Section 6.3: Orthogonal Projections

In this section, we summarize what we discussed in Section 6.2, and provide some applications.

Theorem (orthonormal Decomposition) Let $\mathcal{B}=\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\}$ be an orthonormal basis of a subspace W in R^n. Then, the orthogonal projection of any vector $\vec{y}$ onto $\vec{u}_i$ and orthogonal projection of any vector $\vec{y}$ onto W have the following expressions, respectively,

$$\begin{aligned} \text{Proj}_{\vec{u}_i}(\vec{y}) &= (\vec{y}\cdot\vec{u}_i)\,\vec{u}_i, \quad i=1,2,...,p \\ \text{Proj}_W(\vec{y}) &= \sum_{i=1}^{p}(\vec{y}\cdot\vec{u}_i)\,\vec{u}_i = \sum_{i=1}^{p}\text{Proj}_{\vec{u}_i}(\vec{y}), \end{aligned}$$

and

$$\vec{y} = \text{Proj}_W(\vec{y}) + \vec{z}, \qquad \vec{z}\perp W.$$

Moreover, if we set $U = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_p]_{n\times p}$ to be a matrix whose columns are $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_p\}$, then

$$\text{Proj}_W(\vec{y}) = UU^T\vec{y}. \tag{1}$$

Proof. We only need to show (1). To this end, we see that

$$\begin{aligned} UU^T\vec{y} &= [\vec{u}_1, \vec{u}_2, ..., \vec{u}_p]\left(\begin{bmatrix}(\vec{u}_1)^T\\ (\vec{u}_2)^T\\ ...\\ (\vec{u}_p)^T\end{bmatrix}_{p\times n}\vec{y}\right) = [\vec{u}_1, \vec{u}_2, ..., \vec{u}_p]_{n\times p}\begin{bmatrix}(\vec{u}_1)^T\vec{y}\\ (\vec{u}_2)^T\vec{y}\\ ...\\ (\vec{u}_p)^T\vec{y}\end{bmatrix}_{p\times 1} \\ &= \vec{u}_1(\vec{u}_1)^T\vec{y} + \vec{u}_2(\vec{u}_2)^T\vec{y} + ... + \vec{u}_p(\vec{u}_p)^T\vec{y} \\ &= \left((\vec{u}_1)^T\vec{y}\right)\vec{u}_1 + \left((\vec{u}_2)^T\vec{y}\right)\vec{u}_2 + ... + \left((\vec{u}_p)^T\vec{y}\right)\vec{u}_p = \sum_{i=1}^{p}(\vec{y}\cdot\vec{u}_i)\,\vec{u}_i = \text{Proj}_W(\vec{y}). \end{aligned}$$

■

Theorem (The best approximation Theorem) Let W be a subspace. Then, for any vector $\vec{y}$, $\text{Proj}_W(\vec{y}) \in W$ is the best approximation to $\vec{y}$ by vectors in W. More precisely, $Proj_W(\vec{y})$ is the closest point in W to $\vec{y}$, i.e.,

$$\|\vec{y} - Proj_W(\vec{y})\| \le \|\vec{y} - \vec{w}\|, \quad \text{for any } \vec{w}\in W. \tag{2}$$

This closest distance is defined as the distance from $\vec{y}$ to W :

$$dist(\vec{y}, W) = \min\{\|\vec{y}-\vec{w}\| : \vec{w}\in W\} = \|\vec{y} - Proj_W(\vec{y})\|.$$

Proof. By definition, $\vec{z} = \vec{y} - Proj_W(\vec{y})$ is orthogonal to W. For any $\vec{w} \in W$, $\vec{z}$ is orthogonal to $\vec{w} - Proj_W(\vec{y}) \in W$.

$$\begin{aligned}
\|\vec{y} - \vec{w}\|^2 &= \|(\vec{y} - Proj_W(\vec{y})) - (\vec{w} - Proj_W(\vec{y}))\|^2 \\
&= \|\vec{z} - (\vec{w} - Proj_W(\vec{y}))\|^2 \\
&= \|\vec{z}\|^2 + \|\vec{w} - Proj_W(\vec{y})\|^2 \qquad \text{(by Pythagorean)} \\
&\geq \|\vec{z}\|^2 = \|\vec{y} - Proj_W(\vec{y})\|^2 .
\end{aligned}$$

■

Example 6.3.1. Let

$$\vec{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \ \vec{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \ \vec{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ \vec{z} = \begin{bmatrix} -1 \\ 5 \\ 10 \end{bmatrix}, \ W = Span\{\vec{u}_1, \ \vec{u}_2\}.$$

Determine whether $\vec{y}$ or $\vec{z}$ is closer to W.

