Lecture 10: MATH 342W: Introduction to Data Science and Machine Learning

Sergio E. Garcia Tapia*

March 4, 2025 (last updated March 8, 2025)

Uniqueness of Orthogonal Projection Matrix

Let X be a matrix of dimension $n \times (p+1)$ (where $n > p+1$) and full rank. Then $\text{rank}(X) = p+1$. Let $X_\perp$ be a matrix of dimension $n \times (n - [p+1])$, where the columns $X_\perp$ are a basis of $co\ell[X]^\perp$ (the orthogonal complement of the column space of X). Then $\text{rank}(X_\perp) = n - (p+1)$.

Let H be an orthogonal projection matrix onto the column space of X. Since each column of X is already in the column space of X, we know (and indeed we have shown before) that

$$HX = X$$

On the other hand, since every column of $X_\perp$ is in the orthogonal complement of $co\ell[X]$, projecting the column with H yields $\mathbf{0}_n$ for each column:

$$HX_\perp = \mathbf{0}_{n\times(n-[p+1])}$$

Let $X_{\text{full}} = [X \mid X_\perp]$, meaning X_{full} is the matrix obtained by concatenating the columns of X and columns of $X_\perp$ into an $n \times n$ matrix. Since X and $X_\perp$ are each full rank, and since columns of X and those of $X_\perp$ are orthogonal, it follows that all columns combined make up a basis of $\mathbb{R}^n$. Thus, X_{full} an $n \times n$ matrix of full rank and hence invertible. We will use this fact to show that the orthogonal projection matrix H is unique.

Theorem 1. Let X be a matrix of dimension $n \times (p+1)$ that is full rank. Then the orthogonal projection matrix onto the column space of X is unique.

Proof. We have already shown that an orthogonal projection matrix exists, given by $H = X(X^\top X)^{-1}X^\top$. Suppose H' is also an orthogonal projection matrix onto the column space of X. Let $X_\perp$ be the matrix whose columns contain a basis of the $co\ell[X]^\perp$. Then the matrix $X_{\text{full}} = [X \mid X_\perp]$ is full rank. Since H and H' are both orthogonal projection matrices onto $co\ell[X]$, we have

$$H'X_{\text{full}} = [X \mid \mathbf{0}_{n\times(n-[p+1])}] = HX_{\text{full}}$$

*Based on lectures of Dr. Adam Kapelner at Queens College. See also the course GitHub page.

Since X_{full} is full rank, it is invertible, so

$$\begin{aligned}
H' X_{\text{full}} &= H X_{\text{full}} \\
(H' X_{\text{full}}) X_{\text{full}}^{-1} &= (H X_{\text{full}}) X_{\text{full}}^{-1} \\
H' (X_{\text{full}} X_{\text{full}}^{-1}) &= H (X_{\text{full}} X_{\text{full}}^{-1}) \\
H' I_n &= H I_n && (I_n \text{ is the identity matrix}) \\
H' &= H
\end{aligned}$$

$\square$

Sum of Projections

Let $\mathbf{a} \in \mathbb{R}^n$, and let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two linearly independent vectors in $\mathbb{R}^n$. Let $V = [\mathbf{v}_1 \mid \mathbf{v}_2]$, meaning V is an $n \times 2$ matrix whose columns are $\mathbf{v}_1$ and $\mathbf{v}_2$. Is the following statement true?

$$\underset{\mathcal{col}[V]}{\text{proj}}(\mathbf{a}) \stackrel{?}{=} \underset{\text{span}(\mathbf{v}_1)}{\text{proj}}(\mathbf{a}) + \underset{\text{span}(\mathbf{v}_2)}{\text{proj}}(\mathbf{a}) \tag{1}$$

Recall that if $\mathbf{u}$ is a vector, then

$$\underset{\text{span}(\mathbf{u}}{\text{proj}}(\mathbf{a}) = \mathbf{u}(\mathbf{u}^\top \mathbf{u})^{-1} \mathbf{u}^\top \mathbf{a} = \frac{\mathbf{u}\mathbf{u}^\top}{\|\mathbf{u}\|^2} = H_{\mathbf{u}} \mathbf{a}$$

