Norms and Inner Products

Garrett Thomas

July 21, 2018

1 About

This document is part of a series of notes about math and machine learning. You are free to distribute it as you wish. The latest version can be found at http://gwthomas.github.io/notes. Please report any errors to gwthomas@stanford.edu.

2 Norms

Norms generalize the notion of length from Euclidean space.

A **norm** on a vector space V is a function $\|\cdot\| : V \to \mathbb{R}$ that satisfies

(i) $\|v\| \geq 0$, with equality if and only if $v = 0$

(ii) $\|\alpha v\| = |\alpha|\|v\|$

(iii) $\|u + v\| \leq \|u\| + \|v\|$ (the **triangle inequality**)

for all $u, v \in V$ and all $\alpha \in \mathbb{F}$. A vector space endowed with a norm is called a **normed vector space**, or simply a **normed space**.

An important fact about norms is that they induce metrics, giving a notion of convergence in vector spaces.

Proposition 1. *If a vector space V is equipped with a norm $\|\cdot\| : V \to \mathbb{R}$, then*

$$d(u, v) \triangleq \|u - v\|$$

is a metric on V.

Proof. The axioms for metrics follow directly from those for norms. If $u, v, w \in V$, then

(i) $d(u, v) = \|u - v\| \geq 0$, with equality if and only if $u - v = 0$, i.e. $u = v$

(ii) $d(u, v) = \|u - v\| = \|(-1)(v - u)\| = |-1|\|v - u\| = d(v, u)$

(iii) $d(u, v) = \|u - v\| = \|u - w + w - v\| \leq \|u - w\| + \|w - v\| = d(u, w) + d(w, v)$

so d is a metric as claimed. □

Therefore any normed space is also a metric space. If a normed space is complete with respect to the distance metric induced by its norm, it is said to be a **Banach space**.

The metric induced by a norm automatically has the property of **translation invariance**, meaning that $d(u+w, v+w) = d(u, v)$ for any $u, v, w \in V$:

$$d(u+w, v+w) = \|(u+w) - (v+w)\| = \|u+w-v-w\| = \|u-v\| = d(u, v)$$

3 Inner products

An **inner product** on a vector space V over $\mathbb{F}$ is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ satisfying

(i) $\langle v, v \rangle \geq 0$, with equality if and only if $v = 0$

(ii) Linearity in the first slot: $\langle u+v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ and $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$

(iii) **Conjugate symmetry**: $\langle u, v \rangle = \overline{\langle v, u \rangle}$

for all $u, v, w \in V$ and all $\alpha \in \mathbb{F}$. A vector space endowed with an inner product is called an **inner product space**.

If $\mathbb{F} = \mathbb{R}$, the conjugate symmetry condition reduces to symmetry: $\langle u, v \rangle = \langle v, u \rangle$. Additionally, the combination of linearity and conjugate symmetry imply **sesquilinearity**:

$$\langle u, \alpha v \rangle = \overline{\langle \alpha v, u \rangle} = \overline{\alpha \langle v, u \rangle} = \overline{\alpha}\,\overline{\langle v, u \rangle} = \overline{\alpha} \langle u, v \rangle$$

It is also worth noting that $\langle 0, v \rangle = 0$ for any v because for any $u \in V$

$$\langle u, v \rangle = \langle u+0, v \rangle = \langle u, v \rangle + \langle 0, v \rangle$$

Proposition 2. *The function*

$$\langle x, y \rangle \triangleq x\overline{y}$$

is an inner product on $\mathbb{C}$ (considered as a vector space over $\mathbb{C}$).

Proof. If $x, y, z \in \mathbb{C}$ and $\alpha \in \mathbb{C}$, then

1. $\langle z, z \rangle = z\overline{z} = |z|^2 \geq 0$, with equality if and only if $z = 0$
2. $\langle x+y, z \rangle = (x+y)\overline{z} = x\overline{z} + y\overline{z} = \langle x, z \rangle + \langle y, z \rangle$, and $\langle \alpha x, y \rangle = (\alpha x)\overline{y} = \alpha(x\overline{y}) = \alpha \langle x, y \rangle$
3. $\langle x, y \rangle = x\overline{y} = \overline{y\overline{x}} = \overline{\langle y, x \rangle}$

so the function is an inner product as claimed. □

3.1 Orthogonality

Orthogonality generalizes the notion of perpendicularity from Euclidean space. Two vectors u and v are said to be **orthogonal** if $\langle u, v \rangle = 0$; we write $u \perp v$ for shorthand. A set of vectors $v_1, \dots, v_n$ is described as **pairwise orthogonal** (or just **orthogonal** for short) if $v_j \perp v_k$ for all $j \neq k$.

