

Discussion 8B

The following notes are useful for this discussion: [Note 13](#).

1. Gram-Schmidt Algorithm

Let's apply Gram-Schmidt orthonormalization to a list of three linearly independent vectors $[\vec{s}_1, \vec{s}_2, \vec{s}_3]$.

- (a) Let's say we had two collections of vectors $\{\vec{v}_1, \dots, \vec{v}_n\}$ and $\{\vec{w}_1, \dots, \vec{w}_n\}$. **How can we prove that $\text{Span}(\{\vec{v}_1, \dots, \vec{v}_n\}) = \text{Span}(\{\vec{w}_1, \dots, \vec{w}_n\})$?**

Solution: Notice that taking the span of some vectors gives you a set of vectors. So, when proving two sets S_1 and S_2 are equal, we can show that $S_1 \subseteq S_2$ and $S_2 \subseteq S_1$. We can show $S_1 \subseteq S_2$ by showing that, if $a \in S_1$, then $a \in S_2$. Likewise, we can show $S_2 \subseteq S_1$ by showing that, if $b \in S_2$, then $b \in S_1$.

In the context of the given problem, we can show that $\text{Span}(\{\vec{v}_1, \dots, \vec{v}_n\}) = \text{Span}(\{\vec{w}_1, \dots, \vec{w}_n\})$ by first showing $\text{Span}(\{\vec{v}_1, \dots, \vec{v}_n\}) \subseteq \text{Span}(\{\vec{w}_1, \dots, \vec{w}_n\})$. That is, we can show that $\vec{v}_i \in \text{Span}(\{\vec{w}_1, \dots, \vec{w}_n\})$ for every $i = 1$ to $i = n$. Next, we can show $\text{Span}(\{\vec{w}_1, \dots, \vec{w}_n\}) \subseteq \text{Span}(\{\vec{v}_1, \dots, \vec{v}_n\})$ by showing that $\vec{w}_i \in \text{Span}(\{\vec{v}_1, \dots, \vec{v}_n\})$ for every $i = 1$ to $i = n$.

- (b) **Find unit vector \vec{q}_1 such that $\text{Span}(\{\vec{q}_1\}) = \text{Span}(\{\vec{s}_1\})$, where \vec{s}_1 is nonzero.**

Solution: Note that any $\vec{v} \in \text{Span}(\{\vec{s}_1\})$ can be written as $\vec{v} = a\vec{s}_1$ for some $a \in \mathbb{R}$. We need $\vec{q}_1 \in \text{Span}(\{\vec{s}_1\})$ and we need it to be a unit vector. Hence, we can write

$$\vec{q}_1 = \frac{\vec{s}_1}{\|\vec{s}_1\|}. \quad (1)$$

Next, we need to show $\vec{s}_1 \in \text{Span}(\{\vec{q}_1\})$. We can see that $\vec{s}_1 = a\vec{q}_1$ where $a = \|\vec{s}_1\|$.

- (c) Let's say that we wanted to write

$$\vec{s}_2 = c_1\vec{q}_1 + \vec{z}_2 \quad (2)$$

where $c_1\vec{q}_1$ entirely represents the component of \vec{s}_2 in the direction of \vec{q}_1 , and \vec{z}_2 represents the component of \vec{s}_2 that is distinctly *not* in the direction of \vec{q}_1 (i.e. \vec{z}_2 and \vec{q}_1 are orthogonal).

