_j = \frac{\sigma_j^2}{\sigma_i\sigma_j}\boldsymbol{v}_i \cdot \boldsymbol{v}_j$$

Notice that if $i = j$, then the expression simplifies to just $\|\boldsymbol{v}_i\|^2 = 1$ since the vectors $\boldsymbol{v}_i$ are orthonormal. If $i \neq j$, then the dot product collapses to zero. This behaviour implies that the vectors $\boldsymbol{u}_i$ are orthonormal. Now, following the definition of singular vectors, we consider

$$\boldsymbol{A}^T\boldsymbol{u}_i = \frac{1}{\sigma_i}(\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_i) = \frac{\sigma_i^2}{\sigma_i}\boldsymbol{v}_i = \sigma_i\boldsymbol{v}_i,$$

so $\boldsymbol{v}_i$ are right singular vectors corresponding to singular values σ_i while $\boldsymbol{u}_i$ by our formulation are left singular vectors. Now all that is left is to construct $\boldsymbol{U} = \begin{pmatrix} \boldsymbol{u}_1 & \ldots & \boldsymbol{u}_r & \ldots & \boldsymbol{u}_n \end{pmatrix}$ and $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1, \ldots, \sigma_r, 0, \ldots, 0\}$. □

Note:-

The procedure to compute SVD of A ($A = U\Sigma V^T$) are as follows:

1. Solve for eigenvalues λ of AA^T (or A^TA depending on which is easier) and take the singular values $\sigma = \sqrt{\lambda}$.
2. Solve for left singular vectors by using
$$(AA^T - \lambda I)\boldsymbol{u} = \boldsymbol{0}$$
3. Convert the left singular vectors to unit vectors and construct matrix $\boldsymbol{U}$.
4. Compute $\boldsymbol{V}$ by using the definition of right singular vectors:
$$\sigma_i\boldsymbol{v}_i = A^T\boldsymbol{u}_i.$$
5. Express $A = U\Sigma V^T$ as a sum of rank-one matrices.

8.4 Exercises

Question 1

Compute the singular value decomposition (full and compact) of the following matrices.

- $\begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$

Solution:

- Let $\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$. This matrix is diagonal, so $\boldsymbol{A}\boldsymbol{A}^T = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$ and the eigenvalues are 4 and 1. We take singular values $\sigma = 2, 1$, so $\boldsymbol{\Sigma} = \text{diag}\{2, 1\}$.
  Solving for left singular vectors (eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$) $\boldsymbol{u}$,

  $$(\boldsymbol{A}\boldsymbol{A}^T - 4I_2)\boldsymbol{u} = \boldsymbol{0} \qquad\qquad (\boldsymbol{A}\boldsymbol{A}^T - I_2)\boldsymbol{u} = \boldsymbol{0}$$

  $$\begin{pmatrix} 0 & 0 \\ 0 & -3 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad\qquad \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  Taking the unit eigenvector for each eigenvalue, we have

  $$\boldsymbol{u}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{u}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

  since zero row i indicate that any value of u_i will satisfy the system. Then we solve for $\boldsymbol{V}$,

  $$\boldsymbol{V} = A\boldsymbol{U}^T\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

  So the SVD of $\boldsymbol{A}$ is

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

- Let $\boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$. We take $\boldsymbol{B}^T\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ since the resulting product is 2×2. This is a diagonal matrix, so it follows that the only non-zero eigenvalue is $\lambda = 1$. Then we take the singular value $\sigma = \sqrt{1} = 1$. We now solve for the right singular vectors $\boldsymbol{v}$ which are eigenvectors of $\boldsymbol{B}^T\boldsymbol{B}$:

  $$(\boldsymbol{B}^T\boldsymbol{B} - \boldsymbol{I}_2)\boldsymbol{v} = \boldsymbol{0}$$

  $$\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

  We have $v_2 = 0$ and $v_1 \in \mathbb{R}$. The unit singular vector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We now solve for $\boldsymbol{U}$ in compact form:

  $$\boldsymbol{U} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Hence the SVD of $\boldsymbol{B}$ in compact form is

$$\boldsymbol{B} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix}.$$

For the non-compact form, we just need to choose vectors to fill in $\boldsymbol{U}$ and $\boldsymbol{V}$ so that we have orthonormal columns. An easy choice here is

$$\boldsymbol{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Hence the non-compact form is

$$\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Question 2

Let

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The eigen-decomposition of $\boldsymbol{X}^T\boldsymbol{X}$ is

$$\boldsymbol{X}^T\boldsymbol{X} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^T.$$