Proposition 3. *If $v_1, \dots, v_n$ are nonzero and pairwise orthogonal, then they are linearly independent.*

Proof. Assume $v_1, \dots, v_n$ are nonzero and pairwise orthogonal, and suppose $\alpha_1 v_1 + \dots \alpha_n v_n = 0$. Then for each $j = 1, \dots, n$,

$$0 = \langle 0, v_j \rangle = \langle \alpha_1 v_1 + \dots \alpha_n v_n, v_j \rangle = \alpha_1 \langle v_1, v_j \rangle + \dots + \alpha_n \langle v_n, v_j \rangle = \alpha_j \langle v_j, v_j \rangle$$

Since $v_j \neq 0$, $\langle v_j, v_j \rangle > 0$, so $\alpha_j = 0$. Hence $v_1, \dots, v_n$ are linearly independent. □

4 Norms induced by inner products

Any inner product induces a norm given by

$$\|v\| \triangleq \sqrt{\langle v, v \rangle}$$

Moreover, these norms have certain special properties related to the inner product. The notation $\|\cdot\|$ is not yet justified as we have not yet shown that this is in fact a norm. We need first a couple of intermediate results, which are useful in their own right.

4.1 Pythagorean Theorem

The well-known Pythagorean theorem generalizes naturally to arbitrary inner product spaces.

Proposition 4. *(Pythagorean theorem) If $u \perp v$, then*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Proof. Suppose $u \perp v$, i.e. $\langle u, v \rangle = 0$. Then

$$\|u + v\|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \langle v, u \rangle + \langle u, v \rangle + \langle v, v \rangle = \|u\|^2 + \|v\|^2$$

as claimed. □

It is clear that by repeatedly applying the Pythagorean theorem for two vectors, we can obtain a slight generalization: if $v_1, \dots, v_n$ are orthogonal, then

$$\|v_1 + \dots + v_n\|^2 = \|v_1\|^2 + \dots + \|v_n\|^2$$

4.2 Cauchy-Schwarz inequality

The **Cauchy-Schwarz inequality** is one of the most widespread and useful inequalities in mathematics.

Proposition 5. *If V is an inner product space, then*

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

for all $u, v \in V$. Equality holds exactly when u and v are linearly dependent.

Proof. If $v = 0$, then equality holds, for $|\langle u, 0 \rangle| = 0 = \|u\| 0 = \|u\| \|0\|$. So suppose $v \neq 0$. In this case we can define

$$w \triangleq \frac{\langle u, v \rangle}{\|v\|^2} v$$

which satisfies

$$\begin{aligned}
\langle w, u - w \rangle &= \left\langle \frac{\langle u, v \rangle}{\|v\|^2} v, u - \frac{\langle u, v \rangle}{\|v\|^2} v \right\rangle \\
&= \frac{\langle u, v \rangle}{\|v\|^2} \left(\langle v, u \rangle - \frac{\overline{\langle u, v \rangle}}{\|v\|^2} \langle v, v \rangle \right) \\
&= \frac{\langle u, v \rangle}{\|v\|^2} \left(\langle v, u \rangle - \frac{\langle v, u \rangle}{\|v\|^2} \|v\|^2 \right) \\
&= 0
\end{aligned}$$

Therefore, by the Pythagorean theorem,

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

which, after multiplying through by $\|v\|^2$, yields

$$|\langle u, v \rangle|^2 \leq \|u\|^2 \|v\|^2$$

The claimed inequality follows by taking square roots, since both sides are nonnegative.

