Mostafa Sabri

Lecture 20

6.2 Orthogonal Sets

6.3 Orthogonal Projections

جامعة نيويورك أبوظبي
NYU ABU DHABI

Orthogonal Sets

A set of vectors $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an *orthogonal set* if $\mathbf{u}_i \cdot \mathbf{u}_j = \mathbf{0}$ for any $i \neq j$.

Show that the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthogonal if $\mathbf{u}_1 = (1, 2, 3)$, $\mathbf{u}_2 = (-6, 0, 2)$, $\mathbf{u}_3 = (1, -5, 3)$.

Thm: If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors, then S is linearly independent and hence a basis for $H = \mathrm{Span}(S)$.

Proof. Suppose $\sum_{i=1}^{p} c_i \mathbf{u}_i = \mathbf{0}$. Then $\mathbf{u}_j \cdot \sum_{i=1}^{p} c_i \mathbf{u}_i = 0$, so $\sum_{i=1}^{p} c_i \mathbf{u}_j \cdot \mathbf{u}_i = 0$, so $c_j \mathbf{u}_j \cdot \mathbf{u}_j = 0$, implying $c_j = 0$ since $\mathbf{u}_j \neq \mathbf{0}$. This holds for any j, so S is independent. $\square$

Orthogonal Basis

An orthogonal basis for a subspace $W \subset \mathbb{R}^n$ is a basis of W that is also an orthogonal set.

Thm: If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal basis of W, then any $\mathbf{w} \in W$ has the expansion $\mathbf{w} = \sum_{i=1}^{p} c_i \mathbf{u}_i$ for

$$c_i = \frac{\mathbf{w} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}$$

Proof. Let $\mathbf{w} \in W = \mathrm{Span}(S)$, so $\mathbf{w} = \sum_{i=1}^{p} c_i \mathbf{u}_i$ for some $c_i \in \mathbb{R}$. So $\mathbf{u}_j \cdot \mathbf{w} = \mathbf{u}_j \cdot \sum_{i=1}^{p} c_i \mathbf{u}_i = \sum_{i=1}^{p} c_i \mathbf{u}_j \cdot \mathbf{u}_i = c_j \mathbf{u}_j \cdot \mathbf{u}_j$. $\square$

Orthogonal Decomposition

Express $\mathbf{w} = (1, 1, 1)$ in the basis $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ of the previous exercise.

Fix a vector $\mathbf{u}$ and suppose we want to decompose an arbitrary $\mathbf{y} \in \mathbb{R}^n$ in the form

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z},$$

where $\widehat{\mathbf{y}} = \alpha\mathbf{u}$ is in $\mathrm{Span}\{\mathbf{u}\}$ and $\mathbf{z}$ is orthogonal to $\mathbf{u}$. Assuming this is possible, the vector $\mathbf{y} - \widehat{\mathbf{y}} = \mathbf{z}$ is orthogonal to $\mathbf{u}$. Thus,

$$0 = (\mathbf{y} - \widehat{\mathbf{y}}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \widehat{\mathbf{y}} \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha\mathbf{u} \cdot \mathbf{u}.$$

This shows that we must have $\alpha = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$.

Orthogonal Projection

Let $L = \mathrm{Span}\{\mathbf{u}\}$. We define the *orthogonal projection of* $\mathbf{y}$ *onto* $\mathbf{u}$ to be the vector $\widehat{\mathbf{y}}$ such that

1. $\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$,
2. $\widehat{\mathbf{y}} \in L$ and $\mathbf{z} \in L^{\perp}$.

In other words,

$$\mathrm{proj}_L(\mathbf{y}) = \widehat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$$

Let $\mathbf{u} = (1, 2)$ and $\mathbf{y} = (-3, 5)$. Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Then decompose $\mathbf{y}$ as $\widehat{\mathbf{y}} + \mathbf{z}$ where $\mathbf{z}\cdot\mathbf{u} = 0$.

Geometric interpretation of the weights in ob

If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal basis of W, then we now deduce that any $\mathbf{w} \in W$ can be expanded as

$$\mathbf{w} = \sum_{i=1}^{p} \mathrm{proj}_{L_i}(\mathbf{w})$$

where $L_i = \mathrm{Span}\{\mathbf{u}_i\}$.

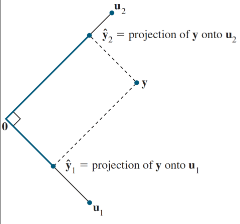

FIGURE 4 A vector decomposed into the sum of two projections.

