Orthogonal Decomposition Theorem

W a subspace of $\mathbb{R}^n$. Each $\boldsymbol{y} \in \mathbb{R}^n$ can be written in a unique way as

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z} \text{ where } \begin{cases} \widehat{\boldsymbol{y}} & \text{is in } W \\ \boldsymbol{z} & \text{is in } W^{\perp} \end{cases}$$

If $W = \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_p}\}$ is an orthogonal basis for W, then formula are

$$\begin{aligned} \widehat{\boldsymbol{y}} &= \tfrac{\boldsymbol{y}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} + \ldots + \tfrac{\boldsymbol{y}\cdot\boldsymbol{v_i}}{\boldsymbol{v_i}\cdot\boldsymbol{v_i}}\boldsymbol{v_i} + \ldots + \tfrac{\boldsymbol{y}\cdot\boldsymbol{v_p}}{\boldsymbol{v_p}\cdot\boldsymbol{v_p}}\boldsymbol{v_p} \\ &\quad \updownarrow \qquad\qquad \updownarrow \qquad\qquad \updownarrow \\ &= \text{proj}_{\boldsymbol{v_1}}\boldsymbol{y} + \ldots + \text{proj}_{\boldsymbol{v_i}}\boldsymbol{y} + \ldots + \text{proj}_{\boldsymbol{v_p}}\boldsymbol{y} \end{aligned}$$

$$= \qquad \text{proj}_W \boldsymbol{y}$$

and

$$\boldsymbol{z} \quad = \quad \boldsymbol{y} - \widehat{\boldsymbol{y}}$$

$\widehat{\boldsymbol{y}}$ is called the orthogonal projection of y

on W : $\widehat{\boldsymbol{y}} = \text{proj}_W \boldsymbol{y}$

$\boldsymbol{z}$ is called the component of $\boldsymbol{y}$ orthogonal to W

Example

Let $W = \text{Span}\{\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}\} \subseteq \mathbb{R}^4$.

$$\boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \notin W$$

Find the point in W closest to $\boldsymbol{y}$ and find the distance from $\boldsymbol{y}$ to W.

$$\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_2}{\boldsymbol{u}_2\cdot\boldsymbol{u}_2}\boldsymbol{u}_2 + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_3}{\boldsymbol{u}_3\cdot\boldsymbol{u}_3}\boldsymbol{u}_3$$

$$= \quad .. \quad + \quad \frac{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}2\\0\\1\\1\end{bmatrix}}{\begin{bmatrix}2\\0\\1\\1\end{bmatrix}\begin{bmatrix}2\\0\\1\\1\end{bmatrix}}\begin{bmatrix}2\\0\\1\\1\end{bmatrix} + \ldots \quad = \quad \begin{bmatrix}\frac{4}{3}\\1\\\frac{2}{3}\\\frac{2}{3}\end{bmatrix}$$

In general, is there a matrix A such that $A\boldsymbol{y} = \widehat{\boldsymbol{y}}$ (is "projection onto a subspace W" represented by some matrix multiplication? Yes.

Suppose $\mathcal{B} = \{\boldsymbol{u}_1, ..., \boldsymbol{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$.

Let U be the $n \times p$ matrix $U = [\boldsymbol{u}_1\ \boldsymbol{u}_2\ ...\ \boldsymbol{u}_p]$

Then UU^T is $n \times n$, and $UU^T\boldsymbol{y} = \widehat{\boldsymbol{y}} = \text{proj}_W \boldsymbol{y}$

$\boldsymbol{y} \mapsto UU^T\boldsymbol{y} = \widehat{\boldsymbol{y}}$ is a transformation from $\mathbb{R}^n$ into $\mathbb{R}^n$;
Since $\widehat{\boldsymbol{y}}$ is a linear combination of the columns of U, the range of the transformation is $\text{col}\, U$, a subspace of $\mathbb{R}^n$.