Given \vec{q}_1 from the previous step, **find c_1 as in eq. (2), and use \vec{z}_2 to find unit vector \vec{q}_2 such that $\text{Span}(\{\vec{q}_1, \vec{q}_2\}) = \text{Span}(\{\vec{s}_1, \vec{s}_2\})$ and \vec{q}_2 is orthogonal to \vec{q}_1 . Show that $\text{Span}(\{\vec{q}_1, \vec{q}_2\}) = \text{Span}(\{\vec{s}_1, \vec{s}_2\})$.**

Solution: To find c_1 , we can compute the projection of \vec{s}_2 onto \vec{q}_1 , namely

$$\text{proj}_{\vec{q}_1}(\vec{s}_2) = \frac{\vec{q}_1^\top \vec{s}_2}{\underbrace{\left(\vec{q}_1^\top \vec{q}_1\right)}_1} \vec{q}_1 = \underbrace{\left(\vec{q}_1^\top \vec{s}_2\right)}_{c_1} \vec{q}_1 \quad (3)$$

This projection represents all the components of \vec{s}_2 that are in the direction of \vec{q}_1 . To find \vec{z}_2 , we can use eq. (2) to obtain

$$\vec{z}_2 = \vec{s}_2 - \left(\vec{q}_1^\top \vec{s}_2\right) \vec{q}_1 \quad (4)$$

which, by design, is orthogonal to \vec{q}_1 since it has no components in the direction of \vec{q}_1 . We have satisfied the orthogonality condition with \vec{z}_2 , so all that is left is to normalize this quantity to find \vec{q}_2 :

$$\vec{q}_2 = \frac{\vec{z}_2}{\|\vec{z}_2\|} \quad (5)$$

Next, we need to show the two spans are equal. First, we can show $\text{Span}(\{\vec{q}_1, \vec{q}_2\}) \subseteq \text{Span}(\{\vec{s}_1, \vec{s}_2\})$. From part 1.b, we already know $\vec{q}_1 \in \text{Span}(\{\vec{s}_1\}) \subseteq \text{Span}(\{\vec{s}_1, \vec{s}_2\})$. We can rewrite \vec{q}_2 as

$$\vec{q}_2 = \alpha \vec{s}_2 + \beta \vec{q}_1 \quad (6)$$

for $\alpha = \frac{1}{\|\vec{z}_2\|}$ and $\beta = \frac{-(\vec{q}_1^T \vec{s}_2)}{\|\vec{z}_2\|}$. We know $\vec{q}_1 = a \vec{s}_1$ for $a = \frac{1}{\|\vec{s}_1\|}$ (from part 1.b), so we can write

$$\vec{q}_2 = \alpha \vec{s}_2 + a\beta \vec{s}_1 \quad (7)$$

so $\vec{q}_2 \in \text{Span}(\{\vec{s}_1, \vec{s}_2\})$.

Next, we can show $\text{Span}(\{\vec{s}_1, \vec{s}_2\}) \subseteq \text{Span}(\{\vec{q}_1, \vec{q}_2\})$. From the 1.b, we know $\vec{s}_1 \in \text{Span}(\{\vec{q}_1\}) \subseteq \text{Span}(\{\vec{q}_1, \vec{q}_2\})$. Now, we can perform algebraic manipulation and rewrite eq. (6) to say

$$\vec{s}_2 = \frac{\vec{q}_2}{\alpha} - \frac{\beta \vec{q}_1}{\alpha} \quad (8)$$

so $\vec{s}_2 \in \text{Span}(\{\vec{q}_1, \vec{q}_2\})$. Hence, we have shown that $\text{Span}(\{\vec{q}_1, \vec{q}_2\}) = \text{Span}(\{\vec{s}_1, \vec{s}_2\})$.

Intuitive Explanation on Projections for Orthogonalization:

The idea behind why we take projections and calculate projection error can be seen as a method to extract \vec{z}_2 from

$$\vec{s}_2 = c_1 \vec{q}_1 + \vec{z}_2 \quad (9)$$

where we choose this decomposition of \vec{s}_2 such that $c_1 \vec{q}_1$ and \vec{z}_2 are orthogonal. That is, we will use the term $c_1 \vec{q}_1$ to represent the component of \vec{s}_2 in the direction of \vec{q}_1 , and \vec{z}_2 to represent the component of \vec{s}_2 that is distinctly *not* in the direction of \vec{q}_1 . We can solve for c_1 using projections. By subtracting this part out as in eq. (4), we are left with a vector \vec{z}_2 that does not have any components in the direction of \vec{q}_1 . Hence, it will be orthogonal to \vec{q}_1 . See fig. 1 for an intuitive plot of what this decomposition could look like.

