

What is on today

1 Orthogonal sets
2 Orthogonal projections

1 Orthogonal sets

Lay–Lay–McDonald §6.2 pp. 344 – 346

A set \( \{u_1, \ldots, u_p\} \) is an orthonormal set if it is an orthogonal set of unit vectors. If \( W \) is the subspace spanned by such a set, then \( \{u_1, \ldots, u_p\} \) is an orthonormal basis for \( W \), since the set is automatically linearly independent, by a theorem we saw in the previous class.

The simplest example of an orthonormal set is the standard basis \( \{e_1, \ldots, e_n\} \) for \( \mathbb{R}^n \). Here is another example:

Example 1. Show that \( \{v_1, v_2, v_3\} \) is an orthonormal basis of \( \mathbb{R}^3 \), where

\[
\begin{align*}
v_1 &= \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \\
v_2 &= \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \\
v_3 &= \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}.
\end{align*}
\]

Matrices whose columns form an orthonormal set are important in applications and in algorithms for matrix computations. Here are some properties of these matrices.

Theorem 2. An \( m \times n \) matrix \( U \) has orthonormal columns if and only if \( U^T U = I \).

Proof. To simplify notation, we suppose that \( U \) has 3 columns, each a vector in \( \mathbb{R}^m \). The
proof of the general case is essentially the same. Let \( U = [u_1 \quad u_2 \quad u_3] \) and compute

\[
U^T U = \begin{bmatrix}
u_1^T \\
u_2^T \\
u_3^T \\
\end{bmatrix}
\begin{bmatrix}
u_1 & \nu_2 & \nu_3 \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
u_1^T u_1 & \nu_1^T u_2 & \nu_1^T u_3 \\
\nu_2^T u_1 & \nu_2^T u_2 & \nu_2^T u_3 \\
\nu_3^T u_1 & \nu_3^T u_2 & \nu_3^T u_3 \\
\end{bmatrix}.
\]

The entries in the matrix are inner products. The columns of \( U \) are orthogonal iff

\[
u_1^T u_2 = \nu_2^T u_1 = 0, \quad \nu_1^T u_3 = \nu_3^T u_1 = 0, \quad \nu_2^T u_3 = \nu_3^T u_2 = 0.
\]

The columns of \( U \) all have unit length iff

\[
u_1^T u_1 = 1, \quad \nu_2^T u_2 = 1, \quad \nu_3^T u_3 = 1.
\]

\[
\square
\]

**Theorem 3.** Let \( U \) be an \( m \times n \) matrix with orthonormal columns, and let \( x \) and \( y \) be in \( \mathbb{R}^n \). Then

1. \( \|Ux\| = \|x\| \)
2. \( (Ux) \cdot (Uy) = x \cdot y \)
3. \( (Ux) \cdot (Uy) = 0 \) iff \( x \cdot y = 0 \).

The first and third properties say that the linear mapping \( x \mapsto Ux \) preserves lengths and orthogonality.

**Example 4.** Let \( U = \begin{bmatrix} 1/\sqrt{2} & 0 & 2/3 \\ 1/\sqrt{2} & 0 & -2/3 \\ 0 & 1/3 \\ \end{bmatrix} \) and \( x = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} \). Verify that \( U \) has orthonormal columns and that \( \|Ux\| = \|x\| \).

\[
v_1 \cdot v_2 = 0 \quad \|v_1\| = 1 \quad \|v_2\| = 1
\]

\[
\|u_1\| = \sqrt{1 + 1 + 1} = \sqrt{3} \quad \|u_2\| = \sqrt{1 + 1 + 1} = \sqrt{3}
\]

\[
\|u_3\| = \sqrt{1 + 1 + 1} = \sqrt{3}
\]

\[
\|u_1\| = \sqrt{2 + 3} = \sqrt{5} \quad \|u_2\| = \sqrt{2 + 3} = \sqrt{5}
\]

\[
\|u_3\| = \sqrt{2 + 3} = \sqrt{5}
\]

The previous two theorems are particularly useful when applied to square matrices. An **orthogonal matrix** is a square invertible matrix \( U \) such that \( U^{-1} = U^T \). (Such a matrix has orthonormal columns.) It is easy to see that any square matrix with orthonormal columns is an orthogonal matrix. Such a matrix must have orthonormal rows as well!
Example 5. Is the matrix \( U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix} \) orthogonal?

Any square matrix with orthonormal columns is an orthogonal matrix. From Ex. 1, we saw that these vectors formed an orthonormal basis of \( \mathbb{R}^3 \). So the matrix is orthogonal.

Example 6. Let \( U \) be an \( n \times n \) matrix with orthonormal columns. Show that \( \det U = \pm 1 \).