Looked at "from the other end":

If U is any $n \times p$ and has orthonormal columns, then

UU^T is $n \times n$, so we have $\mathbb{R}^n \to \mathbb{R}^n$
$\boldsymbol{y} \mapsto UU^T\boldsymbol{y}$

$UU^T\boldsymbol{y} = \widehat{\boldsymbol{y}} = \text{proj}_W \boldsymbol{y}$, where $W = \text{col}\, U$

Example $W = \text{Span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^4.$

Normalize the vectors in $\mathcal{B}$ to get orthonormal basis for W :

$$\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2/\sqrt{6} \\ 0 \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$$

$$U = \begin{bmatrix} 0 & 2/\sqrt{6} & 0 \\ 1 & 0 & 0 \\ 0 & 1/\sqrt{6} & -1/\sqrt{2} \\ 0 & 1/\sqrt{6} & 1/\sqrt{2} \end{bmatrix}_{4\times 3} \quad \text{and} \quad U^T = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 2/\sqrt{6} & 0 & 1/\sqrt{6} & 1/\sqrt{6} \\ 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}_{3\times 4}$$

so $$UU^T = \begin{bmatrix} \frac{2}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} & \frac{2}{3} \end{bmatrix}_{4\times 4}$$

$T(\boldsymbol{y}) = UU^T\boldsymbol{y} \qquad (T: \mathbb{R}^4 \to \mathbb{R}^4)$

$\uparrow$

$=$ orthogonal projection of $\mathbb{R}^4$ onto $W = \text{Col}\, U \subset \mathbb{R}^4$

$$UU^T\boldsymbol{y} = \widehat{\boldsymbol{y}}$$

For example (using same $\boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ as in preceding example)

$$UU^T\boldsymbol{y} = UU^T \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} \\ 1 \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} = \widehat{\boldsymbol{y}}$$

.

Recall: If $W = \{\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}\}$ is an orthogonal basis for W, then formulas for projection of $\boldsymbol{y}$ and component of $\boldsymbol{y}$ orthogonal to W are

$$\begin{aligned}
\widehat{\boldsymbol{y}} &= \tfrac{\boldsymbol{y}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} + \ldots + \tfrac{\boldsymbol{y}\cdot\boldsymbol{v_i}}{\boldsymbol{v_i}\cdot\boldsymbol{v_i}}\boldsymbol{v_i} + \ldots + \tfrac{\boldsymbol{y}\cdot\boldsymbol{v_p}}{\boldsymbol{v_p}\cdot\boldsymbol{v_p}}\boldsymbol{v_p} \\
&\quad \updownarrow \qquad\qquad \updownarrow \qquad\qquad\qquad \updownarrow \\
&= \text{proj}_{\boldsymbol{v_1}}\boldsymbol{y} + \ldots + \text{proj}_{\boldsymbol{v_i}}\boldsymbol{y} + \ldots + \text{proj}_{\boldsymbol{v_p}}\boldsymbol{y} \\
&= \qquad\qquad \text{proj}_{W}\boldsymbol{y}
\end{aligned}$$

and

$$\boldsymbol{z} \quad = \quad \boldsymbol{y} - \widehat{\boldsymbol{y}}$$

Gram Schmidt Formulas

Suppose $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}\}$ is a basis for the subspace W of $\mathbb{R}^n$. Then we can convert to an orthogonal basis $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_p}\}$ for W using the formulas:

$$\begin{aligned}
\boldsymbol{v_1} &= \boldsymbol{x_1} \\
\boldsymbol{v_2} &= \boldsymbol{x_2} - \text{proj}_{W_1}\boldsymbol{x_2} && (W_1 = \text{span}\{\boldsymbol{v_1}\}) \\
&= \boldsymbol{x_2} - \text{proj}_{\boldsymbol{v_1}}\boldsymbol{x_2} = \boldsymbol{x_2} - \tfrac{\boldsymbol{x_2}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} \\
\boldsymbol{v_3} &= \boldsymbol{x_3} - \text{proj}_{W_2}\boldsymbol{x_3} && (W_2 = \text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}) \\
&= \boldsymbol{x_3} - (\text{proj}_{\boldsymbol{v_1}}\boldsymbol{x_3} + \text{proj}_{\boldsymbol{v_2}}\boldsymbol{x_3}) = \boldsymbol{x_3} - \tfrac{\boldsymbol{x_3}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} - \tfrac{\boldsymbol{x_3}\cdot\boldsymbol{v_2}}{\boldsymbol{v_2}\cdot\boldsymbol{v_2}}\boldsymbol{v_2} \\
&\quad \vdots \\
\boldsymbol{v_p} &= \boldsymbol{x_p} - \text{proj}_{W_{p-1}}\{\boldsymbol{x_p}\} && (W_{p-1} = \text{Span}\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_{p-1}}\}) \\
&= \boldsymbol{x_p} - (\text{proj}_{\boldsymbol{v_1}}\boldsymbol{x_p} + \text{proj}_{\boldsymbol{v_2}}\boldsymbol{x_p} + \ldots + \text{proj}_{\boldsymbol{v_{p-1}}}\boldsymbol{x_p}) \\
&= \boldsymbol{x_p} - \tfrac{\boldsymbol{x_p}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} - \tfrac{\boldsymbol{x_p}\cdot\boldsymbol{v_2}}{\boldsymbol{v_2}\cdot\boldsymbol{v_2}}\boldsymbol{v_2} - \ldots - \tfrac{\boldsymbol{x_p}\cdot\boldsymbol{v_{p-1}}}{\boldsymbol{v_{p-1}}\cdot\boldsymbol{v_{p-1}}}\boldsymbol{v_{p-1}}
\end{aligned}$$

Note that each $\boldsymbol{v_{i+1}}$, is just the component of $\boldsymbol{x_{i+1}}$ that is orthogonal to $W_i = \text{Span}\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_i}\}$ = the span of the earlier vectors $\boldsymbol{v_1}, \ldots, \boldsymbol{v_i}$.

$$W = \text{Span}\left\{\begin{bmatrix}0\\2\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\\1\end{bmatrix}\right\} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}\}$$

You can check that $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}$ are linearly independent and therefore form a basis for W.

Find an orthogonal basis for W using Gram-Schmidt Process

1. Let $\boldsymbol{v_1} = \boldsymbol{x_1} = \begin{bmatrix}0\\2\\0\\1\end{bmatrix}$. $W_1 = \text{Span}\{\boldsymbol{x_1}\} = \text{Span}\{\boldsymbol{v_1}\} = \text{Span}(\begin{bmatrix}0\\2\\0\\1\end{bmatrix})$

2. Replace $\boldsymbol{x_2}$ with $\boldsymbol{v_2}$ = the component of $\boldsymbol{x_2}$ orthogonal to $W_1 = \text{span}\{\boldsymbol{v_1}\}$

$$= \boldsymbol{x_2} - \text{proj}_{W_1}\boldsymbol{x_2}$$

$$= \boldsymbol{x_2} - \frac{\boldsymbol{x_2}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1}$$

$$\frac{\boldsymbol{x_2}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} = \frac{\begin{bmatrix}1\\0\\0\\1\end{bmatrix}\cdot\begin{bmatrix}0\\2\\0\\1\end{bmatrix}}{\begin{bmatrix}0\\2\\0\\1\end{bmatrix}\cdot\begin{bmatrix}0\\2\\0\\1\end{bmatrix}}\begin{bmatrix}0\\2\\0\\1\end{bmatrix} = \frac{1}{5}\begin{bmatrix}0\\2\\0\\1\end{bmatrix} = \begin{bmatrix}0\\\frac{2}{5}\\0\\\frac{1}{5}\end{bmatrix}, \quad \text{so}$$

$$\boldsymbol{v_2} = \boldsymbol{x_2} - \text{proj}_{W_1}\boldsymbol{x_2} = \begin{bmatrix}1\\0\\0\\1\end{bmatrix} - \begin{bmatrix}0\\\frac{2}{5}\\0\\\frac{1}{5}\end{bmatrix} = \begin{bmatrix}1\\-\frac{2}{5}\\0\\\frac{4}{5}\end{bmatrix}$$