Observe that the only inequality in the reasoning above comes from $\|u - w\|^2 \geq 0$, so equality holds if and only if

$$0 = u - w = u - \frac{\langle u, v \rangle}{\|v\|^2} v$$

which implies that u and v are linearly dependent. Conversely if u and v are linearly dependent, then $u = \alpha v$ for some $\alpha \in \mathbb{F}$, and

$$w = \frac{\langle \alpha v, v \rangle}{\|v\|^2} v = \frac{\alpha \langle v, v \rangle}{\|v\|^2} v = \alpha v$$

giving $u - w = 0$, so that equality holds. □

We now have all the tools needed to prove that $\| \cdot \|$ is in fact a norm.

Proposition 6. *If V is an inner product space, then*

$$\|v\| \triangleq \sqrt{\langle v, v \rangle}$$

is a norm on V.

Proof. The axioms for norms mostly follow directly from those for inner products, but the triangle inequality requires a bit of work. If $u, v \in V$ and $\alpha \in \mathbb{F}$, then

(i) $\|v\| = \sqrt{\langle v, v \rangle} \geq 0$ since $\langle v, v \rangle \geq 0$, with equality if and only if $v = 0$.

(ii) $\|\alpha v\| = \sqrt{\langle \alpha v, \alpha v\rangle} = \sqrt{\alpha\overline{\alpha}\langle v, v\rangle} = \sqrt{|\alpha|^2\langle v, v\rangle} = |\alpha|\sqrt{\langle v, v\rangle} = |\alpha|\|v\|$.

(iii) Using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle \\
&= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\
&= \|u\|^2 + \langle u, v\rangle + \overline{\langle u, v\rangle} + \|v\|^2 \\
&= \|u\|^2 + 2\,\mathrm{Re}\langle u, v\rangle + \|v\|^2 \\
&\leq \|u\|^2 + 2|\langle u, v\rangle| + \|v\|^2 \\
&\leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\
&= (\|u\| + \|v\|)^2
\end{aligned}$$

Taking square roots yields $\|u+v\| \leq \|u\| + \|v\|$, since both sides are nonnegative.

Hence $\|\cdot\|$ is a norm as claimed. □

Thus every inner product space is a normed space, and hence also a metric space. If an inner product space is complete with respect to the distance metric induced by its inner product, it is said to be a **Hilbert space**.

4.3 Orthonormality

A set of vectors $e_1, \dots, e_n$ are said to be **orthonormal** if they are orthogonal and have unit norm (i.e. $\|e_j\| = 1$ for each j). This pair of conditions can be concisely expressed as $\langle e_j, e_k\rangle = \delta_{jk}$, where δ_{jk} is the Kronecker delta.

Observe that if $v \neq 0$, then $v/\|v\|$ is a unit vector in the same "direction" as v (i.e. a positive scalar multiple of v), for

$$\left\| \frac{v}{\|v\|} \right\| = \frac{1}{\|v\|}\|v\| = 1$$

Recall that an orthogonal set of nonzero vectors are linearly independent. Unit vectors are necessarily nonzero (since $\|0\| = 0 \neq 1$), so an orthonormal set of vectors always forms a basis for its span. Not surprisingly, such a basis is referred to as an **orthonormal basis**. A nice property of orthonormal bases is that vectors' coefficients in terms of this basis can be computed via the inner product.

Proposition 7. *If $e_1, \dots, e_n$ is an orthonormal basis for V, then any $v \in V$ can be written*

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

Proof. Since $e_1, \dots, e_n$ is a basis for V, any $v \in V$ has a unique expansion $v = \alpha_1 e_1 + \cdots + \alpha_n e_n$ where $\alpha_1, \dots, \alpha_n \in \mathbb{F}$. Then for $j = 1, \dots, n$,

$$\begin{aligned}
\langle v, e_j\rangle &= \langle \alpha_1 e_1 + \cdots + \alpha_n e_n, e_j\rangle \\
&= \alpha_1\langle e_1, e_j\rangle + \cdots + \alpha_n\langle e_n, e_j\rangle \\
&= \alpha_j
\end{aligned}$$

since $\langle e_k, e_j\rangle = 0$ for $k \neq j$ and $\langle e_j, e_j\rangle = 1$. □

The norm of vectors expressed in an orthonormal basis is also easily found, for

$$\|\alpha_1 e_1 + \cdots + \alpha_n e_n\|^2 = \|\alpha_1 e_1\|^2 + \cdots + \|\alpha_n e_n\|^2 = |\alpha_1|^2 + \cdots + |\alpha_n|^2$$

by application of the Pythagorean theorem. Combining this with the previous result, we see

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2$$

4.4 The Gram-Schmidt process

The **Gram-Schmidt process** produces an orthonormal basis for a vector space from an arbitrary basis for the space.

