1. Inner Products and Norms

One knows from a basic introduction to vectors in $\mathbb{R}^n$ (Math 254 at OSU) that the length of a vector $\mathbf{x} = (x_1 \ x_2 \ ... \ x_n)^T \in \mathbb{R}^n$, denoted $\|\mathbf{x}\|$, or $|\mathbf{x}|$, is given by the formula

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$ 

We will call this function $\| - \| : \mathbb{R}^n \to [0, \infty)$ the (standard) norm on $\mathbb{R}^n$.

Recall the familiar dot product on $\mathbb{R}^n$:

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then $\mathbf{x} = (x_1 \ x_2 \ ... \ x_n)^T, \mathbf{y} = (y_1 \ y_2 \ ... \ y_n)^T$ where $x_i, y_i \in \mathbb{R}$ for $i = 1, 2, ..., n$. Then

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$ 

We will call the dot product on $\mathbb{R}^n$ the (standard) inner product on $\mathbb{R}^n$, and we will use the fancier notation $(\mathbf{x}, \mathbf{y})$ for $\mathbf{x} \cdot \mathbf{y}$. Formally, we can regard the inner product as a function $(-, -) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

An easy and important consequence is that $(\mathbf{x}, \mathbf{x}) = \|\mathbf{x}\|^2$, equivalently, $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$.

We also have another expression for this inner product:

For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T \mathbf{x}$ (same as $\mathbf{x}^T \mathbf{y}$).
Now let’s explore the (standard) norm and the (standard) inner product on \( \mathbb{C}^n \):

Recall that for \( z \in \mathbb{C}, z = a + bi \), and \( |z|^2 = a^2 + b^2 = z \bar{z} \) (really just the \( \mathbb{R}^2 \) (standard) norm!)

Now let \( z = (z_1 \ z_2 \ldots \ z_n)^T \) where \( z_j = a_j + b_j i \) for \( j = 1, 2, \ldots, n \), that is \( z \in \mathbb{C}^n \).

It is natural to define the (standard) norm on \( \mathbb{C}^n \) by

\[
\|z\| = \sqrt{\sum_{j=1}^{n} |z_j|^2} = \sqrt{\sum_{j=1}^{n} (a_j^2 + b_j^2)} \text{ for } z \in \mathbb{C}^n.
\]

Furthermore, with \( w = (w_1 \ w_2 \ldots \ w_n)^T \) in \( \mathbb{C}^n \), we define the (standard) inner product on \( \mathbb{C}^n \) by

\[
(z, w) = \sum_{i=1}^{n} z_i \bar{w}_i.
\]

Again, an easy and important consequence is that \((z, z) = \|z\|^2\), equivalently, \(\|z\| = \sqrt{(z, z)}\).

Let \( A \in M_{m \times n} \) have entries in \( \mathbb{C} \). Let \( A^* = (\bar{A})^T \) where \( \bar{A} \) determined by applying the conjugate operation entry-wise. This combination (conjugate transpose) is the hermitian adjoint or just the adjoint of \( A \).

For \( w, w^* = (\bar{w})^T = (\bar{w}_1 \ \bar{w}_2 \ldots \ \bar{w}_n) \) and thus

\[
(z, w) = w^* z.
\]
Now let’s get more abstract. We will let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$. Let $V$ be an arbitrary vector space over $\mathbb{F}$. An inner product on $V$ is a function

$$( -, - ) : V \times V \to \mathbb{F},$$

that takes in two vectors $u, v \in V$ and outputs a scalar $(u, v) \in \mathbb{F}$, satisfying these four fundamental properties:

(1) Positivity (aka non-negativity): For all $v \in V$, $(v, v) \in \mathbb{R}$ and $(v, v) \geq 0$.

(2) Definiteness (aka non-degeneracy): Let $v \in V$. Then $(v, v) = 0$ iff $v = 0$.

(3) Linearity: For all $c \in \mathbb{F}$, for all $u, v, w \in V$, $(u + cv, w) = (u, w) + c(v, w)$.

(4) Conjugate symmetry: For all $u, v \in V$, $(u, v) = (v, u)$ (just symmetry if $\mathbb{F} = \mathbb{R}$.)

Let $(-, -)$ be an inner product on a vector space $V$. We show $\forall x \in V$, $(0, x) = (x, 0) = 0$.

Let $x \in V$. By conjugate symmetry, it suffices to establish that $(0, x) = 0$. Well, $(0, x) = (0 + 0, x) = (0, x) + (0, x)$ by linearity. Therefore $(0, x) = 0$.

If $V$ is a vector space over $\mathbb{F}$ and $(-, -)$ is an inner product on $V$ then we call $V$ and inner product space over $\mathbb{F}$.

Examples:

(In all of these, if $\mathbb{F} = \mathbb{R}$ then there is no need for the conjugation.)

(1) Let $V = \mathbb{F}^n$. We have already covered standard (aka Euclidean) inner product on $V$ given by

$$\text{For all } u, v \in V, \ (u, v) = v^* u.$$ 

(ELFY) Show that this function does indeed satisfy the four fundamental properties of an inner product...

(2) Let $V = \mathbb{P}_n$, polynomials of degree at most $n$ over $\mathbb{F}$. Let the interval $I = [a, b]$. Then for $f(x), g(x) \in \mathbb{P}_n$ define

$$(f(x), g(x))_I = \int_a^b f(x) \overline{g(x)} dx.$$ 

(ELFY) Show that this function does indeed satisfy the four fundamental properties of an inner product...
(3) Let $V = M_{m,n}$ (over $\mathbb{F}$). For $A, B \in M_{mn}$ define $(A, B) = \text{trace}(B^*A)$.

(ELFY) Show that this function does indeed satisfy the four fundamental properties of an inner product...

Note:

$$\text{trace}(B^*A) = \sum_{i=1}^{n} (B^*A)_{ii} = \sum_{i=1}^{n} (\overline{b_1}a_{1i} + \overline{b_2}a_{2i} + \cdots + \overline{b_m}a_{mi}).$$

Properties of inner products:

From here until otherwise indicated (page 12), let $V$ be an inner product space (over $\mathbb{F}$).

(ELFY) Show that For all $c \in \mathbb{F}$, for all $u, v, w \in V$, $(u, v + cw) = (u, v) + \overline{c}(u, w)$.

Hint: Use conjugate symmetry, linearity, and then conjugate symmetry again!

**Proposition:** Let $x \in V$. Then,

$x = 0$ iff for all $y \in V$, $(x, y) = 0$.

**Proof:** The forward direction is trivial:

For all $y \in V$, $(0, y) = (0y, y) = 0(y, y) = 0$ (and $(y, 0) = (y, 0y) = 0(y, y) = 0$ as well).

Now assume $x \in V$ satisfies for all $y \in V$, $(x, y) = 0$. Then in particular, $(x, x) = 0$. Using the definite property of an inner product we get $x = 0$ □

(ELFY) Use the proposition above to prove that for all $x, y \in V$, $x = y$ iff for all $z \in V$, $(x, z) = (y, z)$. 
**Proposition:** Let $T, S \in \mathcal{L}(V)$. Suppose that for all $x, y \in V$, $(Tx, y) = (Sx, y)$. Then $T = S$.

**Proof:** Let $T, S \in \mathcal{L}(V)$.

(*) Assume for any $x, y \in V$ that $(Tx, y) = (Sx, y)$.