Orthonormal Sets

A set $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an *orthonormal set* if it is an orthogonal set of unit vectors.

If so, S is an *orthonormal basis* of $H = \mathrm{Span}(S)$.

e.g. $S = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis of $\mathbb{R}^n$.

e.g. By normalizing the vectors in the first exercise we get an orthonormal basis of $\mathbb{R}^3$.

Thm: An $m \times n$ matrix U has orthonormal columns iff $U^T U = I_n$.

Orthonormal Columns

Proof. If $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ then

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \cdots & \mathbf{u}_1^T\mathbf{u}_n \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \cdots & \mathbf{u}_2^T\mathbf{u}_n \\ & & \ddots & \\ \mathbf{u}_n^T\mathbf{u}_1 & \mathbf{u}_n^T\mathbf{u}_2 & \cdots & \mathbf{u}_n^T\mathbf{u}_n \end{bmatrix}.$$

Thus, $U^T U = I$ iff $\mathbf{u}_i^T \mathbf{u}_i = 1 \ \forall i$ and $\mathbf{u}_i^T \mathbf{u}_j = 0 \ \forall i \neq j$. This occurs iff $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal set. $\square$

Thm: If U is an $m \times n$ matrix with orthonormal columns and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then

1. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$,
2. $\|U\mathbf{x}\| = \|\mathbf{x}\|$,
3. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0 \iff \mathbf{x} \cdot \mathbf{y} = 0$.

Orthonormal Columns

Proof. 1. $(U\mathbf{x})\cdot(U\mathbf{y}) = (\sum_{i=1}^{n} x_i\mathbf{u}_i)\cdot(\sum_{j=1}^{n} y_i\mathbf{u}_j) = \sum_{i,j=1}^{n} x_i y_j \mathbf{u}_i\cdot\mathbf{u}_j = \sum_{i=1}^{n} x_i y_i = \mathbf{x}\cdot\mathbf{y}$.

2. $\|U\mathbf{x}\|^2 = (U\mathbf{x})\cdot(U\mathbf{x}) = \mathbf{x}\cdot\mathbf{x} = \|\mathbf{x}\|^2$ by 1.

3. This follows from 1. $\square$

e.g. If $n \leq m$, then from any orthonormal basis $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ of $\mathbb{R}^m$ we can construct such an $m \times n$ matrix U by taking n vectors in S as columns.

If U is a *square matrix* with $U^T U = I$, we call U an *orthogonal matrix*. In this case, U is invertible with $U^{-1} = U^T$.

Orthogonal Decomposition

Thm: Let $W \subset \mathbb{R}^n$ be a subspace. Then any $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$$

where $\widehat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^{\perp}$. Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal basis for W, then

$$\widehat{\mathbf{y}} = \sum_{i=1}^{p} \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \mathbf{u}_i \qquad (1)$$

<u>Proof.</u> We can always find an orthogonal basis $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ for W (Lec 21). Given $\mathbf{y} \in \mathbb{R}^n$ and S, we now define $\widehat{\mathbf{y}}$ by (1) and define $\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}$.

Orthogonal Decomposition

Clearly $\hat{\mathbf{y}} \in W$. Next,

$$\mathbf{u}_j \cdot \mathbf{z} = \mathbf{u}_j \cdot \mathbf{y} - \mathbf{u}_j \cdot \sum_{i=1}^{p} \frac{\mathbf{y} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i} \mathbf{u}_i = \mathbf{u}_j \cdot \mathbf{y} - \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j} \mathbf{u}_j \cdot \mathbf{u}_j = 0$$

so $\mathbf{z} \in S^{\perp} = W^{\perp}$ as required.

To show uniqueness, suppose $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ is another expansion for $\mathbf{y}$ with $\hat{\mathbf{y}}_1 \in W$ and $\mathbf{z}_1 \in W^{\perp}$. Then $\hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$, so $\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$. This implies that $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1 \in W \cap W^{\perp} = \{\mathbf{0}\}$, so $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$. This implies $\mathbf{z}_1 - \mathbf{z} = \mathbf{0}$, so $\mathbf{z}_1 = \mathbf{z}$, completing the proof. $\square$

Find the orthogonal decomposition of $\mathbf{y}$ with respect to $W = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ if $\mathbf{u}_1 = (1, 2, 3)$, $\mathbf{u}_2 = (-6, 0, 2)$ and $\mathbf{y} = (1, 1, 1)$.