Optional: rescale $\boldsymbol{v_2}$ to avoid working with the fractions in the remaining calculations

$$\text{new } \boldsymbol{v_2} = 5\begin{bmatrix}1\\-\frac{2}{5}\\0\\\frac{4}{5}\end{bmatrix} = \begin{bmatrix}5\\-2\\0\\4\end{bmatrix}$$

Then $W_2 = \text{Span}\{\boldsymbol{x_1}, \boldsymbol{x_2}\} = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{x_2}\}$
$= \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}.$

3. Let $W_3 = \text{Span}\{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}\} = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{x_3}\}$

Replace $\boldsymbol{x_3}$ with $\boldsymbol{v_3}$ $= \text{the component of } \boldsymbol{x_3} \text{ orthogonal to } W_2$
$= \boldsymbol{x_3} - \text{proj}_{W_2}\boldsymbol{x_3}$
$= \boldsymbol{x_3} - \text{proj}_{\boldsymbol{v_1}}\boldsymbol{x_3} - \text{proj}_{\boldsymbol{v_2}}\boldsymbol{x_3}$

$$\text{proj}_{W_2}\boldsymbol{x_3} = \frac{\boldsymbol{x_3}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} + \frac{\boldsymbol{x_3}\cdot\boldsymbol{v_2}}{\boldsymbol{v_2}\cdot\boldsymbol{v_2}}\boldsymbol{v_2}$$

$$= \frac{\begin{bmatrix}1\\0\\1\\1\end{bmatrix}\cdot\begin{bmatrix}0\\2\\0\\1\end{bmatrix}}{\begin{bmatrix}0\\2\\0\\1\end{bmatrix}\cdot\begin{bmatrix}0\\2\\0\\1\end{bmatrix}}\begin{bmatrix}0\\2\\0\\1\end{bmatrix} + \frac{\begin{bmatrix}1\\0\\1\\1\end{bmatrix}\cdot\begin{bmatrix}5\\-2\\0\\4\end{bmatrix}}{\begin{bmatrix}5\\-2\\0\\4\end{bmatrix}\cdot\begin{bmatrix}5\\-2\\0\\4\end{bmatrix}}\begin{bmatrix}5\\-2\\0\\4\end{bmatrix}$$

$$= \quad \ldots = \begin{bmatrix}1\\0\\0\\1\end{bmatrix},$$

so $\boldsymbol{v_3} = \boldsymbol{x_3} - \text{proj}_{W_2}\boldsymbol{x_3} = \begin{bmatrix}1\\0\\1\\1\end{bmatrix} - \begin{bmatrix}1\\0\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$

Then $W_3 = \text{Span}\{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}\} = \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$, and we have replaced all the $\boldsymbol{x_i}$'s. W_3 is the same as the subspace W we were originally given, and an orthogonal basis for W is

$$\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\} = \left\{\begin{bmatrix}0\\2\\0\\1\end{bmatrix}, \begin{bmatrix}5\\-2\\0\\4\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}\right\}$$

Note: each time you create a $\boldsymbol{v_i}$, you should mentally calculate dot products to be sure that $\boldsymbol{v_i}$ is perpendicular to each of the preceding $\boldsymbol{v}$'s. ***If there's any arithmetic mistake, you want to find it before making further calculations.***

What would happen if we applied the Gram-Schmidt Process to a linearly dependent set of (nonzero) vectors $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_p}\}$?