Assume $T \neq S$. Then there exists $x \in V$ such that $Tx \neq Sx$. However, $(Tx, y) = (Sx, y)$ for all $y \in V$ by assumption (*). Using the exercise left for you on the previous page, $Tx = Sx$, which is a contradiction. So $T = S$ □

Note: The argument doesn’t have to be done by contradiction.

An inner product on $V$ induces a norm on $V$ corresponding to the inner product, which is a function $\| - \| : V \to \mathbb{R}$ given by

$$\|v\| = \sqrt{(v, v)} \text{ for } v \in V.$$ 

Basic norm property:

$$\|cv\| = \sqrt{(cv, cv)} = \sqrt{|c|^2(v, v)} = |c|\|v\|.$$

Also, we call two vectors $u, v \in V$ orthogonal if $(u, v) = 0$ (as a consequence, by conjugate symmetry, $(v, u) = 0$ would also hold).

(ELFY) if $u, v$ are orthogonal, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$ (Pythagorean Theorem).
Orthogonal decomposition:

Let $u, v \in V$. Assume $v \neq 0$. Then $(v, v) \neq 0$. Let $c \in \mathbb{F}$. Clearly,

$$u = cv + (u - cv).$$

Let's find $c$ such that $v, (u - cv)$ are orthogonal:

$$(u - cv, v) = 0 \rightarrow (u, v) - c(v, v) = 0 \rightarrow (u, v) = c(v, v) \rightarrow c = \frac{(u, v)}{(v, v)} = \frac{(u, v)}{\|v\|^2}.$$ 

We will call $\text{proj}_v(u) = \frac{(u, v)}{\|v\|^2}v$ the orthogonal projection of $u$ onto $v$.

We will call $u = \text{proj}_v(u) + (u - \text{proj}_v(u))$ the orthogonal decomposition of $u$ along $v$.

Consider that

$$(*) \quad \|\text{proj}_v(u)\| = \left(\frac{|(u, v)|}{\|v\|^2}\right)\|v\| = \frac{|(u, v)|}{\|v\|}.$$ 

Notice that since $(u - \text{proj}_v(u), v) = 0$ it follows that

$$(u - \text{proj}_v(u), \text{proj}_v(u)) = (u - \text{proj}_v(u), \frac{(u, v)}{\|v\|^2}v) = \frac{(u, v)}{\|v\|^2}(u - \text{proj}_v(u), v) = 0.$$ 

Thus if $u = \text{proj}_v(u) + (u - \text{proj}_v(u))$ is the orthogonal decomposition of $u$ along $v$, then the vector terms satisfy the Pythagorean Theorem:

$$(**) \quad \|u\|^2 = \|\text{proj}_v(u)\|^2 + \|u - \text{proj}_v(u)\|^2.$$ 

Then $(*)$ and $(**)$ implies

$$\|u - \text{proj}_v(u)\|^2 = \|u\|^2 - \frac{|(u, v)|^2}{\|v\|^2}.$$
You could of course calculate this directly:

\[
\|u - \text{proj}_v(u)\|^2 = (u - \text{proj}_v(u), u - \text{proj}_v(u))
\]

\[
= (u, u - \text{proj}_v(u)) - (\text{proj}_v(u), u - \text{proj}_v(u))
\]

\[
= (u, u) - (u, \text{proj}_v(u)) - (\text{proj}_v(u), u) + (\text{proj}_v(u), \text{proj}_v(u))
\]

\[
= \|u\|^2 + \frac{|(u, v)|^2}{\|v\|^2} - (u, \text{proj}_v(u)) - \frac{(u, v)}{\|v\|^2} \frac{v}{\|v\|^2}
\]

\[
= \|u\|^2 + \frac{|(u, v)|^2}{\|v\|^2} - \frac{(u, v)}{\|v\|^2} \frac{v}{\|v\|^2}
\]

\[
= \|u\|^2 + \frac{|(u, v)|^2}{\|v\|^2} - \frac{|(u, v)|^2}{\|v\|^2} - \frac{|(u, v)|^2}{\|v\|^2}
\]

\[
= \|u\|^2 - \frac{|(u, v)|^2}{\|v\|^2}.
\]

In the last step we used the fact that conjugation does nothing to a real number:

\[
\frac{|(u, v)|^2}{\|v\|^2} = \frac{|(u, v)|^2}{\|v\|^2}.
\]
The following propositions is of great importance.

**Proposition (Cauchy-Schwarz Inequality):** If \( u, v \in V \) then

\[
|(u, v)| \leq \|u\| \|v\|.
\]

**Proof:** Let \( u, v \in V \). If \( v = 0 \) then the inequality holds as both sides are clearly 0.

So we assume that \( v \neq 0 \). Consider the orthogonal decomposition \( u = \text{proj}_v(u) + w \) where \( w = (u - \text{proj}_v(u)) \).

\[
0 \leq \|w\|^2 = \|u\|^2 - \frac{|(u, v)|^2}{\|v\|^2} \rightarrow \\
\frac{|(u, v)|^2}{\|v\|^2} \leq \|u\|^2 \rightarrow \\
\frac{|(u, v)|}{\|v\|} \leq \|u\| \rightarrow \\
|(u, v)| \leq \|u\| \|v\| \square
Proposition (Triangle Inequality): If \( u, v \in V \) then
\[
\|u + v\| \leq \|u\| + \|v\|.
\]

Proof: Let \( u, v \in V \). Consider that for all \( z = a + bi \in \mathbb{C}, \)
\[
\text{Re}(a + bi) = a \leq |a| = \sqrt{a^2} \leq \sqrt{a^2 + b^2} = |a + bi|.
\]

Then
\[
\|u + v\|^2 = (u + v, u + v)
\]
\[
= (u, u) + (v, v) + (u, v) + (v, u)
\]
\[
= \|u\|^2 + \|v\|^2 + (u, v) + (v, u)
\]
\[
= \|u\|^2 + \|v\|^2 + 2\text{Re}(u, v)
\]
\[
\leq \|u\|^2 + \|v\|^2 + 2|u, v|
\]
\[
\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\|
\]
\[
= (\|u\| + \|v\|)^2.
\]

Hence, \( \|u + v\| \leq \|u\| + \|v\| \) □
These are the so-called polarization identities:

**Proposition:** For all $x, y \in V$,

1. $(x, y) = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 \right)$ if $F = \mathbb{R}$.

2. $(x, y) = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2 \right)$ if $F = \mathbb{C}$.

**Proof:** Let $x, y \in V$. Assume $F = \mathbb{R}$. Then conjugation does nothing and $(x, y) = (y, x)$.

Consider that

$$
\frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 \right) = \frac{1}{4} ((x + y, x + y) - (x - y, x - y)) = \frac{1}{4} ((x, x) + 2(x, y) + (y, y) - [(x, x) - 2(x, y) + (y, y)]) = (x, y).
$$

Assume $F = \mathbb{C}$. Then

$$
z = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2 \right) = \frac{1}{4} ((x + y, x + y) - (x - y, x - y) + i(x + iy, x + iy) - i(x - iy, x - iy))
$$

Let’s split $z$ into two parts:

$$
c_1 = \frac{1}{4} ((x + y, x + y) - (x - y, x - y)),
$$

and

$$
c_2 = \frac{1}{4} (i(x + iy, x + iy) - i(x - iy, x - iy)).
$$

Clearly $z = c_1 + c_2$. 
\[ c_1 = \frac{1}{4} ((x, x) + (x, y) + (y, x) + (y, y) - [(x, x) - (x, y) - (y, x) + (y, y)]) \rightarrow \]
\[ c_1 = \frac{1}{4} (2(x, y) + 2(y, x)) = \frac{1}{2} ((x, y) + (y, x)). \]

\[ c_2 = \frac{i}{4} ((x, x) + (x, iy) + (iy, x) + (iy, iy) - [(x, x) - (x, iy) - (iy, x) + (iy, iy)]) \rightarrow \]
\[ c_2 = \frac{i}{4} (2(x, iy) + 2(iy, x)) = \frac{i}{2} [(-i)(x, y) + i(y, x)] = \frac{1}{2} ((x, y) - (y, x)). \]

Finally, \( z = c_1 + c_2 = (x, y) \). \( \square \)

**Proposition (Parallelogram Identity):** For all \( x, y \in V \),

\[ \|x + y\|^2 + \|x - y\|^2 = 2 (\|x\|^2 + \|y\|^2). \]

I leave the proof of this one to you...it’s pretty easy.
Now let $V$ be a vector space over $\mathbb{F}$, but we no longer assume that there is an inner product on $V$.