Hence, if H_1 is the orthogonal projection matrix onto $\text{span}(\mathbf{v}_1)$, H_2 is the orthogonal projection matrix onto $\text{span}(\mathbf{v}_1)$, and H is the orthogonal projection matrix onto $\mathcal{col}[V]$, then Equation 1 is equivalent to

$$H\mathbf{a} \stackrel{?}{=} H_1 \mathbf{a} + H_2 \mathbf{a} = (H_1 + H_2)\mathbf{a}$$

Now, is $H_1 + H_2$ an orthogonal projection matrix? If so, then the uniqueness implied by Theorem 1 would imply $H = H_1 + H_2$. Recall that P is a projection matrix if $(\mathbf{v} - P\mathbf{v})^\top (P\mathbf{w}) = 0$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Here we let $P = H_1 + H_2$ and $\mathbf{v} = \mathbf{w} = \mathbf{a}$:

$$H_1 + H_2 \text{ is an orthogonal projection} \iff ((H_1 + H_2)\mathbf{a})^\top (\mathbf{a} - (H_1 + H_2)\mathbf{a}) = 0$$

In light of this equivalence, we will explore what it takes to make it true.

$$\begin{aligned}
0 &\stackrel{?}{=} ((H_1 + H_2)\mathbf{a})^\top (\mathbf{a} - (H_1 + H_2)\mathbf{a}) && \text{(Unverified conjecture)} \\
&= (H_1\mathbf{a} + H_2\mathbf{a})^\top (\mathbf{a} - (H_1\mathbf{a} + H_2\mathbf{a})) && \text{(Linearity)} \\
&= (H_1\mathbf{a} + H_2\mathbf{a})^\top \mathbf{a} - (H_1\mathbf{a} + H_2\mathbf{a})^\top (H_1\mathbf{a} + H_2\mathbf{a}) && \text{(Distributivity)} \\
&= ((H_1\mathbf{a})^\top + (H_2\mathbf{a})^\top)\mathbf{a} - ((H_1\mathbf{a})^\top + (H_2\mathbf{a})^\top)(H_1\mathbf{a} + H_2\mathbf{a}) && ((A+B)^\top = A^\top + B^\top) \\
&= (H_1\mathbf{a})^\top \mathbf{a} + (H_2\mathbf{a})^\top \mathbf{a} - ((H_1\mathbf{a})^\top H_1\mathbf{a} + (H_1\mathbf{a})^\top H_2\mathbf{a} + (H_2\mathbf{a})^\top H_1\mathbf{a} + (H_2\mathbf{a})^\top H_2\mathbf{a}) \\
&= \mathbf{a}^\top H_1^\top \mathbf{a} + \mathbf{a}^\top H_2^\top \mathbf{a} - \mathbf{a}^\top H_1^\top H_1 \mathbf{a} - \mathbf{a}^\top H_1^\top H_2 \mathbf{a} - \mathbf{a}^\top H_2^\top H_1 \mathbf{a} - \mathbf{a}^\top H_2^\top H_2 \mathbf{a} \\
&= \mathbf{a}^\top \underbrace{(H_1^\top - H_1^\top H_1)}_{0} \mathbf{a} + \mathbf{a}^\top \underbrace{(H_2^\top - H_2^\top H_2)}_{0} \mathbf{a} - \mathbf{a}^\top H_1^\top H_2 \mathbf{a} - \mathbf{a}^\top H_2^\top H_1 \mathbf{a} \\
&= -(H_1\mathbf{a})^\top H_2\mathbf{a} - (H_2\mathbf{a})^\top H_1\mathbf{a} && (H_1 \text{ and } H_2 \text{ are idempotent and symmetric}) \\
&= -(H_1\mathbf{a}) \cdot (H_2\mathbf{a}) - (H_2\mathbf{a}) \cdot (H_1\mathbf{a}) && \text{(Definition of dot product)} \\
&= -2(H_1\mathbf{a}) \cdot (H_2\mathbf{a}) && \text{(Dot product is commutative)}
\end{aligned}$$