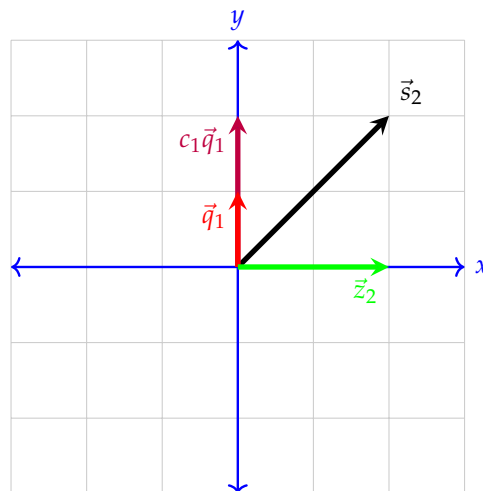


Figure 1: Decomposition of \vec{s}_2

- (d) **What would happen if $\{\vec{s}_1, \vec{s}_2, \vec{s}_3\}$ were *not* linearly independent, but rather \vec{s}_1 were a multiple of \vec{s}_2 ?**

Solution: If \vec{s}_2 is a multiple of \vec{s}_1 , then $\vec{z}_2 = 0$. This means that the projection of \vec{s}_2 onto $\text{Span}(\{\vec{s}_1\})$ is just \vec{s}_2 , so we have found an orthonormal basis for $\text{Span}(\{\vec{s}_1, \vec{s}_2\})$, in particular the basis $\{\vec{q}_1\}$. Hence, we can move onto \vec{s}_3 and continue the algorithm from there.

- (e) Now given \vec{q}_1 and \vec{q}_2 in parts 1.b and 1.c, find \vec{q}_3 such that $\text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}) = \text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\})$, and \vec{q}_3 is orthogonal to both \vec{q}_1 and \vec{q}_2 , and finally $\|\vec{q}_3\| = 1$. You do not have to show that the two spans are equal.

Solution: Based on the intuitive explanation from part 1.c, we would like to write

$$\vec{s}_3 = c_1 \vec{q}_1 + c_2 \vec{q}_2 + \vec{z}_3 \quad (10)$$

where $c_1 \vec{q}_1$ represents the component of \vec{s}_3 that is in the direction of only \vec{q}_1 , $c_2 \vec{q}_2$ represents the component that is in the direction of only \vec{q}_2 , and \vec{z}_3 represents the component that is distinctly *not* in the directions of \vec{q}_1 and \vec{q}_2 . Note that \vec{q}_1 and \vec{q}_2 are in distinctly different directions, since they are orthogonal (this allows us to claim that $c_1 \vec{q}_1$ and $c_2 \vec{q}_2$ represent distinctly different directional components of \vec{s}_3).

We can compute c_1 and c_2 by projections. Namely,

$$c_1 \vec{q}_1 = \text{proj}_{\vec{q}_1}(\vec{s}_3) = \frac{\vec{q}_1^\top \vec{s}_3}{\|\vec{q}_1\|^2} \vec{q}_1 = \underbrace{(\vec{q}_1^\top \vec{s}_3)}_{c_1} \vec{q}_1 \quad (11)$$

$$c_2 \vec{q}_2 = \text{proj}_{\vec{q}_2}(\vec{s}_3) = \frac{\vec{q}_2^\top \vec{s}_3}{\|\vec{q}_2\|^2} \vec{q}_2 = \underbrace{(\vec{q}_2^\top \vec{s}_3)}_{c_2} \vec{q}_2 \quad (12)$$

To find \vec{z}_3 , we can subtract out $c_1 \vec{q}_1$ and $c_2 \vec{q}_2$, namely:

$$\vec{z}_3 = \vec{s}_3 - (\vec{q}_1^\top \vec{s}_3) \vec{q}_1 - (\vec{q}_2^\top \vec{s}_3) \vec{q}_2 \quad (13)$$

All that is left is to normalize this quantity, that is

$$\vec{q}_3 = \frac{\vec{z}_3}{\|\vec{z}_3\|} \quad (14)$$

- (f) **(PRACTICE) Confirm that** $\text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}) = \text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\})$.