The function $n : V \to \mathbb{R}$ is a **norm on $V$** if it satisfies the following properties:

1. **Homogeneity:** For all $v \in V$ and for all $c \in \mathbb{F}$, $n(cv) = |c|n(v)$.
2. **Triangle Inequality:** For all $u, v \in V$, $n(u + v) \leq n(u) + n(v)$.
3. **Positivity (aka non-negativity):** For all $v \in V$, $n(v) \geq 0$.
4. **Definiteness (aka non-degeneracy):** Let $v \in V$. Then $n(v) = 0$ iff $v = \mathbf{0}$.

Of course, the norm $\| - \| : V \to \mathbb{R}$ induced from an inner product space $V$ satisfies these properties:

We have already verified that (1) and (2) hold. Properties (3) and (4) are direct consequences of the positivity and definiteness of the inner product.

So all inner product spaces are normed spaces, via the induced norm $\| v \| = \sqrt{(v, v)}$.

However, there are normed spaces where the norm does not correspond to an inner product.

When $V$ be a normed vector space, with norm $n$, we will still use the notation $n(v) = \| v \|$.
Let $V = \mathbb{R}^n$ or $V = \mathbb{C}^n$. It turns out that for any $p$, $1 \leq p < \infty$, following function $\| - \|_p : V \to \mathbb{R}$ given by

$$\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}},$$

is a norm called the $p$-norm. Of course the 2-norm (aka Euclidean norm) is the norm induced by the standard inner product on $V$.

We can also define a norm called the $\infty$-norm:

$$\|x\|_{\infty} = \max\{|x_i| : i = 1, 2, ..., n\}.$$ 

For $\| - \|_p$ and $\| - \|_{\infty}$ it is easy to check (ELFY) that they satisfy defining properties (1), (3), and (4) of a norm.

The triangle inequality is a different story. Of course we already have proven the triangle inequality for $\| - \|_2$.

The triangle inequality for $\| - \|_p$ isn't too hard to check (ELFY) when $p = 1$ and $p = \infty$.

For arbitrary values of $p$ it is fairly involved to prove the triangle inequality, but it does hold, and is called Minkowski’s Inequality. We won’t do that here...

It turns out that $\| - \|_p$ for $p \neq 2$ cannot be induced by an inner product (they don’t satisfy the parallelogram inequality)!

The proof is beyond the scope of this course (because of the reverse implication), but the following proposition characterizes norms induced from inner products:

**Proposition:** Let $V$ be a normed space with a norm $\| - \|$. The norm $\| - \|$ is induced from an inner product $(-, -)$ on $V$ iff $\| - \|$ satisfies the parallelogram inequality, that is, for all $x, y \in V$,

$$\|x + y\|^2 + \|x - y\|^2 = 2 (\|x\|^2 + \|y\|^2).$$

**Suggested Homework:** 1.1, 1.2, 1.3, 1.5, 1.7, and 1.9.
2. Orthogonal and Orthonormal Bases

Recall two vectors \( u, v \) in an inner product space \( V \) are orthogonal if \( (u, v) = 0 \). We will use the notation \( u \perp v \).

A vector \( v \in V \) is orthogonal to a subspace \( U \) of \( V \) if for all \( u \in U \), \( u \perp v \). We use the notation \( v \perp U \). Two subspaces \( U_1, U_2 \) are orthogonal subspaces if for all \( u_1 \in U_1 \), \( u_1 \) is orthogonal to the subspace \( U_2 \).

Furthermore, we can define for any subspace \( U \) of \( V \), the orthogonal complement to \( U \) (in \( V \)): \( U^\perp = \{ v \in V \mid v \perp U \} \).

Also \( u \) is called a unit vector if \( (u, u) = 1 \) (and thus \( \|u\| = 1 \)).

The vectors \( v_1, v_2, ..., v_k \) are called an orthogonal system if they are pairwise orthogonal, and an orthonormal system if they are an orthogonal system of unit vectors.

**Proposition:** Any orthogonal system of non-zero vectors is linearly independent.

**Proof:** Let \( v_1, v_2, ..., v_k \) be an orthogonal system of non-zero vectors. Suppose

\[
\sum_{i=1}^{k} c_i v_i = 0 \text{ for } c_i \in \mathbb{F}, \ i = 1, 2, ..., k.
\]

Then for any \( j = 1, 2, ..., k \) we get

\[
\left( \sum_{i=1}^{k} c_i v_i, v_j \right) = \sum_{i=1}^{k} c_i (v_i, v_j) = c_j \|v_j\|^2 = 0.
\]

For all \( j = 1, 2, ..., k \), \( \|v_j\|^2 \neq 0 \) as \( v_j \neq 0 \). Therefore, \( c_1 = c_2 = ... = c_k = 0 \), which completes the proof that the orthogonal system is linearly independent \( \square \).
Proposition (General Pythagorean Theorem): Let \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \) be an orthogonal system. Then for all \( c_1, c_2, \ldots, c_k \in F \),

\[
\left\| \sum_{i=1}^{k} c_i \mathbf{v}_i \right\|_2^2 = \sum_{i=1}^{k} |c_i|^2 \|\mathbf{v}_i\|_2^2.
\]

Proof: Let \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \) be an orthogonal. Let \( c_i \in F \) for \( i = 1, 2, \ldots, k \).

\[
\left\| \sum_{i=1}^{k} c_i \mathbf{v}_i \right\|_2^2 = \left( \sum_{i=1}^{k} c_i \mathbf{v}_i, \sum_{j=1}^{k} c_j \mathbf{v}_j \right) = \sum_{i=1}^{k} \sum_{j=1}^{k} c_i \overline{c}_j (\mathbf{v}_i, \mathbf{v}_j) = \sum_{i=1}^{k} c_i \overline{c}_i (\mathbf{v}_i, \mathbf{v}_i) = \sum_{i=1}^{k} |c_i|^2 \|\mathbf{v}_i\|_2^2 \quad \square
\]

A basis \( \mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \) of \( V \) is an orthogonal basis (orthonormal basis) if \( \mathcal{B} \) is an orthogonal system (orthonormal system).

Due to a previous proposition, if \( n = \dim V \), then any orthogonal (orthonormal) system of \( n \) non-zero vectors forms a orthogonal (orthonormal) basis.

