Lecture 32: More on Orthogonality

More on Projections

Last time we discussed how to carry out the orthogonal projection to a line L. Now we want to look at projections (in $\mathbb{R}^n$) to larger-dimensional subspaces W.

Let $\{\vec{v}_1, \dots, \vec{v}_n\} \subset \mathbb{R}^n$ be an orthogonal basis. We have (for any $\vec{y} \in \mathbb{R}^n$)

$$\vec{y} = \frac{\vec{y}\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1}\vec{v}_1 + \cdots + \frac{\vec{y}\cdot\vec{v}_n}{\vec{v}_n\cdot\vec{v}_n}\vec{v}_n$$

by Lecture 31. Suppose we want to project to $W = \text{span}\{\vec{v}_1, \vec{v}_2\}$.

<u>Claim</u>: $z := \frac{\vec{y}\cdot\vec{v}_3}{\vec{v}_3\cdot\vec{v}_3}\vec{v}_3 + \cdots + \frac{\vec{y}\cdot\vec{v}_n}{\vec{v}_n\cdot\vec{v}_n}\vec{v}_n$ belongs to $W^\perp$.

If this is true, then setting $\hat{y} := \frac{\vec{y}\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1}\vec{v}_1 + \frac{\vec{y}\cdot\vec{v}_2}{\vec{v}_2\cdot\vec{v}_2}\vec{v}_2$

we have

$$\vec{y} = \hat{y} + \vec{z} \qquad (\hat{y} \in W,\ \vec{z} \in W^\perp)$$

and define

$$\text{proj}_W \vec{y} := \hat{y}.$$

<u>Proof of Claim</u>: It is enough to check that $\vec{z}$ is $\perp$ to a basis of W, i.e. $\vec{v}_1$ and $\vec{v}_2$. Since the $\vec{v}_i$ are orthogonal, this is obvious. □

In fact — and this is the key point — we don't need to know the rest of the $\vec{v}_i$ at all: just $\vec{v}_1$ & $\vec{v}_2$ will do. (This anticipates knowing that $\vec{v}_1$ & $\vec{v}_2$ can necessarily be extended to an orthogonal basis, which comes courtesy of Gram-Schmidt below.)

Ex 1/ Project $\vec{y} = \begin{pmatrix}1\\3\\5\end{pmatrix}$ onto $W = \text{span}\left\{\overset{\vec{v}_1}{\begin{pmatrix}1\\3\\-2\end{pmatrix}}, \overset{\vec{v}_2}{\begin{pmatrix}5\\1\\4\end{pmatrix}}\right\} \subset \mathbb{R}^3$.

First check $\vec{v}_1 \cdot \vec{v}_2 = 0$. Now

$$\text{proj}_W \vec{y} = \frac{\begin{pmatrix}1\\3\\5\end{pmatrix}\cdot\begin{pmatrix}1\\3\\-2\end{pmatrix}}{\begin{pmatrix}1\\3\\-2\end{pmatrix}\cdot\begin{pmatrix}1\\3\\-2\end{pmatrix}}\begin{pmatrix}1\\3\\-2\end{pmatrix} + \frac{\begin{pmatrix}1\\3\\5\end{pmatrix}\cdot\begin{pmatrix}5\\1\\4\end{pmatrix}}{\begin{pmatrix}5\\1\\4\end{pmatrix}\cdot\begin{pmatrix}5\\1\\4\end{pmatrix}}\begin{pmatrix}5\\1\\4\end{pmatrix} = 0\begin{pmatrix}1\\3\\-2\end{pmatrix} + \frac{28}{42}\begin{pmatrix}5\\1\\4\end{pmatrix} = \begin{pmatrix}10/3\\2/3\\8/3\end{pmatrix}. \quad //$$

Properties:

⓪ The definition of $\hat{y}$ is "independent of choices"

What we mean here is that if $\hat{y}' + \vec{z}\,' = \vec{y} = \hat{y} + \vec{z}$ with $\hat{y}'$ and $\hat{y} \in W$, and $\vec{z}\,', \vec{z} \in W^\perp$, then $\hat{y}' = \hat{y}$ (and $\vec{z}\,' = \vec{z}$). Why? $\hat{y}' + \vec{z}\,' = \hat{y} + \vec{z} \Rightarrow \underbrace{\hat{y}' - \hat{y}}_{\text{in } W} = \underbrace{\vec{z} - \vec{z}\,'}_{\text{in } W^\perp}$

$\Rightarrow \hat{y}' - \hat{y} \in W \cap W^\perp \Rightarrow (\hat{y}' - \hat{y})\cdot(\hat{y}' - \hat{y}) = 0$

$\Rightarrow \hat{y}' - \hat{y} = \vec{0} \Rightarrow \hat{y}' = \hat{y}$.