The last equality is zero if and only if the dot product is zero. Now recall that $\mathbf{a}$ is arbitrary, that $H_1\mathbf{a}$ projects $\mathbf{a}$ onto the one-dimensional subspace span$(\mathbf{v}_1)$, and $H_2\mathbf{a}$ projects $\mathbf{a}$ onto the one-dimensional subspace span$(\mathbf{v}_2)$. Thus there exists $c_1, c_2 \in \mathbb{R}$ such that $H_1\mathbf{a} = c_1\mathbf{v}_1$ and $H_2\mathbf{a} = c_2\mathbf{v}_2$, so we may write

$$\begin{aligned}(H_1\mathbf{a}) \cdot (H_2\mathbf{a}) &= (c_1\mathbf{v}_1) \cdot (c_2\mathbf{v}_2) \\ &= c_1c_2(\mathbf{v}_1 \cdot \mathbf{v}_2) \\ &= c_1c_2(\|\mathbf{v}_1\|\|\mathbf{v}_2\| \cos\theta)\end{aligned}$$

The last equality is a consequence of the Schwarz inequality (see [Axl23]), and θ represents the angle between $\mathbf{v}_1$ and $\mathbf{v}_2$. In order for this to hold for all $\mathbf{a}$, we require $\cos\theta = 0$, and hence $\theta = 90°$. We have verified the following fact.

Theorem 2. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be two linearly independent vectors in $\mathbb{R}^n$. Let $V = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$, $V_1 = \text{span}(\mathbf{v}_1)$, and $V_2 = \text{span}(\mathbf{v}_2)$. Then, for all $\mathbf{a} \in \mathbb{R}^n$,

$$\text{proj}_V(\mathbf{a}) = \text{proj}_{V_1}(\mathbf{a}) + \text{proj}_{V_2}(\mathbf{a})$$

if and only if $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.

Corollary 1. Let $\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_d$ be a linearly independent list of vectors in $\mathbb{R}^n$. Let $V = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d)$, $V_k = \text{span}(\mathbf{v}_k)$ for all k such that $1 \le k \le d$. Then, for all $\mathbf{a} \in \mathbb{R}^n$,

$$\text{proj}_V(\mathbf{a}) = \sum_{k=1}^{d} \text{proj}_{V_k}(\mathbf{a})$$

if and only if the list $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d)$ is orthogonal (meaning all vectors are pairwise orthogonal).

See Figure 1 for a depiction of the statement Theorem 2. Effectively, the theorem says that $H_1 + H_2$ is not necessarily an orthogonal projection.

Projection Matrices onto Dimension 1 Subspaces

Suppose $\mathbf{v}_k$ is a nonzero vector in $\mathbb{R}^n$. Recall from linear algebra that

$$\underset{\text{span}(\mathbf{v}_k)}{\text{proj}}(\mathbf{a}) := \frac{\mathbf{v}_k^\top \mathbf{a}}{\|\mathbf{v}_k\|^2}\mathbf{v}_k$$

Let's focus on the portion that says $(\mathbf{v}_k^\top \mathbf{a})\mathbf{v}_k$. It turns out this is equivalent to $\mathbf{v}_k\mathbf{v}_k^\top \mathbf{a}$. Suppose that we let $\mathbf{u}$ be their difference:

$$\mathbf{u} := (\mathbf{v}_k^\top \mathbf{a})\mathbf{v}_k - (\mathbf{v}_k\mathbf{v}_k^\top \mathbf{a})$$

Now let $\mathbf{b}$ be any vector in $\mathbb{R}^n$. Multiplying by $\mathbf{b}^\top$ on the left we get:

$$\begin{aligned}\mathbf{b}^\top\mathbf{u} &= \mathbf{b}^\top(\mathbf{v}_k^\top \mathbf{a})\mathbf{v}_k - \mathbf{b}^\top(\mathbf{v}_k\mathbf{v}_k^\top \mathbf{a}) \\ &= (\mathbf{v}_k^\top \mathbf{a})\mathbf{b}^\top\mathbf{v}_k - (\mathbf{b}^\top\mathbf{v}_k)(\mathbf{v}_k^\top \mathbf{a}) && \text{(Associativity: } A(BC) = (AB)C\text{)} \\ &= 0\end{aligned}$$