We have \( U^T U = I \) since \( U \) has orthonormal columns.

\[
\det(U^T U) = \det(I) = 1
\]

\[
\det(U^T) \det(U) = 1 \quad \Rightarrow \quad \det(U) \det(U) = 1 \quad \Rightarrow \quad \det(U)^2 = 1 \quad \Rightarrow \quad \det(U) = \pm 1.
\]

2 Orthogonal projections

Lay–Lay–McDonald §6.3 pp. 349 – 353

Example 7. Let \( \{u_1, \ldots, u_5\} \) be an orthogonal basis for \( \mathbb{R}^5 \) and let \( y = c_1u_1 + \cdots + c_5u_5 \). Consider the subspace \( W = \text{Span}\{u_1, u_2\} \) and write \( y \) as the sum of a vector \( z_1 \in W \) and a vector \( z_2 \in W^\perp \).

\[
y = \frac{c_1u_1 + c_2u_2 + c_3u_3 + c_4u_4 + c_5u_5}{z_1} + \frac{z_2}{z_2^\perp}?
\]

Let \( z_1 = c_1u_1 + c_2u_2 \). Check that \( z_2 \in W^\perp ? \) Show \( z_2 \) is \( \perp \) to \( u_1, u_2 \):

\[
z_2 \cdot u_1 = (c_3u_3 + c_4u_4 + c_5u_5) \cdot u_1 = c_3(u_3 \cdot u_1) + c_4(u_4 \cdot u_1) + c_5(u_5 \cdot u_1) = 0
\]

\[
z_2 \cdot u_2 = (c_3u_3 + c_4u_4 + c_5u_5) \cdot u_2 = c_3(u_3 \cdot u_2) + c_4(u_4 \cdot u_2) + c_5(u_5 \cdot u_2) = 0
\]

So \( z_1, z_2 \) have the right properties.

The next theorem shows that the decomposition \( y = z_1 + z_2 \) in the previous example can be computed without having an orthogonal basis for \( \mathbb{R}^n \); it's enough to have an orthogonal basis for \( W \).

Theorem 8 (Orthogonal Decomposition Theorem). Let \( W \) be a subspace of \( \mathbb{R}^n \). Then each \( y \in \mathbb{R}^n \) can be written uniquely in the form \( y = \hat{y} + z \), where \( \hat{y} \in W \) and \( z \in W^\perp \). In fact, if \( \{u_1, \ldots, u_p\} \) is any orthogonal basis of \( W \), then

\[
\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p
\]

and \( z = y - \hat{y} \).

The vector \( \hat{y} \) in the theorem is called the orthogonal projection of \( y \) onto \( W \) and often is written as

\[
\hat{y} = \text{proj}_W y.
\]
Example 9. Let \( \mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \) and \( \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \). Show that \( \{ \mathbf{u}_1, \mathbf{u}_2 \} \) is an orthogonal basis for \( W = \text{Span}\{ \mathbf{u}_1, \mathbf{u}_2 \} \). Write \( \mathbf{y} \) as the sum of a vector in \( W \) and a vector orthogonal to \( W \).

Now we study some properties of orthogonal projections. First, note that if \( \{ \mathbf{u}_1, \ldots, \mathbf{u}_p \} \) is an orthogonal basis of a subspace \( W \) and \( \mathbf{y} \in W \), then \( \text{proj}_W \mathbf{y} = \mathbf{y} \). This also follows from the next theorem:

**Theorem 10** (Best Approximation Theorem). Let \( W \) be a subspace of \( \mathbb{R}^n \), let \( \mathbf{y} \) be any vector in \( \mathbb{R}^n \) and let \( \hat{\mathbf{y}} \) be the orthogonal projection of \( \mathbf{y} \) onto \( W \). Then \( \hat{\mathbf{y}} \) is the closest point in \( W \) to \( \mathbf{y} \), in the sense that \( \| \mathbf{y} - \hat{\mathbf{y}} \| < \| \mathbf{y} - \mathbf{v} \| \) for all \( \mathbf{v} \in W \) distinct from \( \hat{\mathbf{y}} \).

The vector \( \hat{\mathbf{y}} \) is called the best approximation to \( \mathbf{y} \) by elements of \( W \).

Example 11. Let \( \mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \) and let \( W = \text{Span}\{ \mathbf{u}_1, \mathbf{u}_2 \} \). Find the closest point in \( W \) to \( \mathbf{y} \). (This closest point gives us the distance from \( \mathbf{y} \) to \( W \).)

The closest point in \( W \) to \( \mathbf{y} \) is given by

\[
\hat{\mathbf{y}} = \left( \begin{array}{c}
\frac{2}{25} \\
\frac{-5}{25} \\
\frac{-1}{25}
\end{array} \right)
\]

This was computed in Example 9.

The last theorem today shows how the formula for \( \text{proj}_W \mathbf{y} \) is simplified when the basis for \( W \) is an orthonormal set.

**Theorem 12.** If \( \{ \mathbf{u}_1, \ldots, \mathbf{u}_p \} \) is an orthonormal basis for a subspace \( W \) of \( \mathbb{R}^n \), then

\[
\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p.
\]

If \( U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p] \) then

\[
\text{proj}_W \mathbf{y} = UU^T \mathbf{y}
\]

for all \( \mathbf{y} \in \mathbb{R}^n \).