In $\mathbb{R}^n$:

$\boldsymbol{u} \neq 0$ $\quad$ $L = \text{Span}\{\boldsymbol{u}\}$ $\quad$ $\boldsymbol{y}$ in $\mathbb{R}^n$

$$\boldsymbol{y} = \underset{\text{in } L}{\widehat{\boldsymbol{y}}} + \underset{\text{in } L^{\perp}}{\boldsymbol{z}} \quad \text{where } \widehat{\boldsymbol{y}} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$$

$$\widehat{\boldsymbol{y}} = \mathbf{proj}_L \boldsymbol{y} = \mathbf{proj}_{\boldsymbol{u}} \boldsymbol{y}$$

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

A geometric interpretation

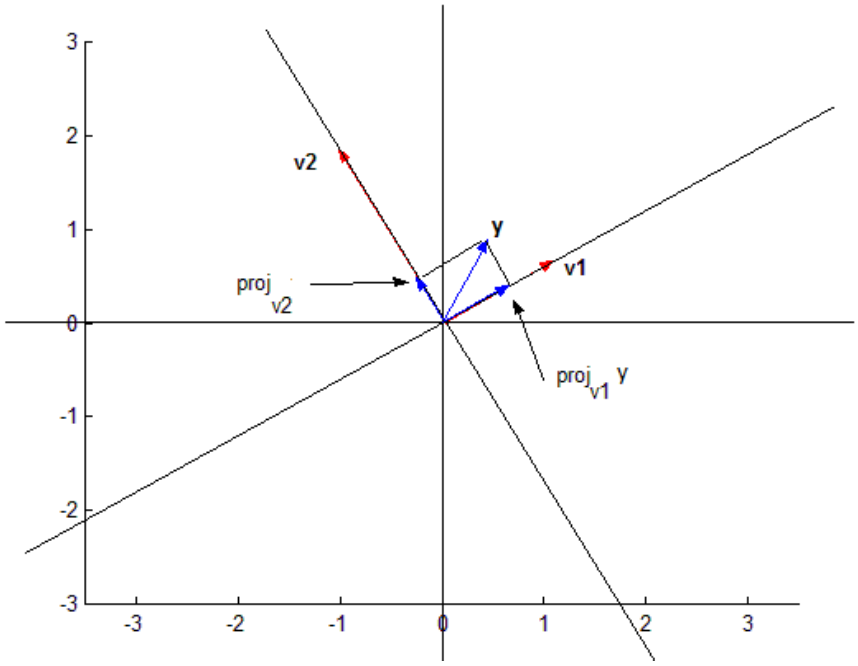
Recall: $\mathcal{B} = \{\boldsymbol{v_1}, ..., \boldsymbol{v_i}, ...\boldsymbol{v_n}\}$ = orthogonal basis for $\mathbb{R}^n$

$$\boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1 + \ldots + \frac{\boldsymbol{y}\cdot\boldsymbol{v}_i}{\boldsymbol{v}_i\cdot\boldsymbol{v}_i}\boldsymbol{v}_i + \ldots + \frac{\boldsymbol{y}\cdot\boldsymbol{v}_n}{\boldsymbol{v}_n\cdot\boldsymbol{v}_n}\boldsymbol{v}_n$$

$$= \text{proj}_{\boldsymbol{v}_1}\boldsymbol{y} + \ldots + \text{proj}_{\boldsymbol{v}_i}\boldsymbol{y} + \ldots + \text{proj}_{\boldsymbol{v}_n}\boldsymbol{y}$$

So (pictured in $\mathbb{R}^2$): for an orthogonal basis $\mathcal{B} = \{\boldsymbol{v_1}, \boldsymbol{v_2}\}$

