Orthogonal Projections for the $\mathbb{R}^n$ case: Let $\vec{v} \in \mathbb{R}^n$ and $\vec{a} \in \mathbb{R}^n$ and $\vec{e} \in \mathbb{R}^n$. Then

$$\text{proj}_{\vec{v}}(\vec{a}) = \frac{\vec{v}\vec{v}^T}{\|v\|^2}\vec{a} = H\vec{a}$$

which lives in $\mathbb{R}^{n\times n}$. H is called a projection matrix.
Notes:

- $\vec{a} = \text{proj}_{\vec{v}}(\vec{a}) + \vec{e} \to \vec{e} = \vec{a} - \text{proj}_{\vec{v}}(\vec{a})$
- $\vec{e} \cdot \text{proj}_{\vec{v}}(\vec{a}) = 0$; e and the projection are orthogonal
- $\text{proj}_{\vec{v}}(\vec{a}) = c\vec{v} = l\vec{v}_0 \in \text{colsp}\left\{\vec{v}\right\}$ (all linear combinations of $\vec{v}$)
- $\vec{e} \cdot \vec{v} = 0$; $\vec{e}$ and $\vec{v}$ are orthogonal

Now project onto an entire space $\mathcal{V} = \text{colsp}\left\{v\right\}$.

$$V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_k \end{bmatrix} \in \mathbb{R}^{n\times k}$$

Several Observations:

- If $\text{proj}_{\vec{v}}(\vec{a}) \in \text{colsp}\left\{V\right\}$, then $\text{proj}_{\vec{v}}(\vec{a}) = w_1\vec{v}_1 + w_2\vec{v}_2 + \cdots + w_k\vec{v}_k = V\vec{w}$
- $\vec{e} = \vec{a} - V\vec{w}$
- $\vec{e} \cdot \vec{v}_i = 0$ for all $i = 1, \dots, k$
- $\vec{v}_1^T(\vec{a} - V\vec{w}) = 0$, $\vec{v}_2^T(\vec{a} - V\vec{w}) = 0$, $\dots$, $\vec{v}_k^T(\vec{a} - V\vec{w}) = 0$

Therefore

$$\begin{aligned}
\vec{V}^T(\vec{a} - V\vec{w}) &= \vec{0}_n \\
V^T\vec{a} - V^TV\vec{w} &= \vec{0}_n \\
\vec{V}^T\vec{a} &= V^TV\vec{w} \\
(V^TV)^{-1}V^T\vec{a} &= \vec{w} \\
\underbrace{V(V^TV)^{-1}}_{H}\vec{a} &= \text{proj}_{\vec{v}}(\vec{a})
\end{aligned}$$

This means that

$$\hat{y} = H\vec{y} = X(X^TX)^{-1}X^T\vec{y} = H\vec{y}$$

where H is a projection matrix.

Properties:

- Symmetric

$$H^T = (V(V^TV)^{-1}V^T)^T = V((V^TV)^{-1})^T(V^T)^T = V(V^TV)^{-1}V^T = H$$

- Idempotent:

$$HH = (V(V^TV)^{-1}V^T)(V(V^TV)^{-1}V^T) = H$$

This means that the solution is the same as orthogonally projecting $\vec{y}$ onto the column space of X.

Residuals:

$$\vec{e} = \vec{y} - \vec{\hat{y}} = I\vec{y} - H\vec{y} = (I - H)\vec{y}$$

Note that

$$(I - H)(I - H) = II - 2IH + HH = I - 2H + H = I - H$$

$I - H$ is the projection matrix onto the column space orthogonal to X. The rank of $I - H$ is $n - (p + 1)$.

By the Pythagorean Theorem,

$$\|\vec{y}\|^2 = \left\|\vec{\hat{y}}\right\|^2 + \|\vec{e}\|^2$$

Then

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum e_i^2$$

Do some more manipulations,

$$\begin{aligned}
\sum y_i^2 - n\bar{y}^2 &= \sum \hat{y}_i^2 - \bar{y}^2 + \sum e_i^2 \\
\sum (y_i - \bar{y})^2 &= \sum y_i^2 - \sum y_i\bar{y} - \sum \bar{y}y_i + \sum \bar{y}^2 \\
\sum y_i^2 - n\bar{y}^2 - n\bar{y}^2 + n\bar{y}^2 &= \sum y_i^2 - n\bar{y}^2 \\
\sum (\hat{y}_i - \bar{y})^2 &= \sum \hat{y}_i^2 - 2\bar{y}\sum \hat{y}_i + n\bar{y}^2 \\
&= \sum \hat{y}_i^2 - n\bar{y}^2 \\
\sum \hat{y}_i &= \hat{y} \cdot \vec{1}_n = \vec{y}^R\vec{1} \\
&= (H\vec{y})^T\vec{1} = \vec{y}^T H^T \vec{1} \\
&= \vec{y}^T H\vec{1} \\
&= \vec{y}^T\vec{1} \\
&= \sum y_i = n\bar{y}
\end{aligned}$$

Using the above formulations, we derive that

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum e_i^2$$

or $SST = SSR + SSE$. Note that SSR (sum of squared regression) is sometimes referred to as SSM (sum of squared model). This is the sum of squares equivalence for linear models Therefore

$$R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = \frac{S_y^2 - S_e^2}{S_y^2}$$

Then
$$\|\vec{y}-\bar{y}\|^2 = \left\|\vec{\hat{y}}-\bar{y}\right\|^2 + \|\vec{e}\|^2$$
In this scenario, a high R^2 can be thought of as $\vec{y}-\bar{y}$ being close to $\hat{y}-\bar{y}$ where the vectors are about the same height. On the other hand, a low R^2 can be thought of as $\vec{v}-\bar{y}$ being perpendicular to $\hat{y}-\bar{y}$. Additionally, R^2 is zero if $\vec{y}-\bar{y}=\vec{e}$ and $\hat{y}-\bar{y}=0$ or $\hat{y}=\bar{y}$.