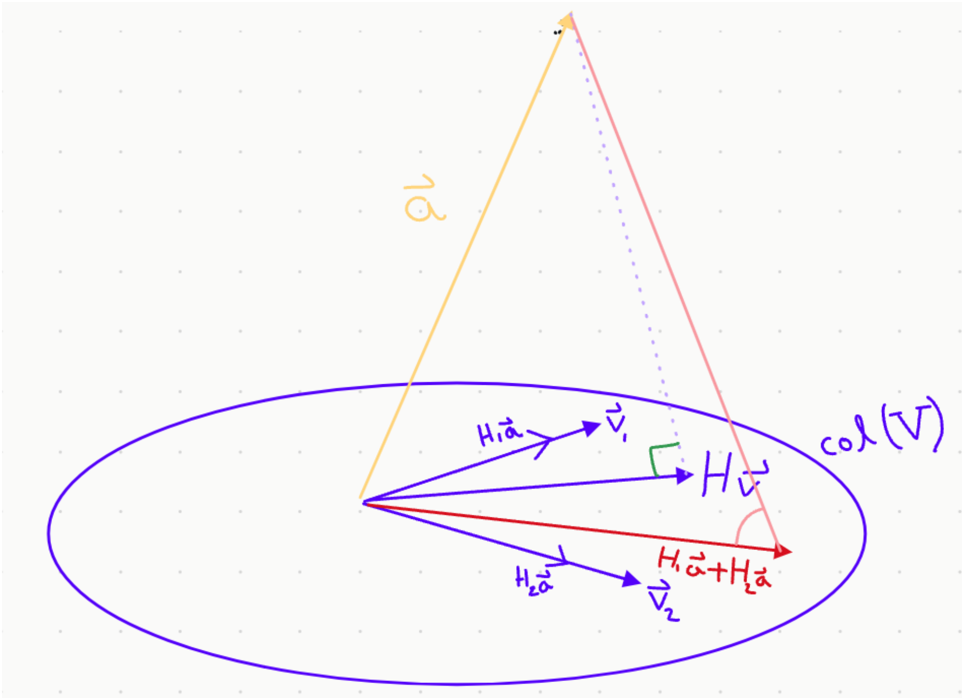

Figure 1: Projecting $\mathbf{a}$ with H_1, H_2, and H.

Since $\mathbf{b}$ is arbitrary, this implies $\mathbf{u}$ is orthogonal to every vector in $\mathbb{R}^n$, and so it must be the zero vector. Thus we have:

$$\operatorname*{proj}_{\text{span}(\mathbf{v}_k)}(\mathbf{a}) = \frac{\mathbf{v}_k^\top \mathbf{a}}{\|\mathbf{v}_k\|^2}\mathbf{v}_k = \frac{\mathbf{v}_k \mathbf{v}_k^\top}{\|\mathbf{v}_k\|^2}\mathbf{a} \tag{2}$$

Notice that since $\mathbf{v}_k$ is $n \times 1$, the product $\mathbf{v}_k \mathbf{v}_k^\top$ is an $n \times n$ matrix, which in fact is of rank 1 because every row is a multiple of $\mathbf{v}_k^\top$.

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$ be a mutually orthogonal list of (nonzero) vectors in $\mathbb{R}^n$, meaning that $\mathbf{v}_k^\top \mathbf{v}_j = 0$ for all $1 \leq k, j \leq d$ where $k \neq j$. Let $V = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_d]$ be an $n \times d$ matrix made by concatenating the vectors in this orthogonal list. Let $\mathbf{a} \in \mathbb{R}^n$. Then

$$\begin{aligned}
\operatorname*{proj}_{\mathit{col}[V]}(\mathbf{a}) &= \sum_{k=1}^{d} \operatorname*{proj}_{\text{span}(\mathbf{v}_k)}(\mathbf{a}) && \text{(Corollary 1)} \\
&= \sum_{k=1}^{d} \frac{\mathbf{v}_k \mathbf{v}_k^\top}{\|\mathbf{v}_k\|^2}\mathbf{a} && \text{(Equation 2)} \\
&= \underbrace{\left(\sum_{k=1}^{d} \frac{\mathbf{v}_k \mathbf{v}_k^\top}{\|\mathbf{v}_k\|^2}\right)}_{H}\mathbf{a} \\
&= H\mathbf{a}
\end{aligned}$$

Again by Corollary 1, H is an orthogonal projection matrix onto $\mathit{col}[V]$. In particular, it can be expressed as a sum of d orthogonal projection matrices, each of rank 1.