Solution 1: We notice that $\vec{u}_1 \cdot \vec{u}_2 = 0$. Thus, $\{\vec{u}_1, \ \vec{u}_2\}$ is an orthogonal basis, and

$$\begin{aligned}
Proj_W(\vec{y}) &= \left(\frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\right)\vec{u}_1 + \left(\frac{\vec{y} \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2}\right)\vec{u}_2 \\
&= \frac{2 + 10 - 3}{4 + 25 + 1}\vec{u}_1 + \frac{-2 + 2 + 3}{4 + 1 + 1}\vec{u}_2 \\
&= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
Proj_W(\vec{z}) &= \left(\frac{\vec{z} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\right)\vec{u}_1 + \left(\frac{\vec{z} \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2}\right)\vec{u}_2 \\
&= \frac{-2 + 25 - 10}{4 + 25 + 1}\vec{u}_1 + \frac{2 + 5 + 10}{4 + 1 + 1}\vec{u}_2 \\
&= \frac{13}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{17}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
dist(\vec{y}, W) &= \|\vec{y} - Proj_W(\vec{y})\| \\
&= \left\| \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \left(\frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}\right) \right\| = \frac{7}{5}\sqrt{5}
\end{aligned}$$

$$\begin{aligned}
dist(\vec{z}, W) &= \|\vec{z} - Proj_W(\vec{z})\| \\
&= \left\| \begin{bmatrix} -1 \\ 5 \\ 10 \end{bmatrix} - \left(\frac{13}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{17}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}\right) \right\| = \frac{19}{5}\sqrt{5}
\end{aligned}$$

Ans: $\vec{y}$ is closer to W.

Solution 2: We first construct an orthonormal basis $\{\vec{v}_1, \vec{v}_2\}$ for W by

$$\vec{v}_1 = \frac{1}{\|\vec{u}_1\|}\vec{u}_1 = \frac{1}{\sqrt{30}}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad \vec{v}_2 = \frac{1}{\|\vec{u}_2\|}\vec{u}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}.$$

Set

$$U = [\vec{v}_1, \vec{v}_2] = \begin{bmatrix} \frac{2}{\sqrt{30}} & \frac{-2}{\sqrt{6}} \\ \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \end{bmatrix}.$$

Then,

$$UU^T = \begin{bmatrix} \frac{2}{\sqrt{30}} & \frac{-2}{\sqrt{6}} \\ \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{30}} & \frac{5}{\sqrt{30}} & -\frac{1}{\sqrt{30}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & 0 & -\frac{2}{5} \\ 0 & 1 & 0 \\ -\frac{2}{5} & 0 & \frac{1}{5} \end{bmatrix}$$

According to (1),

$$Proj_W(\vec{y}) = UU^T\vec{y} = \begin{bmatrix} \frac{4}{5} & 0 & -\frac{2}{5} \\ 0 & 1 & 0 \\ -\frac{2}{5} & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}$$

$$Proj_W(\vec{z}) = UU^T\vec{z} = \begin{bmatrix} \frac{4}{5} & 0 & -\frac{2}{5} \\ 0 & 1 & 0 \\ -\frac{2}{5} & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} -1 \\ 5 \\ 10 \end{bmatrix} = \begin{bmatrix} -\frac{24}{5} \\ 5 \\ \frac{12}{5} \end{bmatrix}.$$

Hence,

$$dist(\vec{y}, W) = \|\vec{y} - Proj_W(\vec{y})\| = \left\| \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix} \right\| = \frac{7}{5}\sqrt{5}$$

$$dist(\vec{z}, W) = \|\vec{z} - Proj_W(\vec{z})\| = \left\| \begin{bmatrix} -1 \\ 5 \\ 10 \end{bmatrix} - \begin{bmatrix} -\frac{24}{5} \\ 5 \\ \frac{12}{5} \end{bmatrix} \right\| = \frac{19}{5}\sqrt{5}$$

Note that the second solution method works better if one needs to computer several orthogonal projections.