① If $\vec{y} \in W$, then $\hat{y} = \vec{y}$. Clear since $\vec{y} = \vec{y} + \vec{0}$ is a decomposition into "W" & "$W^\perp$" pieces, and hence (by ⓪) the only one.

(2) $\hat{y}$ is the closest point to $\vec{y}$ in W. If $\vec{w} \in W$ is arbitrary, then $\vec{y} - \vec{w} = \underbrace{(\vec{y} - \hat{y})}_{\text{in } W^\perp} + \underbrace{(\hat{y} - \vec{w})}_{\text{in } W}$ $\xRightarrow{\text{Pythagorean theorem}}$

$\|\vec{y} - \vec{w}\|^2 = \|\vec{y} - \hat{y}\|^2 + \|\hat{y} - \vec{w}\|^2 \geq \|\vec{y} - \hat{y}\|^2$, with equality if & only if $\vec{w} = \hat{y}$. Write

$$\text{dist}(\vec{y}, W) := \|\vec{y} - \hat{y}\| = \|\vec{z}\|.$$

Ex 2/ Find the closest point to $\vec{y}$ in $W = \text{span}\left\{ \overset{\vec{v}_1}{\begin{pmatrix} 3 \\ 1 \\ -1 \\ 1 \end{pmatrix}}, \overset{\vec{v}_2}{\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}} \right\}$ if $\vec{y} = \begin{pmatrix} 3 \\ 1 \\ 5 \\ 1 \end{pmatrix}$. Compute $\text{dist}(\vec{y}, W)$.

Again, it's essential to check that $\vec{v}_1 \cdot \vec{v}_2$. (Otherwise the formula won't work.) Then the closest point is simply

$$\hat{y} = \frac{\vec{y} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 + \frac{\vec{y} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 = \frac{1}{2} \vec{v}_1 + \frac{3}{2} \vec{v}_2 = \begin{pmatrix} 3 \\ -1 \\ 1 \\ -1 \end{pmatrix},$$

and $\text{dist}(\vec{y}, W) = \left\| \begin{pmatrix} 3 \\ 1 \\ 5 \\ 1 \end{pmatrix} - \begin{pmatrix} 3 \\ -1 \\ 1 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 0 \\ 2 \\ 4 \\ 2 \end{pmatrix} \right\| = \sqrt{2^2 + 4^2 + 2^2} = 2\sqrt{6}$. //

Now suppose $\{\vec{u}_1, \vec{u}_2\}$ was an ortho<u>normal</u> basis of W.

Then

$$\hat{y} = (\vec{y} \cdot \vec{u}_1)\vec{u}_1 + (\vec{y} \cdot \vec{u}_2)\vec{u}_2,$$

and there is an even <u>nicer</u> way to write this: let U denote the $n \times 2$ matrix $\begin{pmatrix} \uparrow & \uparrow \\ \vec{u}_1 & \vec{u}_2 \\ \downarrow & \downarrow \end{pmatrix}$. Then

$$U^T \vec{y} = \begin{pmatrix} \leftarrow \vec{u}_1^T \rightarrow \\ \leftarrow \vec{u}_2^T \rightarrow \end{pmatrix} \vec{y} = \begin{pmatrix} \vec{u}_1 \cdot \vec{y} \\ \vec{u}_2 \cdot \vec{y} \end{pmatrix}$$

$$\phantom{U^T \vec{y} = } \quad 2 \times n \qquad n \times 1 \qquad 2 \times 1$$

and

$$\underbrace{\overset{n\times 2}{U}\,\overset{2\times n}{U^T}}_{n \times n} \vec{y} = \underset{n\times 2}{\begin{pmatrix} \uparrow & \uparrow \\ \vec{u}_1 & \vec{u}_2 \\ \downarrow & \downarrow \end{pmatrix}} \underset{2\times 1}{\begin{pmatrix} \vec{u}_1 \cdot \vec{y} \\ \vec{u}_2 \cdot \vec{y} \end{pmatrix}} = (\vec{u}_1 \cdot \vec{y})\vec{u}_1 + (\vec{u}_2 \cdot \vec{y})\vec{u}_2$$

$$= \hat{y} = \text{proj}_W(\vec{y}),$$

thus arriving at the

Theorem: UU^T is the matrix of proj_W.