Let's simplify the sum by *normalizing* the vectors $\mathbf{v}_k$, which means we re-scale them to have length 1. Let $\mathbf{q}_k := \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$, so $\|\mathbf{q}_k\| = 1$. The list $(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_d)$ is still orthogonal because rescaling a vector does not change its direction, but since they are normalized now we refer to it as an *orthonormal list*. If we let $Q := [\mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots |\mathbf{q}_d]$, then we refer to Q as an *orthonormal matrix*. Now we can write

$$\begin{aligned}
\underset{co\ell[Q]}{\text{proj}}(\mathbf{a}) &= \sum_{k=1}^{d} \underset{\text{span}(\mathbf{q}_k)}{\text{proj}}(\mathbf{a}) && \text{(Corollary 1)} \\
&= \sum_{k=1}^{d} \frac{\mathbf{q}_k \mathbf{q}_k^\top}{\|\mathbf{q}_k\|^2}\mathbf{a} && \text{(Equation 2)} \\
&= \sum_{k=1}^{d} (\mathbf{q}_k \mathbf{q}_k^\top)\mathbf{a} && (\text{since } \|\mathbf{q}_k\| = 1)
\end{aligned}$$

Thus, rather than our old formula $X(X^\top X)^{-1}X^\top$ for a projection matrix, we can use $\sum_{k=1}^{d} \mathbf{q}_k \mathbf{q}_k^\top$, where we sum d matrices each of rank 1.

Properties of Orthonormal Matrices

Let Q by an $n \times d$ orthonormal matrix. Then $Q^\top$ is $d \times n$. Recall that the columns are pairwise orthogonal, so

$$\mathbf{q}_k^\top \mathbf{q}_j = \begin{cases} 1 & k = j \\ 0 & k \neq j \end{cases}$$

so

$$Q^\top Q = \begin{bmatrix} \cdots & \mathbf{q}_1^\top & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{q}_d^\top & \cdots \end{bmatrix} \begin{bmatrix} \vdots & \cdots & \vdots \\ \mathbf{q}_1 & \cdots & \mathbf{q}_d \\ \vdots & \cdots & \vdots \end{bmatrix} = I_{d\times d}$$

where $I_{d\times d}$ is the identity matrix. What about $QQ^\top$? This is an $n \times n$ matrix:

$$\begin{aligned}
QQ^\top &= \begin{bmatrix} \vdots & \cdots & \vdots \\ \mathbf{q}_1 & \cdots & \mathbf{q}_d \\ \vdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} \cdots & \mathbf{q}_1^\top & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \mathbf{q}_d^\top & \cdots \end{bmatrix} \\
&= \mathbf{q}_1\mathbf{q}_1^\top + \mathbf{q}_2\mathbf{q}_2^\top + \cdots \mathbf{q}_d\mathbf{q}_d^\top \\
&= H
\end{aligned}$$

Let's verify that $QQ^\top$ is indeed an orthogonal projection matrix:

(i) **Symmetry**:

$$(QQ^\top)^\top = (Q^\top)^\top Q^\top = QQ^\top$$

(ii) **Idempotency**:

$$(QQ^\top)^2 = (QQ^\top)(QQ^\top) = Q\underbrace{(Q^\top Q)}_{I}Q^\top = QQ^\top$$

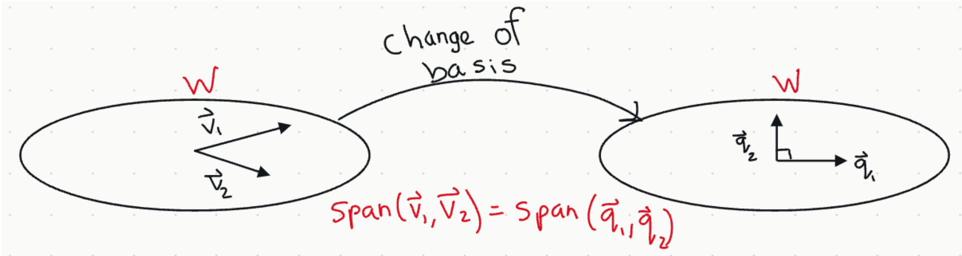

Figure 2: Change from linearly independent basis $\mathbf{v}_1, \mathbf{v}_2$ to orthogonal basis $\mathbf{q}_1, \mathbf{q}_2$ without changing the subspace

