1. $W^\perp$ is a subspace.
2. $(W^\perp)^\perp = W$.
3. Let W be a subspace of R^n with dimension d, $W^\perp$ has dimension $n-d$. Moreover, W and $W^\perp$ separate $\mathbb{R}^n$.

Proof. Let $A \in \mathbb{R}^{n\times d}$ such that $col(A) = W$. By the Complement theorem, $col(A)^\perp = null(A^t)$. It follows from the Rank theorem that,

$$dim(null(A^t)) + dim(col(A^t)) = n. \tag{8}$$

Since $dim(col(A^t)) = dim(row(A)) = d$, this implies that $dim(null(A^t)) = dim(col(A)^\perp) = n-d$. $W \cap W^\perp = 0$, this implies that W and $W^\perp$ separate $\mathbb{R}^n$. □

6 Orthogonal projection

Definition 6.1. An orthogonal basis for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

Theorem 6.2. Let $\{u_1, \cdots, u_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^m$. For each y in W, the weights in the linear combination

$$y = c_1 u_1 + \cdots + c_p u_p \tag{9}$$

are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j}, \tag{10}$$

for all $j = 1, ..., p$. Moreover, if $\{u_1, \cdots, u_p\}$ is orthonormal, $c_j = y \cdot u_j$ for all j.

Proof. Inner product u_i on both side of the equation, this gives,

$$\langle y, u_i\rangle = c_i \langle u_i, u_i\rangle, \forall i = 1, ..., p. \tag{11}$$

It follows that $c_i = \frac{\langle y,u_i\rangle}{\langle u_i,u_i\rangle}$. When the set is orthonormal, $\langle u_i, u_i\rangle = 1$. □

Theorem 6.3. Let W be a subspace of $\mathbb{R}^n$. Then for each $y \in \mathbb{R}^n$, y can be uniquely written as:

$$y = \hat{y} + z,$$

where $\hat{y} \in W$ and $z \in W^\perp$. Moreover, let $\{u_1, \cdots, u_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$, then $\hat{y} = c_1 u_1 + \cdots + c_p u_p$, c_i are defined in theorem 6.2.

Target: want to show $\hat{y} \perp z$.

$z = y - \hat{y}$

$= y - \sum_{j=1}^{p} c_j u_j \quad (*)$

where $c_j = \langle y, u_j \rangle$

Proof. Let us first show that z is orthogonal to $\hat{y}$. For any u_i,

$$u_i \cdot z = y \cdot u_i - \sum_{j=1}^{p} c_j u_j \cdot u_i = 0,$$

where we use the theorem 6.2 in the last step. Next, let us consider the uniqueness. Suppose $y = \hat{y}_1 + z_1$, where $\hat{y}_1 \in W$ and $z_1 \in W^\perp$. We have $\hat{y} - \hat{y}_1 = z - z_1$, but $\hat{y} - \hat{y}_1 \in W$ and $z - z_1 \in W^\perp$ due to the closedness of subspace. This shows that $\hat{y} - \hat{y}_1 = z - z_1 = 0$. □

Inner product $(*)$ with all u_i

$\langle z, u_i \rangle = \langle y, u_i \rangle - c_i \underbrace{\langle u_i, u_i \rangle}_{= 1}$

$= 0$

Remark 5. By the previous theorem, given any y and W (a subspace with an orthonormal basis), there exist unique $\hat{y} \in W$, and $v \in W^{\perp}$, where $W \cap W^{\perp} = 0$. Based on the theorem, we will show that there exists a projector P such that $range(P) = W$ and $range(I - P) = W^{\perp}$. Because W and $W^{\perp}$ are orthogonal to each other, we call P an orthogonal projector. $\hat{y}$ is denoted by $\text{proj}_W y$ and is called the orthogonal projection of y onto W. Let us now find the projector P.