1. What are the singular values of $\boldsymbol{X}$?
2. What are the right singular vectors of $\boldsymbol{X}$?
3. What are the left singular vectors of $\boldsymbol{X}$?
4. Give the compact SVD of $\boldsymbol{X}$. Check your answer, noting that the singular vectors are only specified up to multiplication by -1.
5. Can you compute the full SVD of $\boldsymbol{X}$?
6. What is the eigen-decomposition of $\boldsymbol{X}\boldsymbol{X}^T$?
7. Find a generalised inverse of matrix $\boldsymbol{X}$.

Solution:

1. The eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$ are 3 and 1, so the singular values of $\boldsymbol{X}$ are $\sqrt{3}$ and 1.
2. The unit eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$ give the right singular vectors of $\boldsymbol{X}$:

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

3. We use $\boldsymbol{U} = \boldsymbol{X}\boldsymbol{V}\boldsymbol{\Sigma}^{-1}$ to compute the left singular vectors:

$$\boldsymbol{U} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{3}} & 0\\ 0 & 1\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\frac{2}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{3}} & 1\\ \frac{1}{\sqrt{3}} & -1\end{pmatrix} = \frac{1}{\sqrt{6}}\begin{pmatrix}2 & 0\\ 1 & \sqrt{3}\\ 1 & -\sqrt{3}\end{pmatrix}.$$

4. The compact SVD of $\boldsymbol{X}$ is

$$\boldsymbol{X} = \frac{1}{2\sqrt{3}}\begin{pmatrix}2 & 0\\ 1 & \sqrt{3}\\ 1 & -\sqrt{3}\end{pmatrix}\begin{pmatrix}\sqrt{3} & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}.$$

5. To compute the full SVD of $\boldsymbol{X}$, we need a third column in $\boldsymbol{U}$ which is orthonormal to the other columns. Let such column be $\boldsymbol{v}_3 = \begin{pmatrix}v_1 & v_2 & v_3\end{pmatrix}^T$, then we compute the null space of $\boldsymbol{U}$:

$$\begin{pmatrix}2 & 1 & 1\\ 0 & \sqrt{3} & -\sqrt{3}\end{pmatrix}\begin{pmatrix}v_1\\ v_2\\ v_3\end{pmatrix} = \begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

After performing row operations, we find that the null space is $\langle(-1, 1, 1)\rangle$. We choose one unit vector from this space $\frac{1}{\sqrt{3}}(-1, 1, 1)$ and rewrite the full SVD of $\boldsymbol{X}$:

$$\boldsymbol{X} = \frac{1}{2\sqrt{3}}\begin{pmatrix}2 & 0 & -\sqrt{2}\\ 1 & \sqrt{3} & \sqrt{2}\\ 1 & -\sqrt{3} & \sqrt{2}\end{pmatrix}\begin{pmatrix}\sqrt{3} & 0\\ 0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}$$

6. The eigenvalues of $\boldsymbol{X}\boldsymbol{X}^T$ should be equal to the eigenvalues of $\boldsymbol{X}^T\boldsymbol{X}$, 3 and 1, which are the squares of the singular values of $\boldsymbol{X}$. The corresponding eigenvectors is then equal to the left singular vectors of $\boldsymbol{X}$. Hence the eigen-decomposition of $\boldsymbol{X}\boldsymbol{X}^T$ is

$$\boldsymbol{X}\boldsymbol{X}^T = \begin{pmatrix}2 & 1 & 1\\ 1 & 1 & 0\\ 1 & 0 & 1\end{pmatrix} = \frac{1}{6}\begin{pmatrix}2 & 0\\ 1 & \sqrt{3}\\ 1 & -\sqrt{3}\end{pmatrix}\begin{pmatrix}3 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}2 & 1 & 1\\ 0 & \sqrt{3} & -\sqrt{3}\end{pmatrix}.$$