Gram Schmidt Orthogonalization and QR Decomposition

Suppose X is an $n \times (p+1)$ matrix with linearly independent columns (i.e., full rank). If yu can express X as an orthonormal matrix Q, then you can just use $QQ^\top$ to compute the orthogonal projection matrix rather than $X(X^\top X)^{-1}X$. Can we convert any full rank matrix X into an orthonormal matrix Q while presenting the columns pace? Yes! This entails a change of basis, and this yields a decomposition for X in the form

$$\underbrace{X}_{n\times(p+1)} = \underbrace{Q}_{n\times(p+1)} \underbrace{R}_{(p+1)\times(p+1)} \tag{3}$$

(see Figure 2). Equation 3 is known as the **QR factorization of** X. Here, Q is an orthonormal matrix such that $co\ell[Q] = co\ell[X]$, and R is a full rank matrix. To obtain Q, we can use the **Gram-Schmidt Orthogonalization procedure**. The high-level steps are as follows:

(i) Find an orthogonal set of basis vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_p$.

(ii) Normalize the vectors in the orthogonal set by dividing by their lengths:

$$\mathbf{q}_0 = \frac{\mathbf{v}_0}{\|\mathbf{v}_0\|}, \quad \mathbf{q}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \quad \cdots \quad \mathbf{q}_p = \frac{\mathbf{v}_p}{\|\mathbf{v}_p\|}$$

(iii) Tabulate the entries of R.

The crucial step is the first one, which we explain in detail. Suppose the columns of X are denoted $\mathbf{x}_{\cdot,0}, \mathbf{x}_{\cdot,1}, \ldots, \mathbf{x}_{\cdot,p}$.

(i) We compute the orthogonal set in $p+1$ steps.

- In step 0, set $\mathbf{v}_0 = \mathbf{x}_{\cdot,0}$.
- In step 1, set

$$\mathbf{v}_1 = \mathbf{x}_{\cdot,1} - \underset{\text{span}(\mathbf{v}_0)}{\text{proj}}(\mathbf{x}_{\cdot,1})$$

By removing the component of $\mathbf{x}_{\cdot,1}$ in the direction of $\mathbf{v}_0$, we accomplish the goal of making $\mathbf{v}_0$ and $\mathbf{v}_1$ orthogonal (see Figure 3).

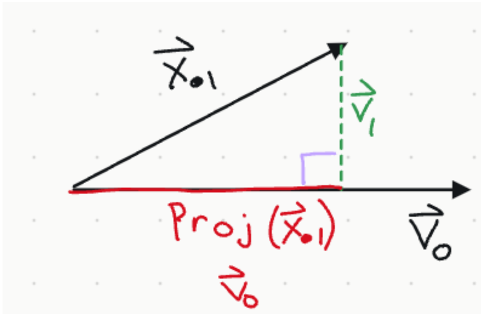

Figure 3: Removing the component of $\mathbf{x}_{\cdot,1}$ in the direction of $\mathbf{v}_0$.

- In step 2, set

$$\mathbf{v}_2 = \mathbf{x}_{\cdot,2} - \underset{\text{span}(\mathbf{v}_0)}{\text{proj}}(\mathbf{x}_{\cdot,2}) - \underset{\text{span}(\mathbf{v}_1)}{\text{proj}}(\mathbf{x}_{\cdot,2})$$

- In general, for step k, where $1 \leq k \leq p$, we have

$$\mathbf{v}_k = \mathbf{x}_{\cdot,k} - \sum_{j=0}^{k-1} \underset{\text{span}(\mathbf{v}_j)}{\text{proj}}(\mathbf{x}_{\cdot,k})$$

(ii) Now we simply normalize $\mathbf{v}_k$ for $0 \leq k \leq p$:

$$\mathbf{q}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$$

(iii) We will discuss this last step next time.

References

[Axl23] Sheldon Axler. *Linear Algebra Done Right*. 4th ed. Springer, 2023. ISBN: 9783031410253.