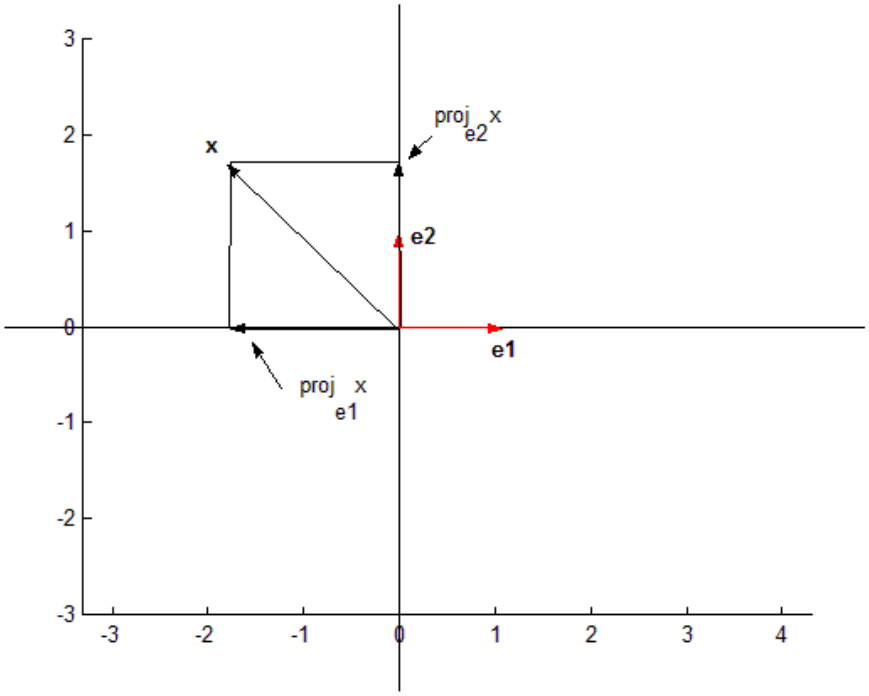
$$\boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1 + \frac{\boldsymbol{xy}\cdot\boldsymbol{v}_2}{\boldsymbol{v}_2\cdot\boldsymbol{v}_2}\boldsymbol{v}_2 = \text{proj}_{\boldsymbol{v}_1}\boldsymbol{y} + \text{proj}_{\boldsymbol{v}_2}\boldsymbol{y}$$



Another (special) illustration in $\mathbb{R}^2$: what if $\mathcal{B} = \{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ = the "standard basis" ?

$$\boldsymbol{x} = (\boldsymbol{x} \cdot \boldsymbol{e_1})\, \boldsymbol{e_1} + \ldots + (\boldsymbol{x} \cdot \boldsymbol{e_2})\, \boldsymbol{e_2}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \boldsymbol{e_1} + x_2 \boldsymbol{e_2}$$



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

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

If $W = \{\boldsymbol{u_1}, \ldots, \boldsymbol{u_p}\}$ is an orthogonal basis for W, then formula are

$$\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u_1} + \ldots + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_p}{\boldsymbol{u}_p\cdot\boldsymbol{u}_p}\boldsymbol{u_p}$$

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

$\widehat{\boldsymbol{y}}$ 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

Best Approximation Theorem

$\widehat{\boldsymbol{y}}$ is the point in W closest to $\boldsymbol{y}$ meaning that if $\boldsymbol{v} \in W$ and $\boldsymbol{v} \neq \widehat{\boldsymbol{y}}$, then

$$||\boldsymbol{y} - \widehat{\boldsymbol{y}}|| < ||\boldsymbol{y} - \boldsymbol{v}||$$

(*so* $\widehat{\boldsymbol{y}}$ *is the vector in* W *that "best approximates"* $\boldsymbol{y}$)

See discussion and proofs in text/class notes.

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.

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

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

so

$\widehat{\boldsymbol{y}} = \frac{1}{1}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \frac{4}{6}\begin{bmatrix} \mathbf{2} \\ \mathbf{0} \\ \mathbf{1} \\ \mathbf{1} \end{bmatrix} + \frac{0}{2}\begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} \\ 1 \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$ = point in W closest to $\boldsymbol{y}$.

Distance from $\boldsymbol{y}$ to $W = ||\boldsymbol{z}|| = ||\boldsymbol{y} - \widehat{\boldsymbol{y}}|| = \| \begin{bmatrix} -\frac{1}{3} \\ 0 \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \|$

$= \sqrt{\frac{1}{9} + \frac{1}{9} + \frac{1}{9}} = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$.

$\widehat{\boldsymbol{y}} = \begin{bmatrix} \frac{4}{3} \\ 1 \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$ = "best approximation" to $\boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ by a point from W.

$||\boldsymbol{y} - \widehat{\boldsymbol{y}}||$ is the size of error when $\widehat{\boldsymbol{y}}$ is used as an approximation for $\boldsymbol{y}$. If $\boldsymbol{w} \in W$ and $\boldsymbol{w} \neq \widehat{\boldsymbol{y}}$, then $||\boldsymbol{y} - \boldsymbol{w}|| > ||\boldsymbol{y} - \widehat{\boldsymbol{y}}|| = \frac{\sqrt{3}}{3}$.