Let \( \mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \) be an orthogonal basis of \( V \). Suppose \( \mathbf{x} \in V \). Usually finding \( [\mathbf{x}]_{\mathcal{B}} = (c_1 \ c_2 \ \cdots \ c_n)^T \) involves solving a linear system, but it is easier when \( \mathcal{B} \) is an orthogonal basis. For any \( j = 1, 2, \ldots, n \),

\[
(\mathbf{x}, \mathbf{v}_j) = \left( \sum_{i=1}^{n} c_i \mathbf{v}_i, \mathbf{v}_j \right) = \sum_{i=1}^{n} c_i (\mathbf{v}_i, \mathbf{v}_j) = c_j \|\mathbf{v}_j\|_2^2 \rightarrow c_j = \frac{(\mathbf{x}, \mathbf{v}_j)}{\|\mathbf{v}_j\|_2^2}.
\]

(This should look familiar! Recall that \( \text{proj}_\mathbf{v}(\mathbf{u}) = \frac{(\mathbf{u}, \mathbf{v})}{\|\mathbf{v}\|_2} \mathbf{v} \).)

Furthermore, if \( \mathcal{B} \) is an orthonormal basis, this result reduces to \( c_j = (\mathbf{x}, \mathbf{v}_j) \).
When $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ is an orthonormal basis, then for any $v \in V$, the formula

$$v = \sum_{i=1}^{n} (v, v_i)v_i,$$

is called an orthogonal Fourier decomposition.

**Suggested Homework:** 2.2, 2.3, 2.5, 2.6, 2.8, and 2.9.
3. Orthogonalization and Gram-Schmidt

Let \( W \) be a subspace of \( V \). For each \( v \in V \), \( w \in W \) is called an orthogonal projection of \( v \) onto \( W \) if \( (v - w) \perp W \).

**Proposition:** Let \( W \) be a subspace of \( V \) and let \( w \) be an orthogonal projection of \( v \) onto \( W \). Then

1. For all \( x \in W \), \( \|v - w\| \leq \|v - x\| \) (\( w \) minimizes distance from \( v \) to \( W \)).
2. If \( x \in W \) and \( \|v - w\| = \|v - x\| \) then \( x = w \) (orthogonal projections are unique).

**Proof:** Let \( W \) be a subspace of \( V \) and let \( w \in W \) be an orthogonal projection of \( v \) onto \( W \). Let \( x \in W \). Set \( y = w - x \in W \). Then

\[
v - x = (v - w) + y.
\]

Since \((v - w) \perp y\) as \( y \in W \), by the Pythagorean Theorem we get

\[
\|v - x\|^2 = \|v - w\|^2 + \|y\|^2 \implies \|v - x\| \geq \|v - w\|,
\]
and \( \|v - x\| = \|v - w\| \) iff \( y = 0 \). Hence \( \|v - x\| = \|v - w\| \) iff \( x = w \). □

(Now we can say the orthogonal projection of \( v \) onto \( W \).)

Let \( W \) be a subspace of \( V \) with an orthogonal basis \( \mathcal{B} = \{v_1, v_2, ..., v_k\} \). Let \( v \in V \) and \( P_W v \) be the orthogonal projection of \( v \) onto \( W \).

Let \( w = \sum_{i=1}^{k} \frac{(v, v_i)}{\|v_i\|^2} v_i \in W \).

Then \( v - w = v - \sum_{i=1}^{k} \frac{(v, v_i)}{\|v_i\|^2} v_i \).
Let \( x \in W \). Then there exists scalars \( c_1, c_2, \ldots, c_k \) such that \( x = \sum_{i=1}^{k} c_i v_i \).

Then

\[
(v - w, x) = \left( v - \sum_{i=1}^{k} \frac{(v, v_i)}{\|v_i\|^2} v_i \sum_{j=1}^{k} c_j v_j \right)
\]

\[
= \left( v, \sum_{i=1}^{k} c_i v_i \right) - \left( \sum_{i=1}^{k} \frac{(v, v_i)}{\|v_i\|^2} v_i ; \sum_{j=1}^{k} c_j v_j \right)
\]

\[
= \sum_{i=1}^{k} \overline{c_i}(v, v_i) - \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{(v, v_i)}{\|v_i\|^2} \overline{c_j} (v_i, v_j)
\]

\[
= \sum_{i=1}^{k} \overline{c_i}(v, v_i) - \sum_{i=1}^{k} \left( \frac{(v, v_i)}{\|v_i\|^2} \overline{c_i} (v_i, v_i) \right)
\]

\[
= \sum_{i=1}^{k} \overline{c_i}(v, v_i) - \sum_{i=1}^{k} \|v_i\|^2 \overline{c_i} (v_i, v_i)
\]

\[
= \sum_{i=1}^{k} \overline{c_i}(v, v_i) - \sum_{i=1}^{k} \overline{c_i}(v, v_i)
\]

\[
= 0.
\]

This shows that \( w \in W \) satisfies \( (v - w) \perp W \), and so, \( w \) is the orthogonal projection of \( v \) onto \( W \). Hence we get the formula:

\[
P_W v = \sum_{i=1}^{k} \frac{(v, v_i)}{\|v_i\|^2} v_i.
\]

**Claim:** Let \( W \) be a subspace of \( V \). \( P_W \in \mathcal{L}(V) \).

Let \( u, v \in V \) and \( c \in F \). Let \( w \in W \). Then \( P_W u + cP_W v \in W \) and

\[
([u + cv] - [P_W u + cP_W v], w) = ([u - P_W u] + c[v - P_W v], w) = (u - P_W u, w) - c(v - P_W v, w) = 0.
\]

Thus \( P_W (u + cv) = P_W u + cP_W v \) showing that \( P_W \) is a linear operator on \( V \).
The Gram-Schmidt algorithm:

Input: A linearly independent set of vectors \( \{x_1, x_2, ..., x_k\} \) spanning a subspace \( W \) of \( V \).

Output: An orthogonal system \( \{v_1, v_2, ..., v_k\} \) such that the span is \( W \).

Here are the steps:

1. Set \( v_1 = c x_1 \) for any non-zero \( c \in \mathbb{F} \) (usually 1).
   Set \( W_1 = \text{span}\{x_1\} = \text{span}\{v_1\} \).

2. Set \( v_2 = c (x_2 - P_{W_1} x_2) = c \left( x_2 - \frac{(x_2, v_1)}{\|v_1\|^2} v_1 \right) \) for any non-zero \( c \in \mathbb{F} \).
   Set \( W_2 = \text{span}\{v_1, v_2\} \).
   Note that \( W_2 = \text{span}\{x_1, x_2\} \).

3. Set \( v_3 = c (x_3 - P_{W_2} x_3) = c \left( x_3 - \frac{(x_3, v_1)}{\|v_1\|^2} v_1 - \frac{(x_3, v_2)}{\|v_2\|^2} v_2 \right) \) for any non-zero \( c \in \mathbb{F} \).
   Set \( W_3 = \text{span}\{v_1, v_2, v_3\} \).
   Note that \( W_3 = \text{span}\{x_1, x_2, x_3\} \).

4. Continue in this manner until \( v_k \) is determined (it’s not necessary to determine \( W_k \)).

Suppose \( V = \mathbb{R}^3 \).

Consider the basis \( \mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} \right\} \).