Proposition 8. *Let $v_1, \dots, v_n \in V$ be linearly independent. Define*

$$\begin{aligned} w_j &\triangleq v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1} \\ e_j &\triangleq \frac{w_j}{\|w_j\|} \end{aligned}$$

for $j = 1, \dots, n$. Then for each $j = 1, \dots, n$, the set $w_1, \dots, w_j$ is an orthonormal basis for $\operatorname{span}\{v_1, \dots, v_j\}$.

Proof. We use induction on j. In the base case $j = 1$, we have $w_1 = v_1$, and $e_1 = w_1/\|w_1\| = v_1/\|v_1\|$. It is clear that $\|e_1\| = 1$ by construction, and that $\operatorname{span}\{e_1\} = \operatorname{span}\{v_1\}$ because e_1 is a nonzero scalar multiple of v_1.

Now suppose that $e_1, \dots, e_{j-1}$ is an orthonormal basis for $\operatorname{span}\{v_1, \dots, v_{j-1}\}$. First observe that w_j so defined cannot be 0, as $v_j \notin \operatorname{span}\{e_1, \dots, e_{j-1}\} = \operatorname{span}\{v_1, \dots, v_{j-1}\}$ due to the assumption of linear independence. Hence $\|w_j\| > 0$ and e_j is well-defined. Moreover, it is clear that $\|e_j\| = 1$ by construction.

We also see that w_j is orthogonal to $e_1, \dots, e_{j-1}$, since for $1 \leq k < j$ we have

$$\begin{aligned} \langle w_j, e_k\rangle &= \langle v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}, e_k\rangle \\ &= \langle v_j, e_k\rangle - \langle v_j, e_1\rangle\langle e_1, e_k\rangle - \cdots - \langle v_j, e_{j-1}\rangle\langle e_{j-1}, e_k\rangle \\ &= \langle v_j, e_k\rangle - \langle v_j, e_k\rangle \\ &= 0 \end{aligned}$$

since $e_1, \dots e_{j-1}$ are orthonormal. It follows that

$$\langle e_j, e_k\rangle = \left\langle \frac{w_j}{\|w_j\|}, e_k \right\rangle = \frac{\langle w_j, e_k\rangle}{\|w_j\|} = 0$$

for $k = 1, \dots, j - 1$, so $e_1, \dots, e_j$ are indeed orthonormal.

It remains to show that $\operatorname{span}\{e_1, \dots, e_j\} = \operatorname{span}\{v_1, \dots, v_j\}$. To this end, note that $w_j \in \operatorname{span}\{v_j, e_1, \dots, e_{j-1}\} = \operatorname{span}\{v_1, \dots, v_j\}$, where the equality of the spans follows from the assumption that $e_1, \dots, e_{j-1}$ is an orthonormal basis for $\operatorname{span}\{v_1, \dots, v_{j-1}\}$ and the fact that rescaling vectors by a nonzero amount does not affect their span. Thus $e_j \in \operatorname{span}\{v_1, \dots, v_j\}$ as well since it is a scalar multiple of w_j, so $\operatorname{span}\{e_1, \dots, e_j\} \subseteq \operatorname{span}\{v_1, \dots, v_j\}$. In the other direction, we have by definition that

$$v_j = w_j + \langle v_j, e_1\rangle e_1 + \cdots + \langle v_j, e_{j-1}\rangle e_{j-1}$$

so $v_j \in \operatorname{span}\{w_j, e_1, \dots, e_{j-1}\} = \operatorname{span}\{e_1, \dots, e_j\}$, again using the scale invariance of span. Hence $\operatorname{span}\{v_1, \dots, v_j\} \subseteq \operatorname{span}\{e_1, \dots, e_j\}$, which proves the result. □

Note that this result guarantees the existence of an orthonormal basis for any finite-dimensional vector space. Any such vector space has a basis, from which an orthonormal basis can be produced via the Gram-Schmidt process.