Geometric Interpretation

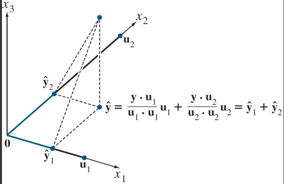

Figure: $\widehat{\mathbf{y}}$ is the sum of its projections onto mutually orthogonal lines.

We call $\widehat{\mathbf{y}}$ in the previous theorem the *orthogonal projection of* $\mathbf{y}$ *onto* W and denote it $\text{proj}_W(\mathbf{y}) = \widehat{\mathbf{y}}$.

Thm: If $\mathbf{y} \in W$, then $\text{proj}_W(\mathbf{y}) = \mathbf{y}$.

Proof. This follows from uniqueness of representation since we can write $\mathbf{y} = \mathbf{y} + \mathbf{0}$ with $\mathbf{0} \in W^\perp$. □

Best Approximation Theorem

Thm: If $W \subseteq \mathbb{R}^n$ is a subspace and $\mathbf{y} \in \mathbb{R}^n$, then $\widehat{\mathbf{y}} = \mathrm{proj}_W(\mathbf{y})$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \widehat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{w}\|$$

for all $\mathbf{w} \in W$, $\mathbf{w} \neq \widehat{\mathbf{y}}$.

This says that the best approximation of $\mathbf{y}$ inside W is $\widehat{\mathbf{y}}$.

<u>Proof.</u> Let $\mathbf{w} \in W$, $\mathbf{w} \neq \widehat{\mathbf{y}}$. Since $\widehat{\mathbf{y}} - \mathbf{w} \in W$, then $\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}$ is orthogonal to $\widehat{\mathbf{y}} - \mathbf{w}$. By Pythagoras, we get

$$\|\mathbf{y} - \mathbf{w}\|^2 = \|\mathbf{y} - \widehat{\mathbf{y}} + \widehat{\mathbf{y}} - \mathbf{w}\|^2 = \|\mathbf{y} - \widehat{\mathbf{y}}\|^2 + \|\widehat{\mathbf{y}} - \mathbf{w}\|^2$$

Since $\widehat{\mathbf{y}} - \mathbf{w} \neq \mathbf{0}$, this implies $\|\mathbf{y} - \mathbf{w}\|^2 > \|\mathbf{y} - \widehat{\mathbf{y}}\|^2$. $\square$

Best Approximation

1. If $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, $\mathbf{u}_1 = (1, 2, 3)$, $\mathbf{u}_2 = (-6, 0, 2)$ and $\mathbf{y} = (1, 1, 1)$, find the closest point in W to $\mathbf{y}$.

We define the distance from $\mathbf{y}$ to a subspace W to be the smallest distance of $\mathbf{y}$ to an element in W.

2. Find the distance of $\mathbf{y}$ to W if $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, $\mathbf{u}_1 = (-2, 1, 4)$, $\mathbf{u}_2 = (3, -2, 2)$ and $\mathbf{y} = (1, 5, -3)$.

Thm: If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a susbpace W of $\mathbb{R}^n$, then

$$\text{proj}_W(\mathbf{y}) = \sum_{i=1}^{p} (\mathbf{y} \cdot \mathbf{u}_i)\mathbf{u}_i\,. \tag{2}$$

Moreover, if $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix}$ then

$$\operatorname{proj}_W(\mathbf{y}) = UU^T\mathbf{y} \qquad \forall \mathbf{y} \in \mathbb{R}^n$$

Proof. Equation (2) follows from equation (1). Next, $UU^T\mathbf{y} = \sum_{i=1}^n (U^T\mathbf{y})_i \mathbf{u}_i = \sum_{i,j=1}^n u_{ji} y_j \mathbf{u}_i = \sum_{i=1}^n (\mathbf{u}_i \cdot \mathbf{y})\mathbf{u}_i = \operatorname{proj}_W(\mathbf{y})$. $\square$

Conclusion: If U is an $n \times p$ matrix with orthonormal columns and $W = \operatorname{Col}(U)$, then we have

$$U^T U = I_p \qquad \text{and} \qquad UU^T = \operatorname{proj}_W$$

If $p = n$, i.e. U is an orthogonal matrix, then $\operatorname{proj}_W = I_p$.