7. A generalised inverse $\boldsymbol{G}$ of $\boldsymbol{X}$ has the property

$$\boldsymbol{X}\boldsymbol{G}\boldsymbol{X} = \boldsymbol{X}.$$

Since $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ is invertible, a generalised inverse is $\boldsymbol{G} = \begin{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}^{-1} & \boldsymbol{O}\end{pmatrix} = \begin{pmatrix}1 & -1 & 0\\ 0 & 1 & 0\end{pmatrix}$ which we verify by

$$\boldsymbol{X}\boldsymbol{G}\boldsymbol{X} = \begin{pmatrix}1 & 1\\ 0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & -1 & 0\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 1 & -1 & 0\end{pmatrix}\begin{pmatrix}1 & 1\\ 0 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}1 & 1\\ 0 & 1\\ 1 & 0\end{pmatrix} = \boldsymbol{X}.$$

Question 3

The SVD can be used to solve linear systems of the form

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

where $\boldsymbol{A}$ is a $n \times p$ matrix, with compact SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$.

1. If $\boldsymbol{A}$ is a square invertible matrix, show that

$$\bar{\boldsymbol{x}} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T\boldsymbol{y}$$

is the unique solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$, i.e. show that $\boldsymbol{A}^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T$.

2. If $\boldsymbol{A}$ is not a square matrix, then $\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T$ is a pseudoinverse (not a true inverse) matrix, and $\bar{\boldsymbol{x}} = \boldsymbol{A}^+\boldsymbol{y}$ is still a useful quantity to consider as we shall now see. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$. Then $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ is an over-determined system in that there are 3 equations in 2 unknowns. Compute $\bar{\boldsymbol{x}} = \boldsymbol{A}^+\boldsymbol{y}$. Is this a solution to the equation?

Solution:

1. Since $\boldsymbol{A}$ is a square invertible matrix, from the SVD, we have

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A}^{-1} &= \boldsymbol{I} \\ \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{A}^{-1} &= \boldsymbol{I} \\ \boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{A}^{-1} &= \boldsymbol{U}^{-1} = \boldsymbol{U}^T \\ \boldsymbol{V}^T\boldsymbol{A}^{-1} &= \boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T \\ \boldsymbol{A}^{-1} &= (\boldsymbol{V}^T)^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T. \end{aligned}$$

2. From the previous question, we have computed the SVD of

$$\boldsymbol{X} = \frac{1}{2\sqrt{3}}\begin{pmatrix} 2 & 0 \\ 1 & \sqrt{3} \\ 1 & -\sqrt{3} \end{pmatrix}\begin{pmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

The pseudoinverse is hence

$$\boldsymbol{A}^+ = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \\ 0 & \sqrt{3} & -\sqrt{3} \end{pmatrix} = \frac{1}{2\sqrt{3}}\begin{pmatrix} \frac{2}{\sqrt{3}} & -\frac{2}{\sqrt{3}} & \frac{4}{\sqrt{3}} \\ \frac{2}{\sqrt{3}} & \frac{4}{\sqrt{3}} & -\frac{2}{\sqrt{3}} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & -1 & 2 \\ 1 & 2 & -1 \end{pmatrix}$$

We now verify that $\boldsymbol{A}^+\boldsymbol{y} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is a solution:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \boldsymbol{y}.$$

Chapter 9

Generalised inverse and pseudoinverse

9.1 Matrix generalised inverse

Definition 9.1.1: Generalised inverse

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, then $\boldsymbol{G} \in \mathbb{R}^{n \times m}$ is a **generalised inverse** of $\boldsymbol{A}$ iff

$$\boldsymbol{AGA} = \boldsymbol{A}.$$

If $\boldsymbol{A}$ is square and invertible, then it has a unique generalised inverse $\boldsymbol{A}^{-1}$. First we can see that $\boldsymbol{AA}^{-1}\boldsymbol{A} = \boldsymbol{AI} = \boldsymbol{A}$ and also that

$$\boldsymbol{G} = \boldsymbol{A}^{-1}(\boldsymbol{AGA})\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}(\boldsymbol{A})\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}.$$

To compute the generalised inverse of any matrix we use the following theorem:

Theorem 9.1.1

Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{pmatrix} \in \mathbb{R}^{m \times n}$ be a matrix of rank r, and $\boldsymbol{A}_{11} \in \mathbb{R}^{r \times r}$. If $\boldsymbol{A}_{11}$ is invertible, then $\boldsymbol{G} = \begin{pmatrix} \boldsymbol{A}_{11}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \in \mathbb{R}^{n \times m}$ is a generalised inverse of $\boldsymbol{A}$.