Let $V=[\vec{v}_1,\vec{v}_2]$. Then
$$\text{proj}_v(\vec{a}) = \text{proj}_{\vec{v}_1}(\vec{a}) = \text{proj}_{\vec{v}_2}(\vec{a})$$
Note that $\|a+b\|^2 = \|a\|^2+\|b\|^2+2\|a\|\|b\|\cos\theta$. Then
$$\begin{aligned}\|\vec{a}\|^2 = \|\text{proj}_v(\vec{a})\|^2 = \left\|\text{proj}_{\vec{v}_1}(\vec{a})\right\|^2 + \left\|\text{proj}_{\vec{v}_2}(\vec{a})\right\|^2 + \|\vec{e}\|^2 + 2\left\|\text{proj}_{\vec{v}_1}(\vec{a})\right\|\left\|\text{proj}_{\vec{v}_2}(\vec{a})\right\|\cos\theta_{\vec{v}_1,\vec{v}_2} \\ + 2\left\|\text{proj}_{\vec{v}_1}(\vec{a})\right\|\|\vec{e}\|\cos\theta_{\vec{v}_1,\vec{e}} + 2\left\|\text{proj}_{\vec{v}_2}(\vec{a})\right\|\|\vec{e}\|\cos\theta_{\vec{v}_2,\vec{e}}\end{aligned}$$
The first cos value is only zero if $v_1\perp v_2$ while the latter two are always zero.
Assume V is orthogonal, then
$$\|\vec{a}\| = \sum_{i=1}^{k}\left\|\text{proj}_{\vec{v}_i}(\vec{a})\right\|^2 + \|\vec{v}\|^2$$
If V is orthogonal,
$$\begin{aligned}\text{proj}_v(\vec{a}) &= \text{proj}_{\vec{v}_1}(\vec{a}) + \text{proj}_{\vec{v}_2}(\vec{a}) \\ &= \frac{\vec{v}_1\vec{v}^T}{\|\vec{v}_1\|^2}\vec{a} + \frac{\vec{v}_2\vec{v}_2^T}{\|\vec{v}_2\|^2}\vec{a} \\ &= \Big(\frac{\vec{v}_1\vec{v}_1^T}{\|\vec{v}_1\|^2} + \frac{\vec{v}_2\vec{v}_2^T}{\|\vec{v}_2\|^2}\Big)\vec{a}\end{aligned}$$
If V is orthonormal (orthogonal and each column is normalized to length 1), then
$$\begin{aligned}\mathbb{P}(v)\vec{a} &= (\vec{v}_1\vec{v}_1^T + \vec{v}_2\vec{v}_2^T)\vec{a} \\ &= \left(\begin{bmatrix} v_{11}^2 & v_{11}v_{12} & \dots & v_{11}v_{1n} \\ v_{12}v_{11} & v_{12}^2 & \dots & v_{12}v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1n}v_{11} & \dots & \dots & v_{1n}^2\end{bmatrix} + \begin{bmatrix} v_{21}^2 & v_{21}v_{23} & \dots & v_{21}v_{2n} \\ v_{22}v_{23} & v_{22}^2 & \dots & v_{22}v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{2n}v_{21} & \dots & \dots & v_{2n}^2\end{bmatrix}\right)\vec{a} \\ &= \left(\begin{bmatrix}\uparrow & \uparrow \\ \vec{v}_1 & \vec{v}_2 \\ \downarrow & \downarrow\end{bmatrix}\begin{bmatrix}\leftarrow & \vec{v}_1^T & \rightarrow \\ \leftarrow & \vec{v}_2^T & \rightarrow\end{bmatrix}\right)\vec{a} \\ &= VV^T\vec{a}\end{aligned}$$
Consider $X\in\mathbb{R}^{n\times(p+1)}$ full rank and $Q\in\mathbb{R}^{n\times(p+1)}$ being orthonormal and thus full rank as well. Then
$$\hat{y} = X(X^TX)^{-1}X^T\vec{y} = QQ^T\vec{y}$$
Given X, to compute a Q with the same column space and orthonormal column vectors, use QR decomposition where $X=QR$.

Gram-Schmidt Algorithm (QR Decomposition): Denote Q as $[\vec{q}_{\cdot 1}, \vec{q}_{\cdot 2}, \ldots, \vec{q}_{\cdot p+1}]$.

1. Let $\vec{v}_{\cdot 1} = \vec{x}_{\cdot 1}$.
2. Let $\vec{q}_{\cdot 1} = \frac{\vec{v}_{\cdot 1}}{\|\vec{v}_{\cdot 1}\|^2}$.
3. Let $\vec{v}_{\cdot 2} = \vec{x}_{\cdot 2} - \text{proj}_{\vec{q}_{\cdot 1}}(\vec{x}_{\cdot 2})$.
4. Let $\vec{q}_{\cdot 2} = \frac{\vec{v}_{\cdot 2}}{\|\vec{v}_{\cdot 2}\|^2}$.
5. Let $\vec{v}_{\cdot 3} = \vec{x}_{\cdot 3} - \text{proj}_{\vec{v}_{\cdot 1}}(\vec{x}_{\cdot 3}) - \text{proj}_{\vec{v}_{\cdot 2}}(\vec{x}_{\cdot 3})$.
6. Let $\vec{q}_{\cdot 3} = \frac{\vec{v}_{\cdot 3}}{\|\vec{v}_{\cdot 3}\|^2}$.
7. Continue until $\vec{q}_{\cdot p+1}$ is created.