5 Orthogonal complements and projections

If $S \subseteq V$ where V is an inner product space, then the **orthogonal complement** of S, denoted $S^\perp$, is the set of all vectors in V that are orthogonal to every element of S:

$$S^\perp = \{v \in V : v \perp s \text{ for all } s \in S\}$$

Proposition 9. *If V is a vector space and $S \subseteq V$, then $S^\perp$ is a subspace of V.*

Proof. (i) $\langle 0, s\rangle = 0$ for any $s \in S$, so $0 \in S^\perp$.

(ii) If $u, v \in S^\perp$, then $\langle u, s\rangle = 0$ and $\langle v, s\rangle = 0$ for all $s \in S$, so

$$\langle u + v, s\rangle = \langle u, s\rangle + \langle v, s\rangle = 0 + 0 = 0$$

for all $s \in S$, i.e. $u + v \in S^\perp$.

(iii) If $v \in S^\perp$ and $\alpha \in \mathbb{F}$, then $\langle v, s\rangle = 0$ for all $s \in S$, so

$$\langle \alpha v, s\rangle = \alpha\langle v, s\rangle = \alpha 0 = 0$$

for all $s \in S$, i.e. $\alpha v \in S^\perp$.

Hence $S^\perp$ is a subspace of V. □

Note that there is no requirement that S itself be a subspace of V. However, if S *is* a (finite-dimensional) subspace of V, we have the following important decomposition.

Proposition 10. *Let V be an inner product space and S be a finite-dimensional subspace of V. Then every $v \in V$ can be written uniquely in the form*

$$v = v_S + v_\perp$$

where $v_S \in S$ and $v_\perp \in S^\perp$.

Proof. Let $e_1, \dots, e_n$ be an orthonormal basis for S, and suppose $v \in V$. Define

$$v_S = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

and

$$v_\perp = v - v_S$$

It is clear that $v_S \in S$ since it is in the span of the chosen basis. We also have, for all $j = 1, \dots, n$,

$$\begin{aligned}
\langle v_\perp, e_j\rangle &= \big\langle v - (\langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n), e_j\big\rangle \\
&= \langle v, e_j\rangle - \langle v, e_1\rangle\langle e_1, e_j\rangle - \cdots - \langle v, e_n\rangle\langle e_n, e_j\rangle \\
&= \langle v, e_j\rangle - \langle v, e_j\rangle \\
&= 0
\end{aligned}$$

which implies $v_\perp \in S^\perp$.

It remains to show that this decomposition is unique, i.e. doesn't depend on the choice of basis. To this end, let $\tilde{u}_1, \dots, \tilde{u}_m$ be another orthonormal basis for S, and define $\tilde{v}_S$ and $\tilde{v}_\perp$ analogously. We claim that $\tilde{v}_S = v_S$ and $\tilde{v}_\perp = v_\perp$.

By definition,

$$v_S + v_\perp = v = \tilde{v}_S + \tilde{v}_\perp$$

so

$$\underbrace{v_S - \tilde{v}_S}_{\in S} = \underbrace{\tilde{v}_\perp - v_\perp}_{\in S^\perp}$$

From the orthogonality of these subspaces, we have

$$0 = \langle v_S - \tilde{v}_S, \tilde{v}_\perp - v_\perp \rangle = \langle v_S - \tilde{v}_S, v_S - \tilde{v}_S \rangle = \|v_S - \tilde{v}_S\|^2$$

It follows that $v_S - \tilde{v}_S = 0$, i.e. $v_S = \tilde{v}_S$. Then $\tilde{v}_\perp = v - \tilde{v}_S = v - v_S = v_\perp$ as well. □

The existence and uniqueness of the decomposition above mean that

$$V = S \oplus S^\perp$$

whenever S is a subspace.

Since the mapping from v to v_S in the decomposition above always exists and is unique, we have a well-defined function

$$\begin{aligned} P_S : V &\to S \\ v &\mapsto v_S \end{aligned}$$

which is called the **orthogonal projection** onto S. We give the most important properties of this function below.