The generalised inverse can be used to find a solution to a consistent linear system:

Theorem 9.1.2

Consider the linear system $\boldsymbol{Ax} = \boldsymbol{b}$. Suppose $\boldsymbol{b} \in \mathrm{Col}(\boldsymbol{A})$ such that the system is consistent. Let $\boldsymbol{G}$ be a generalised inverse of $\boldsymbol{A}$, then $\boldsymbol{x}^* = \boldsymbol{Gb}$ is a particular solution to the system.

Proof: Multiplying both sides of $\boldsymbol{Ax} = \boldsymbol{b}$ by $\boldsymbol{AG}$ gives

$$\begin{aligned} (\boldsymbol{AG})\boldsymbol{Ax} &= (\boldsymbol{AG})\boldsymbol{b} \\ (\boldsymbol{AGA})\boldsymbol{x} &= \boldsymbol{A}(\boldsymbol{Gb}) \\ \boldsymbol{Ax} &= \boldsymbol{Ax}^* \end{aligned}$$

□

Example 9.1.1

Consider the linear system $\boldsymbol{Ax} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} 6 \\ 15 \\ 24 \end{pmatrix}.$$

$\boldsymbol{A}$ has a rank of 2, since after eliminating the first entries in row 2 and row 3 using row 1, we are left with linearly dependent rows in row 2 and 3. Hence we now turn to generalised inverses. By Theorem 8.1.1, since $\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$ is invertible, a generalised matrix of $\boldsymbol{A}$ is

$$\boldsymbol{G} = \frac{1}{3}\begin{pmatrix} -5 & 2 & 0 \\ 4 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then, a particular solution to the system is

$$\boldsymbol{x}^* = \boldsymbol{Gb} = \frac{1}{3}\begin{pmatrix} -5 & 2 & 0 \\ 4 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 6 \\ 15 \\ 24 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix},$$

valid since $\boldsymbol{b}$ is in the column span of $\boldsymbol{A}$, computed as follows:

$$k_1\begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} + k_2\begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} + k_3\begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \boldsymbol{a},$$

for any $\boldsymbol{a}$ in the column space of $\boldsymbol{A}$. Then we try to solve for k_1, k_2, k_3 by reducing

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} \xrightarrow[A_{13}(-3)]{A_{12}(-2)} \begin{pmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix} \xrightarrow[M_2(-\frac{1}{3}), A_{21}(-4)]{A_{23}(-2)} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence, the column space of $\boldsymbol{A}$ is

$$\mathrm{Col}(\boldsymbol{A}) = \langle (1, 0, -1)^T, (0, 1, 2)^T \rangle$$

and we can write

$$\boldsymbol{b} = 6(1, 0, -1)^T + 15(0, 1, 2)^T.$$

9.2 Projection matrices

Definition 9.2.1: Projection matrix

A square matrix $\boldsymbol{P}$ is a projection matrix if $\boldsymbol{P} = \boldsymbol{P}^2$.

Example 9.2.1 (Projection matrices)

$$I, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, O$$

Notice that

$$\det(P) = \det(P^2) = \det(P)^2 \Rightarrow \det(P) = 0, 1.$$

We will give a geometric interpretation of this definition here: Matrices project/map any vector in $\mathbb{R}^n$ onto its range (column space). This is obvious as applying a matrix to a vector gives a linear combination of the columns of the matrix. In addition to this, a projection matrix also keeps all points from its range in their original places.

Theorem 9.2.1

Let $A \in \mathbb{R}^{m\times n}$ with a generalised inverse $G \in \mathbb{R}^{n\times m}$. Then $AG \in \mathbb{R}^{m\times m}$ is a projection matrix.

Proof: From $AGA = A$, we obtain

$$(AG)(AG) = (AGA)G = AG.$$

□

Proposition 9.2.1

Let $A \in \mathbb{R}^{m\times n}$, $G \in \mathbb{R}^{n\times m}$ is a generalised inverse of A, then

$$\text{Col}(AG) = \text{Col}(A)$$

Proof: We prove this in two cases:

- Suppose that $y \in \text{Col}(AG)$, then there exists some $x \in \mathbb{R}^m$ such that $y = (AG)x = A(Gx) \in \text{Col}(A)$. Hence $\text{Col}(AG) \subseteq \text{Col}(A)$.
- Suppose that $y \in \text{Col}(A)$, then there exists $x \in \mathbb{R}^n$ such that $y = Ax = AGAx = (AG)(Ax) \in \text{Col}(AG)$. Hence $\text{Col}(A) \subseteq \text{Col}(AG)$.