Let’s apply the Gram-Schmidt procedure to turn this into an orthogonal basis:

1. Set \( v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \) and \( W_1 = \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\} \).
(2) Consider that
\[
\begin{pmatrix}
1 \\
-1
\end{pmatrix}
- P_{W_1}
\begin{pmatrix}
1 \\
-1
\end{pmatrix}
= \begin{pmatrix}
1 \\
0
\end{pmatrix}
- \frac{1}{14}\left(\begin{pmatrix}
1 \\
-1
\end{pmatrix}, \begin{pmatrix}
1 \\
2
\end{pmatrix}\right)\begin{pmatrix}
1 \\
2
\end{pmatrix}
= \begin{pmatrix}
1 \\
-1
\end{pmatrix}
+ \frac{1}{14}\begin{pmatrix}
1 \\
2
\end{pmatrix}
= \begin{pmatrix}
\frac{15}{14} \\
\frac{3}{14}
\end{pmatrix}.
\]

So for convenience set
\[
v_2 = \begin{pmatrix}
5 \\
-4 \\
1
\end{pmatrix},
\]
and set \(W_2 = \text{span}\left\{\begin{pmatrix}1 \\ 2 \\ 3\end{pmatrix}, \begin{pmatrix}5 \\ -4 \\ 1\end{pmatrix}\right\}.
\]

(3) Consider that
\[
\begin{pmatrix}
2 \\
-1
\end{pmatrix}
- P_{W_2}
\begin{pmatrix}
2 \\
-1
\end{pmatrix}
= \begin{pmatrix}
2 \\
1
\end{pmatrix}
- \frac{1}{14}\left(\begin{pmatrix}
2 \\
-1
\end{pmatrix}, \begin{pmatrix}
1 \\
2
\end{pmatrix}\right)\begin{pmatrix}
1 \\
2
\end{pmatrix}
- \frac{1}{42}\left(\begin{pmatrix}
2 \\
-1
\end{pmatrix}, \begin{pmatrix}
5 \\
-4
\end{pmatrix}\right)\begin{pmatrix}
5 \\
-4
\end{pmatrix}
= \begin{pmatrix}
2 \\
-1
\end{pmatrix}
- \frac{1}{14}\begin{pmatrix}
1 \\
2
\end{pmatrix}
- \frac{5}{42}\begin{pmatrix}
5 \\
-4
\end{pmatrix}
= \begin{pmatrix}
\frac{4}{3} \\
\frac{-4}{3}
\end{pmatrix}.
\]

So for convenience set
\[
v_3 = \begin{pmatrix}
1 \\
1 \\
-1
\end{pmatrix}.
\]
In conclusion \( \mathcal{B}' = \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ \end{pmatrix}, \begin{pmatrix} 5 \\ -4 \\ 1 \\ \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ -1 \\ \end{pmatrix} \right\} \), is an orthogonal basis such that the first \( i \) vectors of \( \mathcal{B} \) and \( \mathcal{B}' \) have the same spans for \( i = 1, 2, 3 \).

Recall if \( W \) is a subspace of \( V \), then \( W^\perp = \{ v \in V | v \perp W \} \).

**Claim:** Let \( W \) be a subspace of \( V \). \( W^\perp \) (orthogonal complement) is also a subspace of \( V \).

This is pretty easy to justify:

First notice \( W^\perp \) is not empty as \( 0 \in W^\perp \).

Let \( u, v \in W^\perp \) and \( c \in \mathbb{F} \). Let \( w \in W \). Then

\[
(u + cv, w) = (u, w) + c(v, w) = 0.
\]

Thus \( u + cv \in W^\perp \) proving that \( W^\perp \) is a subspace of \( V \).

More generally, if \( U, W \) are subspaces of \( V \) then \( U + W = \{ u + w | u \in U, w \in W \} \) is called the **subspace sum**.

(ELFY) \( U + W \) is a subspace of \( V \) for any subspaces \( U, W \).

A subspace sum \( U + W \) is called a **direct sum**, denoted \( U \oplus W \), if \( U \cap W = \{ 0 \} \).

**Proposition:** Let \( U, W \) be subspaces of \( V \). \( V = U \oplus W \) iff each vector in \( v \in V \) has a unique representation \( v = u + w \) for \( u \in U, w \in W \).

**Proof:** Let \( U, W \) be subspaces of \( V \). Assume \( V = U \oplus W \).

Let \( v \in V \). Then there exists \( u \in U \) and \( w \in W \) such that \( v = u + w \). Suppose \( v = u_1 + w_1 \) where \( u_1 \in U \) and \( w_1 \in W \).

Then \( v = u + w = u_1 + w_1 \) implies \( u - u_1 = w - w_1 \).

Then \( u + u_1 = w - w_1 \in U \cap V = \{ 0 \} \). Hence \( u = u_1 \) and \( w = w_1 \), so we have uniqueness.
Now assume for all $v \in V$ there exists unique vectors $u \in U, w \in W$ such that $v = u + w$. We are assuming each vector in $V$ has a unique decomposition into a vector from $U$ plus a vector from $W$.

Clearly $V = U + W$. The question is whether it’s a direct sum. To yield a contradiction, suppose $U \cap W \neq \{0\}$. Let $y \in U \cap W$ such that $y \neq 0$.

Let $v \in V$. Since $V = U + W$ there exists $u \in U, w \in W$ such that $v = u + w$.

Set $u_1 = u + y \in U$. Since $y \neq 0, u_1 \neq u$.

Set $w_1 = w - y \in U$. Since $y \neq 0, w_1 \neq w$.

Then $u_1 + w_1 = u + w = v$, which contradicts the uniqueness of the decomposition. This completes the proof $\square$

Let $v \in V$ and $W$ be a subspace at $V$. Then let $v_1 = P_W v \in W$ and $v_2 = v - P_W v \in W^\perp$. We have already shown that $P_W v \in W$ is the unique vector for which $v - P_W v \in W^\perp$. Therefore we have a unique decomposition of $v$ into a vector in $W$ plus a vector in $W^\perp$. Hence $V = W \oplus W^\perp$.

(ELFY) For any subspace $W$ of $V$, we get $(W^\perp)^\perp = W$.

Suggested Homework: 3.2, 3.3, 3.4, 3.5, 3.7, and 3.8.
4. Least Squares Solution

Let $A \in M_{mn}$ and $b \in \mathbb{F}^m$. How do we determine an $x \in \mathbb{F}^n$ such that $Ax$ is as close as possible to $b$?

If we can find $x$ such that $\|Ax - b\| = 0$ then $Ax = b$ is consistent and we say $x$ is an exact solution.

In the case where we cannot find $x$ such that $\|Ax - b\| = 0$ then we call an $x$ which minimizes $\|Ax - b\|$ a least squares solution to $Ax = b$ (which is inconsistent).

Why least squares? Well, minimizing $\|Ax - b\|$ is equivalent to minimizing

$$\|Ax - b\|^2 = \sum_{i=1}^{m} |(Ax)_i - b_i|^2 = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{ij} x_j - b_i \right|^2.$$ 

Let $W = \text{range } A = \{Ax | x \in \mathbb{F}^n\}$. $W$ is a subspace of $\mathbb{F}^m$.

In particular, the minimum value of $\|Ax - b\|$ is the distance from $b$ to $W$, which is satisfied by an $x$ which solves the equation (*) $Ax = P_W b$.

Furthermore, if $v_1, v_2, ..., v_k$ is an orthogonal basis for $W = \text{range } A$ then we get an explicit formula for $P_W b$:

$$P_W b = \sum_{i=1}^{k} \frac{(b, v_i)}{\|v_i\|^2} v_i.$$ 

So, if we have a basis for $W$ (which we can create by taking columns of $A$ that are pivot columns) then with Gram-Schmidt we can create an orthogonal basis and compute $P_W b$ and then solve (*). That is actually a lot of work...maybe there is an easier method?