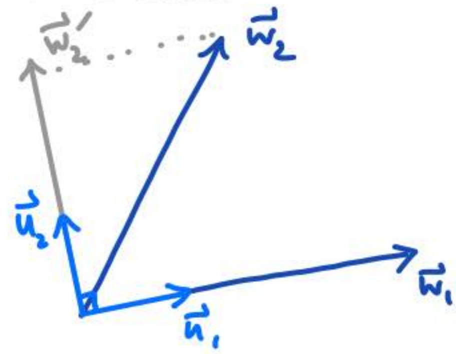
None of the above has anything to do with W being 2-dimensional. Instead of "$\vec{v}_1, \vec{v}_2$" (or "$\vec{u}_1, \vec{u}_2$"), everything works with 1 vector, or 3 vectors, or k vectors. The only difference is that U will be an $n \times k$ matrix; but once more UU^T is $(n\times k)(k \times n) = n \times n$.

Gram-Schmidt orthogonalization Let $W \subset \mathbb{R}^n$, $\vec{y} \in \mathbb{R}^n$.

Since the formula for $\hat{y}$ only applies in case we have an orthogonal basis for W, we need a way of turning an arbitrary basis $\{\vec{w}_1, \ldots, \vec{w}_k\}$ into an orthogonal one. In fact, the following method turns it into an <u>orthonormal</u> basis:

begin by normalizing $\vec{w}_1$. Set

$$\vec{u}_1 := \frac{\vec{w}_1}{\|\vec{w}_1\|}.$$



Referring to the picture, we'd like to make $\vec{w}_2 \perp \vec{u}_1$ by getting rid of its "horizontal" component, $(\vec{w}_2 \cdot \vec{u}_1)\vec{u}_1$. So put

$$\vec{w}_2' := \vec{w}_2 - (\vec{w}_2 \cdot \vec{u}_1)\vec{u}_1$$

and normalize to

$$\vec{u}_2 := \frac{\vec{w}_2'}{\|\vec{w}_2'\|}.$$

Note that $\vec{w}_2' \cdot \vec{u}_1 = \vec{w}_2 \cdot \vec{u}_1 - (\vec{w}_2 \cdot \vec{u}_1)\underbrace{\vec{u}_1 \cdot \vec{u}_1}_{1} = 0$

Next, to make $\vec{w}_3 \perp$ to both $\vec{u}_1$ & $\vec{u}_2$, we take

$$\vec{w}_3' := \vec{w}_3 - (\vec{w}_3 \cdot \vec{u}_1)\vec{u}_1 - (\vec{w}_3 \cdot \vec{u}_2)\vec{u}_2 \quad \text{and} \quad \vec{u}_3 := \frac{\vec{w}_3'}{\|\vec{w}_3'\|}.$$

Continuing like this, we eventually get an orthonormal basis $\{\vec{u}_1, \ldots, \vec{u}_k\}$ of W.

We can "invert" the equations defining the $\{\vec{u}_i\}$ as follows:

$$\vec{w}_1 = \|\vec{w}_1\|\,\vec{u}_1$$
$$\vec{w}_2 = (\vec{w}_2\cdot\vec{u}_1)\,\vec{u}_1 + \|\vec{w}_2{}'\|\,\vec{u}_2$$
$$\vec{w}_3 = (\vec{w}_3\cdot\vec{u}_1)\,\vec{u}_1 + (\vec{w}_3\cdot\vec{u}_2)\,\vec{u}_2 + \|\vec{w}_3{}'\|\,\vec{u}_3$$

etc.,

which looks especially nice in matrix form:

$$\underbrace{\begin{pmatrix} \uparrow & & \uparrow \\ \vec{w}_1 & \cdots & \vec{w}_k \\ \downarrow & & \downarrow \end{pmatrix}}_{A \ \ n\times k} = \underbrace{\begin{pmatrix} \uparrow & & \uparrow \\ \vec{u}_1 & \cdots & \vec{u}_k \\ \downarrow & & \downarrow \end{pmatrix}}_{Q \ \ n\times k} \underbrace{\begin{pmatrix} \|\vec{w}_1\| & \vec{w}_2\cdot\vec{u}_1 & \vec{w}_3\cdot\vec{u}_1 & \cdots & \vec{w}_k\cdot\vec{u}_1 \\ & \|\vec{w}_2{}'\| & \vec{w}_3\cdot\vec{u}_2 & \cdots & \vec{w}_k\cdot\vec{u}_2 \\ & & \|\vec{w}_3{}'\| & & \vdots \\ & 0 & & \ddots & \\ & & & & \|\vec{w}_k{}'\| \end{pmatrix}}_{R \ \ k\times k}$$

A: columns = original basis

Q: columns = o.n. basis

R: upper triangular with positive diagonal entries

If $k=n$, i.e. $W = (\text{all of})\ \mathbb{R}^n$, then this is called the "Iwasawa decomposition" — it takes any invertible $n\times n$ matrix and writes it (uniquely) as the product of an orthogonal matrix and an upper triangular matrix.

Your book breaks the "Gram–Schmidt" algorithm into 2 pieces: first, it produces a merely orthogonal basis (this avoids square roots, and is suitable for some purposes), **then** it normalizes the vectors to get orthonormal ones: Starting again with $\{\vec{w}_1, \dots, \vec{w}_k\}$, write

$$\vec{v}_1 = \vec{w}_1$$
$$\vec{v}_2 = \vec{w}_2 - \frac{\vec{w}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1$$
$$\vec{v}_3 = \vec{w}_3 - \frac{\vec{w}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \frac{\vec{w}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2$$

etc.

for the orthogonal basis $\{\vec{v}_1, \dots, \vec{v}_k\}$ of W; then set

$$\vec{u}_j = \frac{\vec{v}_j}{\|\vec{v}_j\|} \quad \text{for each } j,$$

if an o.n. basis is needed.

Ex 3/ Find an o.n. basis for $W := \operatorname{span}\left\{ \underset{\vec{w}_1}{\begin{pmatrix} 3 \\ 1 \\ -1 \\ 3 \end{pmatrix}}, \underset{\vec{w}_2}{\begin{pmatrix} -5 \\ 1 \\ 5 \\ -7 \end{pmatrix}}, \underset{\vec{w}_3}{\begin{pmatrix} 1 \\ 1 \\ -2 \\ 8 \end{pmatrix}} \right\} \subset \mathbb{R}^4$.

$$\vec{v}_1 = \begin{pmatrix} 3 \\ 1 \\ -1 \\ 3 \end{pmatrix},$$

$$\vec{v}_2 = \begin{pmatrix} -5 \\ 1 \\ 5 \\ -7 \end{pmatrix} - \underbrace{\frac{\vec{w}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}}_{-40/20} \begin{pmatrix} 3 \\ 1 \\ -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 3 \\ -1 \end{pmatrix},$$

$$\vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \\ 8 \end{pmatrix} - \underbrace{\frac{\vec{w}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}}_{30/20} \begin{pmatrix} 3 \\ 1 \\ -1 \\ 3 \end{pmatrix} - \underbrace{\frac{\vec{w}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}}_{-10/20} \begin{pmatrix} 1 \\ 3 \\ 3 \\ -1 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \\ 1 \\ 3 \end{pmatrix}.$$

(Check they're $\perp$!)

That's the orthogonal basis. For orthonormal, put

$$\vec{u}_j = \frac{\vec{v}_j}{\|\vec{v}_j\|} = \frac{1}{\sqrt{20}} \vec{v}_j \quad \text{for each } j.$$

So the QR-factorization here would look like

$$\underset{4\times 3}{\begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{w}_1 & \vec{w}_2 & \vec{w}_3 \\ \downarrow & \downarrow & \downarrow \end{pmatrix}} = \frac{1}{\sqrt{20}} \underset{4\times 3}{\begin{pmatrix} \uparrow & \uparrow & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \\ \downarrow & \downarrow & \downarrow \end{pmatrix}} \underset{3\times 3}{\begin{pmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & \blacksquare \end{pmatrix}}$$