Combining both cases, $\text{Col}(AG) = \text{Col}(A)$. □

Since $(AG)^2 = (AGA)G = AG$, AG is a projection matrix, and from the previous proposition, it projects onto the column space of A. Similarly, GA is a projection matrix onto the row space of A.

9.3 Pseudoinverse

Definition 9.3.1: Pseudoinverse

Let $A \in \mathbb{R}^{m\times n}$. $A^+ \in \mathbb{R}^{n\times m}$ is the **pseudoinverse** (or the **Moore-Penrose inverse**) of A iif

1. A^+ is a generalised inverse of A: $AA^+A = A$;
2. A is a generalised inverse of A^+: $A^+AB = A^+$;
3. AA^+ is symmetric: $(AA^+)^T = AA^+$;
4. A^+A is symmetric: $(A^+A)^T = A^+A$.

If $\boldsymbol{A}^+$ only satisfies the first two conditions, it is a **reflexive generalised inverse**. Furthermore, it can be shown that for any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the pseudoinverse always exists and is unique.

9.4 Orthogonal projection matrices

Definition 9.4.1: Orthogonal projection matrix

A square matrix $\boldsymbol{P}$ is an orthogonal projection matrix iff $\boldsymbol{P} = \boldsymbol{P}^2$ and $\boldsymbol{P} = \boldsymbol{P}^T$.

Theorem 9.4.1

For any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and its pseudoinverse $\boldsymbol{A}^+$, $\boldsymbol{A}\boldsymbol{A}^+$ is an orthogonal projection matrix (onto the column space of $\boldsymbol{A}$).

Proof: Since $\boldsymbol{A}^+$ is a generalised inverse, it follows that $\boldsymbol{A}\boldsymbol{A}^+$ is a projection matrix onto the columns of $\boldsymbol{A}$. Furthermore, by definition of pseudoinverses, $\boldsymbol{A}\boldsymbol{A}^+$ is symmetric. □

Similarly, $\boldsymbol{A}^+\boldsymbol{A}$ is an orthogonal projection matrix onto the row space of $\boldsymbol{A}$.

Theorem 9.4.2

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be any tall matrix with full column rank, i.e. $\text{rank}(\boldsymbol{A}) = n \le m$, then the pseudoinverse of $\boldsymbol{A}$ is

$$\boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$$

If we have that the SVD of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, then

$$\begin{aligned}
\boldsymbol{A}^+ &= ((\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^{-1}(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T \\
&= (\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^{-1}(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T \\
&= (\boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^T)^{-1}(\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T) \\
&= (\boldsymbol{V}^T)^{-1}(\boldsymbol{\Sigma}^{-1})^2(\boldsymbol{V})^{-1}\boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T \\
&= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^T
\end{aligned}$$

9.5 Application to solving linear systems of equations

Theorem 9.5.1

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. If the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has solutions, then $\boldsymbol{x}^* = \boldsymbol{A}^+\boldsymbol{b}$ is an exact solution and has the smallest possible norm, i.e. $\|\boldsymbol{x}^*\| \le \|\boldsymbol{x}\|\ \forall \boldsymbol{x}$.

Proof: First, since $\boldsymbol{A}^+$ is a generalised inverse, $\boldsymbol{A}^+\boldsymbol{b}$ must be a solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Now, for any solution $\boldsymbol{x} \in \mathbb{R}^n$, consider its orthogonal decomposition via $\boldsymbol{A}^+\boldsymbol{A} \in \mathbb{R}^{n\times n}$:

$$\boldsymbol{x} = (\boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x} + (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x} = \boldsymbol{A}^+\boldsymbol{b} + (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x}.$$

Then by the Pythagorean theorem, we have

$$\|\boldsymbol{x}\|^2 = \|\boldsymbol{A}^+\boldsymbol{b}\|^2 + \|(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x}\|^2 \ge \|\boldsymbol{A}^+\boldsymbol{b}\|^2.$$

Hence $\|\boldsymbol{x}\| \ge \|\boldsymbol{A}^+\boldsymbol{b}\|$. □