Solution: We already showed that $\vec{q}_1, \vec{q}_2 \in \text{Span}(\{\vec{s}_1, \vec{s}_2\}) \subseteq \text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\})$ and also $\vec{s}_1, \vec{s}_2 \in \text{Span}(\{\vec{q}_1, \vec{q}_2\}) \subseteq \text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\})$. It remains to show that $\vec{q}_3 \in \text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\})$ (so we can show $\text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}) \subseteq \text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\})$) and that $\vec{s}_3 \in \text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\})$ (so we can show $\text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\}) \subseteq \text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\})$).

To show $\vec{q}_3 \in \text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\})$:

$$\vec{q}_3 = \frac{\vec{z}_3}{\|\vec{z}_3\|} = \underbrace{\gamma}_{\frac{1}{\|\vec{z}_3\|}} \left(\vec{s}_3 - (\vec{s}_3^\top \vec{q}_1) \vec{q}_1 - (\vec{s}_3^\top \vec{q}_2) \vec{q}_2 \right) \quad (15)$$

$$= \gamma \left(\vec{s}_3 - (\vec{s}_3^\top \vec{q}_1) \underbrace{\vec{q}_1}_{a \vec{s}_1} - (\vec{s}_3^\top \vec{q}_2) \underbrace{\vec{q}_2}_{\alpha \vec{s}_2 + a \beta \vec{s}_1} \right) \quad (16)$$

$$= \gamma \vec{s}_3 + \left(-\alpha (\vec{s}_3^\top \vec{q}_2) \right) \vec{s}_2 + \left(-a (\vec{s}_3^\top \vec{q}_1) - a \beta (\vec{s}_3^\top \vec{q}_2) \right) \vec{s}_1 \quad (17)$$

where $a = \frac{1}{\|\vec{s}_1\|}$, $\alpha = \frac{1}{\|\vec{s}_2\|}$, and $\beta = \frac{-(\vec{q}_1^\top \vec{s}_2)}{\|\vec{s}_2\|}$ (taken from eq. (7)). So, $\vec{q}_3 \in \text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\})$. Now, to show $\vec{s}_3 \in \text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\})$, we can perform algebraic manipulation on eq. (16):

$$\vec{s}_3 = \frac{1}{\gamma} \left(\vec{q}_3 + (\vec{s}_3^\top \vec{q}_1) \vec{q}_1 + (\vec{s}_3^\top \vec{q}_2) \vec{q}_2 \right) \quad (18)$$

so $\vec{s}_3 \in \text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\})$. Hence, we conclude that $\text{Span}(\{\vec{s}_1, \vec{s}_2, \vec{s}_3\}) = \text{Span}(\{\vec{q}_1, \vec{q}_2, \vec{q}_3\})$.

2. Orthonormal Matrices and Projections

A matrix A has orthonormal columns, \vec{a}_i , if they are:

- Orthogonal (ie. $\langle \vec{a}_i, \vec{a}_j \rangle = \vec{a}_j^\top \vec{a}_i = 0$ when $i \neq j$)
 - Normalized (ie. vectors with length equal to 1, $\|\vec{a}_i\| = 1$). This implies that $\|\vec{a}_i\|_2 = \langle \vec{a}_i, \vec{a}_i \rangle = \vec{a}_i^\top \vec{a}_i = 1$.
- (a) When $A \in \mathbb{R}^{n \times m}$ and $n \geq m$ (i.e. for tall matrices), **show that if the matrix is orthonormal, then $A^\top A = I_{m \times m}$.**