Here is a nice approach:

Let $A = [a_1 \ a_2 \ \cdots \ a_n]$.

$Ax = P_W b$ iff $(b - Ax) \perp W$ iff $(b - Ax) \perp a_i$ for $i = 1, 2, ..., n$.

Hence $Ax = P_W b$ iff $(b - Ax, a_i) = a_i^*(b - Ax) = 0$ for $i = 1, 2, ..., n$.

Here is the clever part: $(A^*(b - Ax))_i = a_i^*(b - Ax) = 0$ for $i = 1, 2, ..., n$.

So $Ax = P_W b$ iff $A^*(b - Ax) = 0$ iff $A^*Ax = A^*b$. 

So we get the following result:

So the easiest way to find a least squares solution to \( Ax = b \) is to solve the so-called normal equation \( A^*Ax = A^*b \).

Also for a least squares solution \( x \), \( P_Wb = Ax \)

In particular, this means that a least squares solution is unique iff \( A^*A \in M_{nn} \) is invertible.

In this case \( P_Wb = A(A^*A)^{-1}A^*b \), which would hold for all \( b \in \mathbb{F}^m \), so then the matrix of \( P_W \) is \( A(A^*A)^{-1}A^* \).

**Proposition:** Let \( A \in M_{mn} \). \( \text{Ker}\ A = \text{Ker}(A^*A) \). In other words, \( A^*A \) is invertible iff rank \( A = n \).

**Proof:** Let \( A \in M_{mn} \).

Suppose \( x \in \text{Ker} \ A. \) Then \( Ax = 0 \) and thus \( (A^*A)x = A^*(Ax) = A^*0 = 0 \). Hence \( x \in \text{Ker}(A^*A) \) and \( \text{Ker} \ A \subseteq \text{Ker}(A^*A) \).

Suppose \( x \in \text{Ker}(A^*A) \). Then \( (A^*A)x = A^*Ax = 0 \).

Thus \( \|Ax\|^2 = (Ax, Ax) = (Ax)^*Ax = (x^*A^*)Ax = x^*(A^*Ax) = x^*0 = 0 \). This implies that \( Ax = 0 \). So \( x \in \text{Ker} \ A \) and \( \text{Ker}(A^*A) \subseteq \text{Ker} \ A \).

This completes the proof \( □ \)

Line of best fit:

A classic application is finding a line \( y = mx + b \) to best fit some given data \( \{(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)\} \).

The best solution minimizes the following sum of square differences:

\[
\sum_{i=1}^{k} |mx_i + b - y_i|^2.
\]
Equivalently, we are looking for a least squares solution to

\[
\begin{pmatrix}
 x_1 & 1 \\
 x_2 & 1 \\
 . & . \\
 . & . \\
 . & . \\
 x_k & 1
\end{pmatrix}
\begin{pmatrix}
 m \\
 b
\end{pmatrix}
=
\begin{pmatrix}
 y_1 \\
 y_2 \\
 . \\
 . \\
 . \\
 y_k
\end{pmatrix}.
\]

As long as not all \( x \)-coordinates are the same, a least squares solution is unique.

**EXAMPLE:**

Suppose we have the following data points in the \( xy \)-plane: \{\( (1, 2), (3, 5), (4, 5), (6, 8), (6, 9), (7, 10) \}\). Find an equation for the line of best fit.

So we are looking for the least squares solution to

\[
\begin{pmatrix}
 1 & 1 \\
 3 & 1 \\
 4 & 1 \\
 6 & 1 \\
 6 & 1 \\
 7 & 1
\end{pmatrix}
\begin{pmatrix}
 m \\
 b
\end{pmatrix}
=
\begin{pmatrix}
 2 \\
 5 \\
 5 \\
 8 \\
 9 \\
 10
\end{pmatrix}.
\]

First we create the normal equation:

\[
\begin{pmatrix}
 1 & 3 & 4 & 6 & 6 & 7 \\
 1 & 1 & 1 & 1 & 1 & 1 \\
 3 & 1 \\
 4 & 1 \\
 6 & 1 \\
 6 & 1 \\
 7 & 1
\end{pmatrix}
\begin{pmatrix}
 m \\
 b
\end{pmatrix}
=
\begin{pmatrix}
 1 & 3 & 4 & 6 & 6 & 7 \\
 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
 2 \\
 5 \\
 5 \\
 8 \\
 9 \\
 10
\end{pmatrix}
\rightarrow
\begin{pmatrix}
 147 & 27 \\
 27 & 6
\end{pmatrix}
\begin{pmatrix}
 m \\
 b
\end{pmatrix}
=
\begin{pmatrix}
 209 \\
 39
\end{pmatrix}
\rightarrow
\begin{pmatrix}
 m \\
 b
\end{pmatrix}
=
\frac{1}{153}
\begin{pmatrix}
 6 & -27 \\
 -27 & 147
\end{pmatrix}
\begin{pmatrix}
 209 \\
 39
\end{pmatrix}
=
\begin{pmatrix}
 \frac{67}{51} \\
 \frac{39}{51}
\end{pmatrix}.
Hence the line of best fit is...

\[ y = \frac{67}{51} x + \frac{30}{51}. \]

General approach to curve-fitting or surface-fitting:

(1) Find equations your data should fit if there was an exact solution. (The equations must be linear with respect to the parameters.)

(2) Write these equations as a linear system in a matrix form (inconsistent unless there is an exact solution).

(3) Find the least squares solution (left-multiply both sides by the conjugate transpose, aka adjoint, of the coefficient matrix and solve).

Suggested Homework: 4.2, 4.3, and 4.4.
5. Adjoint of a Linear Transformation

Let $V = \mathbb{F}^n$. Let $A \in M_{mn}$ with entries in $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$). Recall the adjoint of $A$ is $A^* = \overline{A^T}$.

**Proposition:** Let $A \in M_{mn}$. For all $x, y \in V$,

$$(Ax, y) = (x, A^*y).$$

**Proof:** Let $A \in M_{mn}$ and $x, y \in V$. Then

$$(Ax, y) = (y, Ax) = (Ax)^*y = (x^*A^*)y = x^*(A^*y) = (A^*y, x) = (x, A^*y) \square$$

This proposition gives rise to the notion of an adjoint operator. Now let $V, W$ be (finite dimensional) inner product spaces and $T \in \mathcal{L}(V, W)$. $S \in \mathcal{L}(W, V)$ is called an adjoint operator if $(Tx, y) = (x, Sy)$ for all $x \in V$ and $y \in W$. We will use the notation $T^*$ for $S$.

Does it exist? Yes. Let $T \in \mathcal{L}(V, W)$. By having $\mathcal{B}, \mathcal{C}$ be orthonormal bases for $V, W$ respectively, define $T^* \in \mathcal{L}(W, V)$ by having $[T^*]_{\mathcal{B} \mathcal{C}} = ([T]_{\mathcal{C} \mathcal{B}})^*$.

Let's justify that it is indeed adjoint:

Let $x \in V$ and $y \in W$. Let $\mathcal{B} = \{v_1, v_2, ..., v_n\}$ and $\mathcal{C} = \{w_1, w_2, ..., w_m\}$ be orthonormal bases for $V, W$ respectively. Then $Tv_i = c_{i1}w_1 + c_{i2}w_2 + \cdots + c_{im}w_m$ and $y = d_1w_1 + d_2w_2 + \cdots + d_mw_m$, where $c_{ij}, d_k$ are scalars for $i = 1, 2, ..., m$, $j = 1, 2, ..., n$, and $k = 1, 2, ..., m$.