Proposition 11. *Let S be a finite-dimensional subspace of V. Then*

(i) *For any $v \in V$ and orthonormal basis $e_1, \dots, e_n$ of S,*

$$P_S v = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_n \rangle e_n$$

(ii) *For any $v \in V$, $v - P_S v \perp S$.*

(iii) *P_S is a linear map.*

(iv) *P_S is the identity when restricted to S (i.e. $P_S s = s$ for all $s \in S$).*

(v) $\text{range}(P_S) = S$ *and* $\text{null}(P_S) = S^\perp$.

(vi) $P_S^2 = P_S$.

(vii) *For any $v \in V$, $\|P_S v\| \leq \|v\|$.*

(viii) *For any $v \in V$ and $s \in S$,*

$$\|v - P_S v\| \leq \|v - s\|$$

with equality if and only if $s = P_S v$. That is,

$$P_S v = \arg\min_{s \in S} \|v - s\|$$

Proof. The first two statements are immediate from the definition of P_S and the work done in the proof of the previous proposition.

In this proof, we abbreviate $P = P_S$ for brevity.

(iii) Suppose $u, v \in V$ and $\alpha \in \mathbb{R}$. Write $u = u_S + u_\perp$ and $v = v_S + v_\perp$, where $u_S, v_S \in S$ and $u_\perp, v_\perp \in S^\perp$. Then

$$u + v = \underbrace{u_S + v_S}_{\in S} + \underbrace{u_\perp + v_\perp}_{\in S^\perp}$$

so $P(u + v) = u_S + v_S = Pu + Pv$, and

$$\alpha v = \underbrace{\alpha v_S}_{\in S} + \underbrace{\alpha v_\perp}_{\in S^\perp}$$

so $P(\alpha v) = \alpha v_S = \alpha P v$. Thus P is linear.

(iv) If $s \in S$, then we can write $s = s + 0$ where $s \in S$ and $0 \in S^\perp$, so $Ps = s$.

(v) $\text{range}(P) \subseteq S$: By definition.

$\text{range}(P) \supseteq S$: Using the previous result, any $s \in S$ satisfies $s = Pv$ for some $v \in V$ (specifically, $v = s$).

$\text{null}(P) \subseteq S^\perp$: Suppose $v \in \text{null}(P)$. Write $v = v_S + v_\perp$ where $v_S \in S$ and $v_\perp \in S^\perp$. Then $0 = Pv = v_S$, so $v = v_\perp \in S^\perp$.

$\text{null}(P) \supseteq S^\perp$: If $v \in S^\perp$, then $v = 0 + v$ where $0 \in S$ and $v \in S^\perp$, so $Pv = 0$.

(vi) For any $v \in V$,

$$P^2 v = P(Pv) = Pv$$

since $Pv \in S$ and P is the identity on S. Hence $P^2 = P$.

(vii) Suppose $v \in V$. Then by the Pythagorean theorem,

$$\|v\|^2 = \|Pv + (v - Pv)\|^2 = \|Pv\|^2 + \|v - Pv\|^2 \geq \|Pv\|^2$$

The result follows by taking square roots.

(viii) Suppose $v \in V$ and $s \in S$. Then by the Pythagorean theorem,-

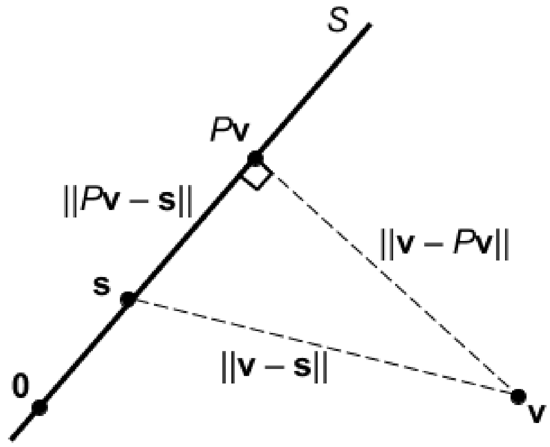
$$\|v - s\|^2 = \|(v - Pv) + (Pv - s)\|^2 = \|v - Pv\|^2 + \|Pv - s\|^2 \geq \|v - Pv\|^2$$

We obtain $\|v - s\| \geq \|v - Pv\|$ by taking square roots. Equality holds iff $\|Pv - s\|^2 = 0$, which is true iff $Pv = s$.