Solution: We want to show $A^\top A = I_{m \times m}$. We proceed directly from the definition of matrix multiplication, using that the columns of A are indexed by \vec{a}_i :

$$A^\top A = \begin{bmatrix} - & \vec{a}_1^\top & - \\ - & \vec{a}_2^\top & - \\ & \vdots & \\ - & \vec{a}_m^\top & - \end{bmatrix} \begin{bmatrix} | & | & \cdots & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_m \\ | & | & \cdots & | \end{bmatrix} \quad (19)$$

$$= \begin{bmatrix} \vec{a}_1^\top \vec{a}_1 & \vec{a}_1^\top \vec{a}_2 & \cdots & \vec{a}_1^\top \vec{a}_m \\ \vec{a}_2^\top \vec{a}_1 & \vec{a}_2^\top \vec{a}_2 & \cdots & \vec{a}_2^\top \vec{a}_m \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_m^\top \vec{a}_1 & \vec{a}_m^\top \vec{a}_2 & \cdots & \vec{a}_m^\top \vec{a}_m \end{bmatrix} \quad (20)$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (21)$$

$$= I_{m \times m} \quad (22)$$

When $\vec{a}_i^\top \vec{a}_i = \|\vec{a}_i\|^2 = 1$ and when $i \neq j$, $\vec{a}_i^\top \vec{a}_j = 0$ because the column vectors are orthogonal.

- (b) Again, suppose $A \in \mathbb{R}^{n \times m}$ where $n \geq m$ is an orthonormal matrix. **Show that the projection of \vec{y} onto the subspace spanned by the columns of A is now $AA^\top \vec{y}$.**

Solution: Recall from 16A, that in order to project onto the column space of a matrix we use the least squares formula. By applying this result, we have that

$$\text{proj}_{\text{Col}(A)}(\vec{y}) = A\hat{x} = A(A^\top A)^{-1}A^\top \vec{y} \quad (23)$$

Plugging in the result from part 2.a,

$$\text{proj}_{\text{Col}(A)}(\vec{y}) = A \left(\underbrace{A^\top A}_{I_{m \times m}} \right)^{-1} A^\top \vec{y} \quad (24)$$

$$= AA^\top \vec{y} \quad (25)$$

- (c) **(PRACTICE) Show if $A \in \mathbb{R}^{n \times n}$ is an orthonormal matrix then the columns, \vec{a}_i , form a basis for \mathbb{R}^n .**

Solution: Recall that, if we would like to show that a set of vectors are linearly independent, then the only β_i 's satisfying

$$\beta_1 \vec{a}_1 + \beta_2 \vec{a}_2 + \cdots + \beta_n \vec{a}_n = \vec{0} \quad (26)$$

would be $\beta_i = 0$ for $i = 1$ to $i = n$. To show that $\beta_i = 0$ for the given instance, we can left multiply eq. (26) by \vec{a}_i^\top (for any $i = 1$ to $i = n$):

$$\vec{a}_i^\top (\beta_1 \vec{a}_1 + \beta_2 \vec{a}_2 + \dots + \beta_n \vec{a}_n) = \vec{a}_i^\top \vec{0} \quad (27)$$

$$\sum_{j=1}^n \beta_j \vec{a}_i^\top \vec{a}_j = 0 \quad (28)$$

$$\beta_i \underbrace{\vec{a}_i^\top \vec{a}_i}_1 = 0 \quad (29)$$

$$\implies \beta_i = 0 \quad (30)$$

where we get to eq. (29) by using the fact that $\vec{a}_i^\top \vec{a}_j = 0$ for $i \neq j$. Hence, $\beta_i = 0$ for $i = 1$ to $i = n$.

Contributors:

- Anish Muthali.
- Regina Eckert.
- Druv Pai.
- Neelesh Ramachandran.