Then $(Tv_i, y) = \sum_{k=1}^{m} c_{ki}d_k = ([T]_{\mathcal{B} \mathcal{C}}[v_i]_{\mathcal{B}}, [y]_{\mathcal{C}}) = ([v_i]_{\mathcal{B}}, ([T]_{\mathcal{C} \mathcal{B}})^*[y]_{\mathcal{C}})$ for $i = 1, 2, ..., n$.

By linearity, we get $(Tx, y) = ([T]_{\mathcal{B} \mathcal{C}}[x]_{\mathcal{B}}, [y]_{\mathcal{C}}) = ([x]_{\mathcal{B}}, ([T]_{\mathcal{C} \mathcal{B}})^*[y]_{\mathcal{C}})$.

Also $T^*w_i = a_{i1}v_1 + a_{i2}v_2 + \cdots + a_{in}v_n$ and $x = b_1v_1 + b_2v_2 + \cdots + b_nv_n$, where $a_{ij}, b_k$ are scalars for $i = 1, 2, ..., n$, $j = 1, 2, ..., m$, and $k = 1, 2, ..., n$.

Then $(x, T^*w_i) = \sum_{k=1}^{n} a_{ki}b_k = ([x]_{\mathcal{B}}, [T^*]_{\mathcal{C} \mathcal{B}}[w_i]_{\mathcal{C}}) = ([x]_{\mathcal{B}}, ([T]_{\mathcal{B} \mathcal{C}})^*[w_i]_{\mathcal{C}})$ for $i = 1, 2, ..., m$.

By linearity, we get $(x, T^*y) = ([x]_{\mathcal{B}}, ([T]_{\mathcal{B} \mathcal{C}})^*[y]_{\mathcal{C}}) = (Tx, y)$.

So $T^*$ is indeed an adjoint operator to $T$. Is it unique? Suppose $S$ is another adjoint operator. Then for all $x \in V$ and $y \in W$, $(Tx, y) = (x, T^*y) = (x, Sy)$ implies that for all $y \in W$, $T^* = Sy$. Thus $S = T^*$. So the adjoint operator is unique.
**Proposition:** Let $T \in \mathcal{L}(V,W)$ where $V, W$ are inner product spaces. Then

1. $\text{Ker } T^* = (\text{range } T)^\perp$.
2. $\text{Ker } T = (\text{range } T^*)^\perp$.
3. $\text{range } T = (\text{Ker } T^*)^\perp$.
4. $\text{range } T^* = (\text{Ker } T)^\perp$.

**Proof:** First recall that for any subspace $U$, $(U^\perp)^\perp = U$.
Thus (1) and (3) are equivalent as well as (2) and (4).

It turns out for any linear transformation $T$, $(T^*)^* = T$ (ELFY - Use conjugate symmetry!).

Thus Thus (1) and (2) are equivalent as well as (3) and (4).

So we can just prove (1). It’s pretty straightforward:

$y \in (\text{range } T)^\perp$ iff $(Tx, y) = (x, T^*y) = 0$ for all $x \in V$. Also, $(x, T^*y) = 0$ for all $x \in V$ iff $T^*y = 0$, which means $y \in \text{Ker } T^*$
This completes the proof □

**Suggested Homework:** 5.2. (Hint: If $P_W$ is the orthogonal projection onto a subspace $W$ of an inner product space $V$ then $I_V - P_W$ is the orthogonal projection onto $W^\perp$.)
6. Isometries and Unitary Operators

Let \( V \) and \( W \) be inner product spaces over \( \mathbb{F} \). \( T \in \mathcal{L}(V, W) \) is an isometry if for all \( x \in V \), \( \|Tx\| = \|x\| \).

**Proposition:** \( T \in \mathcal{L}(V, W) \) is an isometry iff for all \( x, y \in V \),

\[
(Tx, Ty) = (x, y).
\]

**Proof:** Let’s start with the reverse direction. Suppose for all \( x, y \in V \),

\[
(Tx, Ty) = (x, y).
\]

Then for all \( x \in V \), \( \|Tx\| = \sqrt{(Tx, Tx)} = \sqrt{(x, x)} = \|x\| \). So \( T \) is an isometry.

Now assume \( T \) is an isometry. We split this direction into two cases, \( \mathbb{F} = \mathbb{R} \) and \( \mathbb{F} = \mathbb{C} \):

**Assume** \( \mathbb{F} = \mathbb{R} \). Let \( x, y \in V \). Then by the polarization identity,

\[
(Tx, Ty) = \frac{1}{4} \left( \|Tx + Ty\|^2 - \|Tx - Ty\|^2 \right) = \frac{1}{4} \left( \|T(x + y)\|^2 - \|T(x - y)\|^2 \right) = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 \right) = (x, y).
\]

**Assume** \( \mathbb{F} = \mathbb{C} \). Let \( x, y \in V \). Then by the polarization identity,

\[
(Tx, Ty) = \frac{1}{4} \left( \|Tx + Ty\|^2 - \|Tx - Ty\|^2 + i\|Tx + iTy\|^2 - i\|Tx - iTy\|^2 \right) = \frac{1}{4} \left( \|T(x + y)\|^2 - \|T(x - y)\|^2 + i\|T(x + iy)\|^2 - i\|T(x - iy)\|^2 \right) = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2 \right) = (x, y)
\]

So an isometry preserves the inner product.
Proposition: \( T \in \mathcal{L}(V,W) \) is an isometry iff \( T^*T = I_V \).

Proof: Let \( T \in \mathcal{L}(V,W) \) be an isometry. Let \( y \in V \). Then for all \( x \in V \), \( (x, T^*Ty) = (Tx, Ty) \). Hence, \( T^*Ty = y \) for all \( y \in V \). Thus \( T^*T = I_V \).

Now assume \( T^*T = I_V \). Let \( x, y \in V \). Then \( (Tx, Ty) = (T^*Ty, Ty) = (x, I_Vy) = (x, y) \). Hence, \( T^*T = I_V \) for all \( y \in V \). Thus \( T \) is an isometry. □

The proposition above guarantees an isometry is left-invertible, but not necessarily invertible. An isometry \( T \in \mathcal{L}(V,W) \) is a unitary operator if it is invertible.

An isometry \( T \in \mathcal{L}(V,W) \) is a unitary operator if it is invertible.

A matrix \( A \in M_{mn} \) is an isometry if \( A^*A = I_n \). Also, a square matrix \( U \in M_{nn} \) is unitary if \( U^*U = I_n \).

(ELFY) The columns of a matrix \( A \in M_{nn} \) form an orthonormal basis iff \( A \) is unitary.

If \( \mathbb{F} = \mathbb{R} \) then a unitary matrix is also called orthogonal.

Basic properties of unitary matrices: Let \( U \) be a unitary matrix.

1. \( U^{-1} = U^* \) is also unitary.

2. If \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \) is an orthonormal basis of \( \mathbb{F}^n \) then \( U\mathbf{v}_1, U\mathbf{v}_2, \ldots, U\mathbf{v}_n \) is an orthonormal basis of \( \mathbb{F}^n \).