6.1 Orthogonal projector

range(Q)

An orthogonal projector projects onto S_1 along S_2, where S_1 and S_2 are orthogonal. Let $Q = [q_1, ..., q_n] \in \mathbb{R}^{m\times n}$ and q_i are orthonormal, and $S_1 = col(Q)$. In this section, we will construct an orthogonal projection onto the column space of Q. Let $v \in \mathbb{R}^m$, by Theorem $\boxed{6.3}$, we have

$$v = r + \sum_{i=1}^{n} (q_i q_i^t) v, \qquad \sum_{i=1}^{n} q_i q_i^t v = \underbrace{QQ^t}_{P} v = Pv = \hat{y} \tag{12}$$

$\| \langle q_i, q_i \rangle$

col-row form

and r is orthgonal with $\sum_{i=1}^{n} (q_i q_i^t) v$. Thus, the linear transform from v to $\sum_{i=1}^{n} (q_i q_i^t) v = QQ^t v$ is an orthogonal projection onto $range(Q)$. We claim that QQ^t is an orthogonal projector onto the column space of Q.

$P^2 = P$

Proof. First, we need to verify that QQ^T satisfies the definition of the projection. We then need to verify $col(QQ^T) = col(Q)$. Let $y = Qx$ for any $x \in \mathbb{R}^n$. Need to find a $y \in \mathbb{R}^m$ such that $x = Q^T y$. Since $rank(Q^t) = n$ by Theorem 6.7 in the last note, this implies that $\mathbb{R}^n = col(Q^t)$. That is, there always exists $y \in \mathbb{R}^m$ such that $x = Q^t y$. □

Remark 6. $dim(QQ^T) = n$, the projector has rank n. We claim that the complement projector has rank $m - n$. $\Leftrightarrow \dim(col(I - \underbrace{P}_{QQ^t})) = m - n$, $col(I-P) = null(P)$

$\dim(col(P))$ $\downarrow$

by the rank theorem, $\dim(col(I-P)) = m - n$.

An important special case is the rank-one projector that isolates the component in a single direction q. Specifically, $Q = [q]$, the projector onto the column space of Q is:

$$P_q = qq^t, \tag{13}$$

where $rank(P_q) = 1$ and SVD theorem explains this directly. The complements are the rank $m - 1$ orthogonal projectors:

$$\boxed{P_{\perp q}} = I - qq^t. \tag{14}$$

6.2 Projection with an arbitrary basis

We have discussed the projector construction if a set of orthonormal basis is given. Let us now consider the general case. Specifically, let $A \in \mathbb{R}^{m\times n}$, $S_1 = col(a_1, ..., a_n)$ and $rank(A) = n$. Given v, denote its projection onto $range(A)$ as y. Moreover, let $x \in \mathbb{R}^n$ such that $y = Ax$. Due to the orthogonal assumption (orthogonal projection is our target), we have $y - v$ orthogonal to $col(A)$, or it is orthogonal to all columns of A. That is, $a_i^t(y - v) = 0$ for all i. Equivalently $a_i^t(Ax - v) = 0$. Recall matrix multiplication (row-column rule); we have

$$A^t(Ax - v) = 0, \tag{15}$$

or $A^t Ax = A^t v$. Since $A^t A$ is invertible, $x = (A^t A)^{-1} A^t v$. The projection then follows:

$$y = A(A^t A)^{-1} A^t v, \tag{16}$$

or $A(A^t A)^{-1} A^t v$ is the projector onto $col(A)$.

Remark 5.

Question: S_1 & S_2, $\mathbb{R}^n = S_1 + S_2$, $S_1 \cap S_2 = 0$

Can we find a projector P, s.t. $\text{range}(P) = S_1$

$\text{range}(I-P) = S_2$

Now, how to find a projector P, s.t.

$$\begin{cases} \hat{y} = Py \\ z = (I-P)y \end{cases}$$

Def: orthogonal projection; given a projector P,

P is called an orthogonal projection, if

$\text{range}(P) \perp \text{range}(I-P)$

$\text{range}(QQ^+) = \text{range}(P)$

$\quad \parallel$

Pf $\quad \text{col}(QQ^+) = \text{col}(Q)$

It is easy to see $\text{col}(QQ^+) \subseteq \text{col}(Q)$

Let $y = Qx$, $x \in \mathbb{R}^n$ (arbitary vector in $\text{col}(Q)$)

Need to that, $\exists z \in \mathbb{R}^m$, s.t. $x = Q^+ z$. [then $y = QQ^+ z$,

or $y \in \text{col}(QQ^+)$]

Since $\text{rank}(Q^+) = \text{rank}(Q) = n$.

