Professor Jennifer Balakrishnan, *jbala@bu.edu*

What is on today

1 Orthogonal sets

Lay–Lay–McDonald §6.2 pp. 344 – 346

A set $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an *orthonormal set* if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ 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 $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Here is another example:

Example 1. *Show that* $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ *is an orthonormal basis of* $\mathbb{R}^3$, *where* $\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$.

$\boxed{\text{Check orthonormal = check pairwise orthogonal \& each vector has norm 1}}$

orthogonal?

$v_1 \cdot v_2 = \frac{1}{\sqrt{66}}(-3+2+1) = 0$ ✓

$v_1 \cdot v_3 = \frac{1}{\sqrt{66}} \cdot \frac{1}{\sqrt{11}}(-3-4+7) = 0$ ✓

$v_2 \cdot v_3 = \frac{1}{\sqrt{6}} \cdot \frac{1}{\sqrt{66}}(1-8+7) = 0$ ✓

unit vectors?

$\|v_1\| = \sqrt{v_1 \cdot v_1} = \sqrt{\frac{1}{11}(3^2+1^2+1^2)} = \sqrt{1} = 1$ ✓

$\|v_2\| = \sqrt{v_2 \cdot v_2} = \sqrt{\frac{1}{6}(1+4+1)} = 1$ ✓

$\|v_3\| = \sqrt{v_3 \cdot v_3} = \sqrt{\frac{1}{66}(1+16+49)} = 1$ ✓

$\Rightarrow \{v_1, v_2, v_3\}$ is an orthonormal set $\Rightarrow$ orthonormal basis of $\mathbb{R}^3$.

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 $U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3]$ and compute

$$\begin{aligned} U^T U &= \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] \leftarrow u_i \in \mathbb{R}^m \\ &= \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix}. \end{aligned}$$

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

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0.$$

The columns of U all have unit length iff

$$\underbrace{\mathbf{u}_1^T\mathbf{u}_1}_{= u_1 \cdot u_1} = 1, \quad \underbrace{\mathbf{u}_2^T\mathbf{u}_2}_{= u_2 \cdot u_2} = 1, \quad \underbrace{\mathbf{u}_3^T\mathbf{u}_3}_{= u_3 \cdot u_3} = 1.$$

□

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

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

The first and third properties say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality.

Example 4. *Let* $U = \begin{array}{c} \begin{array}{cc} v_1 & v_2 \end{array} \\ \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \end{array}$ *and* $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. *Verify that U has orthonormal columns and that $||U\mathbf{x}|| = ||\mathbf{x}||$.*

orthonormal columns: $[v_1 \ v_2]$

want to check: $v_1 \cdot v_2 \stackrel{?}{=} 0$ + $\|v_1\| \stackrel{?}{=} 1$, $\|v_2\| \stackrel{?}{=} 1$

$v_1 \cdot v_2 = \frac{1}{\sqrt{2}} \cdot \frac{1}{3}(2 - 2 + 0) = 0$ ✓

$\|v_1\| = \sqrt{\frac{1}{2}(1+1+0)} = 1$ ✓

$\|v_2\| = \sqrt{\frac{1}{9}(4+4+1)} = 1$ ✓

$\|Ux\| = \left\| \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} \right\| = \sqrt{9+1+1} = \sqrt{11}$ $\quad \|x\| = \sqrt{2+3^2} = \sqrt{11} \Rightarrow \|Ux\| = \|x\|$.

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
$\Rightarrow$ from Ex. 1, we saw that these vectors formed an orthonormal basis of $\mathbb{R}^3$
$\Rightarrow$ 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.

$\det(U^TU) = \det(I) = 1$

$\det(U^TU) = 1$

$\det(U^T)\det(U) = 1 \Rightarrow \det(U)\det(U) = 1 \Rightarrow (\det U)^2 = 1 \Rightarrow \det U = \pm 1.$ $\square$

$\det(AB) = \det A \det B$

$\det(A^T) = \det(A)$

2 Orthogonal projections

Lay–Lay–McDonald §6.3 pp. 349 – 353

Example 7. *Let $\{\mathbf{u}_1, \dots, \mathbf{u}_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and let $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_5\mathbf{u}_5$. Consider the subspace $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ and write $\mathbf{y}$ as the sum of a vector $\mathbf{z}_1 \in W$ and a vector $\mathbf{z}_2 \in W^\perp$.*

$y = \underbrace{c_1u_1 + c_2u_2}_{= z_1} + \underbrace{c_3u_3 + c_4u_4 + c_5u_5}_{= z_2 ?}$

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:

$z_2 \cdot u_1 = (c_3u_3 + c_4u_4 + c_5u_5)\cdot u_1 = c_3 \cancel{u_3 \cdot u_1} + c_4 \cancel{u_4 \cdot u_1} + c_5 \cancel{u_5 \cdot u_1} = 0$

$z_2 \cdot u_2 = (c_3u_3 + c_4u_4 + c_5u_5)\cdot u_2 = c_3(\cancel{u_3 \cdot u_2}) + c_4(\cancel{u_4 \cdot u_2}) + c_5(\cancel{u_5 \cdot u_2}) = 0$

So z_1, z_2 have the right properties.

The next theorem shows that the decomposition $y = \mathbf{z}_1 + \mathbf{z}_2$ 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 $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. In fact, if $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is any orthogonal basis of W, then*

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

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

$$\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}.$$

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

① $u_1 \cdot u_2 = -4+5-1 = 0$ ✓
u_1 and u_2 are orthogonal
$u_1 \neq k u_2$, so they're lin. indep
$\Rightarrow \{u_1, u_2\}$ form an orthogonal basis for W.

② $y = \overset{\text{vector in } W}{\hat{y}} + z \leftarrow$ orthog to W.

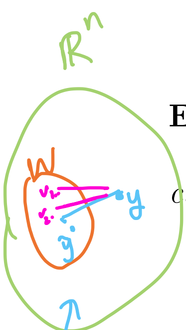
$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2$$
$$= \frac{2+10-3}{4+25+1}\begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + \frac{-2+2+3}{4+1+1}\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$
$$= \frac{3}{10}\begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{6}{10} - 1 \\ \frac{15}{10} + \frac{1}{2} \\ -\frac{3}{10} + \frac{1}{2} \end{pmatrix} = \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix}$$

$$\overset{\parallel}{y - \hat{y}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix} = \begin{pmatrix} 7/5 \\ 0 \\ 14/5 \end{pmatrix}$$

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 $\operatorname{proj}_W \mathbf{y} = \mathbf{y}$. 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.



$\hat{y}$ minimizes distance to y.

Example 11. *Let* $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, *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 y is given by
$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2$$
This was computed in Example 9:
$$\hat{y} = \begin{pmatrix} -2/5 \\ 2 \\ 1/5 \end{pmatrix}.$$

The last theorem today shows how the formula for $\operatorname{proj}_W \mathbf{y}$ 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*

$$\operatorname{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p.$$

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

$$\operatorname{proj}_W \mathbf{y} = UU^T\mathbf{y}$$

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