Suppose \( V, W \) have the same dimension. Let \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \) be an orthonormal basis of \( V \) and \( \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n \) be an orthonormal basis of \( W \). Define \( T \in \mathcal{L}(V,W) \) by \( T\mathbf{v}_i = \mathbf{w}_i \) for \( i = 1, 2, \ldots, n \). Then \( T \) is a unitary operator:

Let \( \mathbf{x} \in V \). There exists \( c_1, c_2, \ldots, c_n \in \mathbb{F} \) such that \( \mathbf{x} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_n \mathbf{v}_n \) and \( \| \mathbf{x} \|^2 = |c_1|^2 + |c_2|^2 + \cdots + |c_n|^2 \).

Then

\[
\|Tx\|^2 = \left\| T \left( \sum_{i=1}^n c_i \mathbf{v}_i \right) \right\|^2 = \left\| \sum_{i=1}^n c_i T\mathbf{v}_i \right\|^2 = \left\| \sum_{i=1}^n c_i \mathbf{w}_i \right\|^2 = \sum_{i=1}^n |c_i|^2 = \| \mathbf{x} \|^2.
\]
**Proposition:** Let \( U \in M_{nn} \) be a unitary matrix.

1. \(|\det U| = 1\). (For an orthogonal matrix \( U \), \( \det U = \pm 1 \).)
2. If \( \lambda \) is an eigenvalue of \( U \) then \(|\lambda| = 1\).

**Proof:** Let \( U \in M_{nn} \) be a unitary matrix.

Since \( \det U^* = \overline{\det U} \), it follows that
\[
1 = \det I_n = \det(U^*U) = (\det U^*) (\det U) = \overline{\det U} \det U = |\det U|^2.
\]

Thus \(|\det U| = 1\), which proves (1).

Let \( \lambda \) be an eigenvalue of \( U \). Then there is a non-zero \( x \in \mathbb{F}^n \) such that \( Ux = \lambda x \).

Then \( \|x\| = \|Ux\| = \|\lambda x\| = |\lambda|\|x\| \) implies \(|\lambda| = 1\), which proves (2) \(\square\)

Two operators \( S, T \in \mathcal{L}(V) \) (or matrices \( S, T \in M_{nn} \)) are called unitarily equivalent if there exists and unitary operator \( U \in \mathcal{L}(V) \) (or matrix \( U \in M_{nn} \)) such that \( S = UTU^* \). In particular, since \( U^{-1} = U^* \), the operators/matrices would also be similar. However, not all similar operators/matrices are unitarily equivalent.

**Proposition:** \( A \in M_{nn} \) is unitarily equivalent to a diagonal matrix \( D \) iff \( A \) has an orthogonal basis of eigenvectors.

**Proof:** Let \( A \in M_{nn} \).

Assume \( A \) is unitarily equivalent to a diagonal matrix \( D \). Then there is a unitary matrix \( U \) such that \( A = UDU^* \). Since \( U^*U = I_n \), it is easy to check that the vectors \( Ue_i \) for \( i = 1, 2, ..., n \) are eigenvectors of \( A \) (ELFY). Finally, since the vectors \( e_1, e_2, ..., e_n \) form an orthogonal (orthonormal) basis, so do the eigenvectors \( Ue_1, Ue_2, ..., Ue_n \) since \( U \) is unitary.

Now assume \( A \) has an orthogonal basis of eigenvectors \( v_1, v_2, ..., v_n \). Let \( u_i = \frac{1}{\|v_i\|} v_i \) for \( i = 1, 2, ..., n \). Then \( B = \{u_1, u_2, ..., u_n\} \) is an orthonormal basis of eigenvectors of \( A \). Let \( D = [A]_B^B \) and \( U \) be the matrix whose columns are \( u_1, u_2, ..., u_n \). Clearly \( D \) is diagonal and \( U \) is unitary (columns are an orthonormal basis).

In particular, \( U = [I_n]_B^S \), where \( S \) is the standard basis. Hence \( U^* = U^{-1} = [I_n]_S^B \) and \( A = [A]_S^B = [I_n]_S^B [A]_B^B [I_n]_B^S = UDU^* \). So \( A \) is unitarily equivalent to \( D \) \(\square\)

**Suggested Homework:** 6.1, 6.4, 6.6, and 6.8.
7. Rigid Motions in $\mathbb{R}^n$

A rigid motion in an inner product space $V$ is a transformation $f : V \to V$ (not necessarily linear) that preserves distance, that is, for all $x, y \in V$, $\|f(x) - f(y)\| = \|x - y\|$.

Clearly every unitary operator $T \in \mathcal{L}(V)$ is a rigid motion.

A nice example of a rigid motion that is not a linear operator is translation:

Let $y \in V$ be non-zero. Let $f_y : V \to V$ be given by $f_y(x) = x + y$. This operator is not linear since $f_y(0) \neq 0$.

**Proposition:** Let $f$ be a rigid motion in a real inner product space $V$. Let $T : V \to V$ be given by $T(x) = f(x) - f(0)$. Then for all $x, y \in V$,

1. $\|T(x)\| = \|x\|$.
2. $\|T(x) - T(y)\| = \|x - y\|$.
3. $(T(x), T(y)) = (x, y)$.

**Proof:** Let $f$ be a rigid motion in a real inner product space $V$. Let $T : V \to V$ be given by $T(x) = f(x) - f(0)$. Let $x, y \in V$.

$\|T(x)\| = \|f(x) - f(0)\| = \|x - 0\| = \|x\|$.

$\|T(x) - T(y)\| = \|(f(x) - f(0)) - (f(y) - f(0))\| = \|f(x) - f(y)\| = \|x - y\|$.

Since $V$ is a real inner product space, it follows that

$\|x - y\|^2 = \|x\|^2 - 2(x, y) + \|y\|^2$ and $\|T(x) - T(y)\|^2 = \|T(x)\|^2 - 2(T(x), T(y)) + \|T(y)\|^2$.

By using (1) and (2) in the equations above, $(T(x), T(y)) = (x, y)$.
The following proposition states that every rigid motion in a real inner product space is a composition of a translation and an orthogonal transformation.

**Proposition:** Let \( f \) be a rigid motion in a real inner product space \( V \). Let \( T : V \to V \) be given by \( T(x) = f(x) - f(0) \). \( T \) is an orthogonal transformation (unitary operator in a real inner product space).

**Proof:** Let \( x, y \in V \) and \( c \in \mathbb{R} \). Consider that

\[
\|T(x + cy) - (T(x) + cT(y))\|^2 = \|(T(x + cy) - T(x)) - cT(y)\|^2 = \\
= \|T(x + cy) - T(x)\|^2 - 2c(T(x + cy) - T(x), T(y)) + c^2\|T(y)\|^2 = \\
= \|x + cy - x\|^2 - 2c(T(x + cy), T(y)) + 2c(T(x), T(y)) + c^2\|y\|^2 = \\
= c^2\|y\|^2 - 2c(x + cy, y) + 2c(x, y) + c^2\|y\|^2 = \\
= 2c^2\|y\|^2 - 2c(x, y) - 2c(cy, y) + 2c(x, y) = 2c^2\|y\|^2 - 2c^2(y, y) = 0.
\]

So \( T(x + cy) - (T(x) + cT(y)) = 0 \), which means \( T(x + cy) = T(x) + cT(y) \). Therefore \( T \) is linear, and thus by the previous proposition, \( T \) is an isometry. However, since \( T \) is an isometry from \( V \) to itself, it follows that \( T \) is invertible and thus unitary, which in a real inner product space is called an orthogonal transformation \( \square \)

Note: We will not cover section 8 of this chapter.