$\Rightarrow$ $\text{col}(Q) = \mathbb{R}^n$.

$\Rightarrow$ there always exists z s.t. $x = Q^+ z$. $\boxtimes$

6.2.

$S_1 = \text{col}(A)$.

do not require orthonormal.

$A = \{a_1\ a_2 \cdots a_n\} \in \mathbb{R}^{m \cdot n}$, $a_1\ a_2 \cdots a_n$ $\boxed{\text{are linear indep}}$

$v \in \mathbb{R}^m$. denote the projection of v onto $\text{col}(A)$ as y.

$\Leftrightarrow \exists\, x \in \mathbb{R}^n$, s.t. $y = Ax$

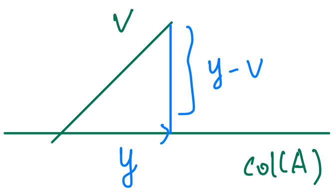
thm 6.3.

$v = y + z$

$y \perp z$

By the property of the orthogonal projection.

$y - v$ is orthogonal to $\text{col}(A)$.



$$\begin{matrix} \langle a_1, y-v \rangle = 0 \\ \cdots \quad \cdots \\ \langle a_n, y-v \rangle = 0 \end{matrix} \quad \Leftrightarrow \quad \begin{matrix} a_1^t (y-v) = 0 \\ \vdots \\ a_n^t (y-v) = 0 \end{matrix}$$

$$A^t (y-v) = 0$$

$$\ll \begin{pmatrix} a_1^t \\ \cdots \\ a_n^t \end{pmatrix}$$

Because $y = Ax$

$$\Rightarrow \quad A^t (Ax - v) = 0$$

$$A^t A x = A^t v$$

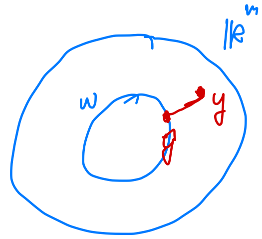
You need to show $A^t A$ is invertible.

$$x = (A^t A)^{-1} A^t v$$

$$y = Ax = \underbrace{A (A^t A)^{-1} A^t}_{P \text{ (orthogonal projector)}} v$$

$$= Pv$$

Theorem 6.4 (The best approximation theorem)**.** Let W be a subspace of $\mathbb{R}^m$, let y in $\mathbb{R}^m$ and $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the closest point in W to y in the sense that:



$$\|y - \hat{y}\| < \|y - v\|,$$

for any $v \in W$ and $v \neq \hat{y}$.

Proof. Let $v \in W$ and $v \neq \hat{y}$. It follows that,

$$y - v = y - \hat{y} + \hat{y} - v \tag{17}$$

We have $y - \hat{y}$ is orthogonal to W, by the Pythagorean theorem,

$$\|y - v\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - v\|^2.$$

□

6.3 Gram Schmidt process

The Gram–Schmidt process is a simple algorithm for producing an orthogonal or an orthonormal basis for any nonzero subspace of $\mathbb{R}^n$. Given a basis $x_1, ..., x_p$ for a nonzero subspace W, define

$$\begin{aligned}
v_1 &= x_1 \\
v_2 &= x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \\
v_3 &= x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2 \\
&\cdots \\
v_p &= x_p - -\frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}.
\end{aligned}$$

Then $v_1, ..., v_p$ is orthognal basis for W, i.e., $span\{x_1, x_2, ..., x_p\} = span\{v_1, v_2, ..., v_p\}$.

Pf of 6.4.

Let $v \in W$, $v \neq \hat{y}$

Need to prove $\underbrace{\|v - y\|}_{\text{distance between } y \,\&\, v.}$ is always bigger $\|v - \hat{y}\|$

$$y - v = \underbrace{y - \hat{y}}_{\in W^{\perp}} + \underbrace{\hat{y} - v}_{\in W}$$

orthogonal to each other.

Take $\| \ \|^2$,

$$\|y - v\|^2 = \|y - \hat{y} + \hat{y} - v\|^2$$

Pythagorean theorem

$$= \underbrace{\|y - \hat{y}\|^2}_{\text{target}} + \underbrace{\|\hat{y} - v\|^2}_{\geq 0}$$

$$\Rightarrow \quad \|y - v\|^2 > \|y - \hat{y}\|^2$$