Section 6.4: The Gram-Schmidt Process

The Gram-Schmidt process is an algorithm to produce an orthogonal/orthonormal basis.
Example 6.4.1 Find an orthonormal basis $\{\vec{v}_1, \vec{v}_2\}$ for W =Span$\{\vec{x}_1, \vec{x}_2\}$, where

$$\vec{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \ \vec{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Sol: We choose $\vec{v}_1 = \vec{x}_1 / \|\vec{x}_1\| = \vec{x}_1/\sqrt{45}$. Next, let $\vec{p}$ be the projection of $\vec{x}_2$ onto $\vec{v}_1$, i.e.,

$$\vec{p} = (\vec{x}_2 \cdot \vec{v}_1)\,\vec{v}_1 = \left(\frac{15}{\sqrt{45}}\right)\frac{1}{\sqrt{45}}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}.$$

We know that $\vec{x}_2 - \vec{p}$ is orthogonal to $\vec{v}_1$. So $\vec{v}_2 = (\vec{x}_2 - \vec{p}) / \|\vec{x}_2 - \vec{p}\|$, or

$$\vec{v}_2 = \left(\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right)/2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Theorem (The Gram-Schmidt process) ■ Given a basis $\{\vec{x}_1, \vec{x}_2, ..., \vec{x}_p\}$ for any subspace W of R^n, an orthonormal basis $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_p\}$ can be constructed as follows:

$$\vec{v}_1 = \frac{\vec{x}_1}{\|\vec{x}_1\|}, \ W_1 = Span\{\vec{v}_1\} = Span\{\vec{x}_1\}$$

$$\vec{u}_2 = \vec{x}_2 - \Pr oj_{W_1}\vec{x}_2 = \vec{x}_2 - (\vec{x}_2 \cdot \vec{v}_1)\,\vec{v}_1,$$
$$\vec{v}_2 = \frac{\vec{u}_2}{\|\vec{u}_2\|}$$
$$W_2 = Span\{\vec{v}_1, \vec{v}_2\} = Span\{\vec{x}_1, \vec{x}_2\}$$

$$\vec{u}_3 = \vec{x}_3 - \Pr oj_{W_2}\vec{x}_3 = \vec{x}_3 - (\vec{x}_3 \cdot \vec{v}_1)\,\vec{v}_1 - (\vec{x}_3 \cdot \vec{v}_2)\,\vec{v}_2,$$
$$\vec{v}_3 = \frac{\vec{u}_3}{\|\vec{u}_3\|}$$
$$W_3 = Span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = Span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\},$$

for *any* $k = 1, 2, 3, ..., p$,

$$\vec{u}_k = \vec{x}_k - \Pr oj_{W_{k-1}}\vec{x}_k = \vec{x}_k - (\vec{x}_k \cdot \vec{v}_1)\,\vec{v}_1 - ... - (\vec{x}_k \cdot \vec{v}_{k-1})\,\vec{v}_{k-1},$$
$$\vec{v}_k = \frac{\vec{u}_k}{\|\vec{u}_k\|}$$

$$W_k = Span\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_k\} = Span\{\vec{x}_1, \vec{x}_2, ..., \vec{x}_k\}.$$

Example 6.4.2. Find an orthonormal basis for the subspace $W = Span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, where

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Sol:

$$\vec{v}_1 = \frac{\vec{x}_1}{\|\vec{x}_1\|} = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

$$\vec{u}_2 = \vec{x}_2 - (\vec{x}_2 \cdot \vec{v}_1)\,\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \left(\frac{3}{2}\right)\frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v}_2 = \frac{\vec{u}_2}{\|\vec{u}_2\|} = \frac{1}{2\sqrt{3}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{u}_3 = \vec{x}_3 - (\vec{x}_3 \cdot \vec{v}_1)\,\vec{v}_1 - (\vec{x}_3 \cdot \vec{v}_2)\,\vec{v}_2 = \ \frac{1}{3}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} - -$$

$$\vec{v}_3 = \frac{\vec{u}_3}{\|\vec{u}_3\|} = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.$$

Theorem (The QR Factorization): Let A be any $m \times n$ matrix A with $\text{rank}(A) = n$. Then A can be factored as $A = QR$, where Q is an $m \times n$ matrix *whose* columns form an orthonormal basis for $Col\,(A)$ and R is an $n \times n$ upper triangle matrix with positive entries on its diagonal.

Proof: Using the Gram-Schmidt process, we can find an orthonormal basis $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ for $Col\,(A)$. Let Q be its column matrix:

$$Q = [\vec{v}_1 \ \vec{v}_2 \ ... \ \vec{v}_n]$$