$\square$

Any linear map P that satisfies $P^2 = P$ is called a **projection**, so we have shown that P_S is a projection (hence the name).

The last part of the previous result shows that orthogonal projection solves the optimization problem of finding the closest point in S to a given $v \in V$. This makes intuitive sense from a pictorial representation of the orthogonal projection:



6 Products of inner product spaces

The following result shows that inner products can be combined additively when dealing with products of inner product spaces.

Proposition 12. *Let $V_1, \dots, V_n$ be inner product spaces. Then*

$$\langle (u_1, \dots, u_n), (v_1, \dots, v_n) \rangle \triangleq \langle u_1, v_1 \rangle_{V_1} + \dots + \langle u_n, v_n \rangle_{V_n}$$

is an inner product on $V_1 \times \dots \times V_n$.

Proof. (i) If $(v_1, \dots, v_n) \in V_1 \times \dots \times V_n$, then

$$\langle (v_1, \dots, v_n), (v_1, \dots, v_n) \rangle = \underbrace{\langle v_1, v_1 \rangle_{V_1}}_{\geq 0} + \dots + \underbrace{\langle v_n, v_n \rangle_{V_n}}_{\geq 0} \geq 0$$

and moreover if $\langle (v_1, \dots, v_n), (v_1, \dots, v_n) \rangle = 0$ we must have $\langle v_j, v_j \rangle_{V_j} = 0$ for all $j = 1, \dots, n$, implying that $v_j = 0$ in each case, and thus $(v_1, \dots, v_n) = 0$.

(ii) If $(u_1, \dots, u_n), (v_1, \dots, v_n), (w_1, \dots, w_n) \in V_1 \times \dots \times V_n$ and $\alpha \in \mathbb{F}$, then

$$\begin{aligned}
\langle (u_1, \dots, u_n) + (v_1, \dots, v_n), (w_1, \dots, w_n) \rangle &= \langle (u_1 + v_1, \dots, u_n + v_n), (w_1, \dots, w_n) \rangle \\
&= \langle u_1 + v_1, w_1 \rangle_{V_1} + \dots + \langle u_n + v_n, w_n \rangle_{V_n} \\
&= \langle u_1, w_1 \rangle_{V_1} + \langle v_1, w_1 \rangle_{V_1} + \dots + \langle u_n, w_n \rangle_{V_n} + \langle v_n, w_n \rangle_{V_n} \\
&= \langle (u_1, \dots, u_n), (w_1, \dots, w_n) \rangle + \langle (v_1, \dots, v_n), (w_1, \dots, w_n) \rangle
\end{aligned}$$

and

$$\begin{aligned}
\langle \alpha(u_1, \dots, u_n), (v_1, \dots, v_n) \rangle &= \langle (\alpha u_1, \dots, \alpha u_n), (v_1, \dots, v_n) \rangle \\
&= \langle \alpha u_1, v_1 \rangle_{V_1} + \dots + \langle \alpha u_n, v_n \rangle_{V_n} \\
&= \alpha \langle u_1, v_1 \rangle_{V_1} + \dots + \alpha \langle u_n, v_n \rangle_{V_n} \\
&= \alpha \langle (u_1, \dots, u_n), (v_1, \dots, v_n) \rangle
\end{aligned}$$

(iii) If $(u_1, \dots, u_n), (v_1, \dots, v_n) \in V_1 \times \cdots \times V_n$, then

$$\begin{aligned}
\langle (u_1, \dots, u_n), (v_1, \dots, v_n) \rangle &= \langle u_1, v_1 \rangle_{V_1} + \cdots + \langle u_n, v_n \rangle_{V_n} \\
&= \overline{\langle v_1, u_1 \rangle_{V_1}} + \cdots + \overline{\langle v_n, u_n \rangle_{V_n}} \\
&= \overline{\langle v_1, u_1 \rangle_{V_1} + \cdots + \langle v_n, u_n \rangle_{V_n}} \\
&= \overline{\langle (v_1, \dots, v_n), (u_1, \dots, u_n) \rangle}
\end{aligned}$$

Thus the function is an inner product as claimed. □

As a corollary of the above, we have that

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^{n} x_j \overline{y_j}$$

is an inner product on $\mathbb{C}^n$.