Remember from last week:

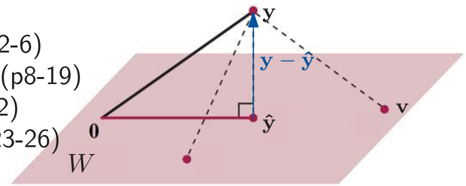
Theorem 9: Best Approximation Theorem: Let W be a subspace of $\mathbb{R}^n$, and $\mathbf{y}$ a vector in $\mathbb{R}^n$. Then the closest point in W to $\mathbf{y}$ is the unique point $\hat{\mathbf{y}}$ in W such that $\mathbf{y}-\hat{\mathbf{y}}$ is in $W^{\perp}$. In other words, $\|\mathbf{y}-\hat{\mathbf{y}}\| < \|\mathbf{y}-\mathbf{v}\|$ for all $\mathbf{v}$ in W with $\mathbf{v} \neq \hat{\mathbf{y}}$.

We proved last week that, if $\hat{\mathbf{y}}$ is in W, and $\mathbf{y}-\hat{\mathbf{y}}$ is in $W^{\perp}$, then $\hat{\mathbf{y}}$ is the unique closest point in W to $\mathbf{y}$. But we did not prove that a $\hat{\mathbf{y}}$ satisfying these conditions always exist.

We will show that the function $\mathbf{y} \mapsto \hat{\mathbf{y}}$ is a linear transformation, called the orthogonal projection onto W, and calculate it using an orthogonal basis for W.

Our remaining goals:

§6.2 The properties of orthogonal bases (p2-6)
§6.3 Calculating the orthogonal projection (p8-19)
§6.4 Constructing orthogonal bases (p20-22)
§6.2 Matrices with orthogonal columns (p23-26)



§6.2: Orthogonal Bases

Definition:
- A set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an *orthogonal set* if each pair of distinct vectors from the set is orthogonal, i.e. if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$.
- A set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an *orthonormal set* if it is an orthogonal set and each $\mathbf{u}_i$ is a unit vector.

Example: $\left\{ \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\5\\-3\end{bmatrix}, \begin{bmatrix}3\\0\\-1\end{bmatrix} \right\}$ is an orthogonal set, because

$$\begin{bmatrix}1\\2\\3\end{bmatrix} \cdot \begin{bmatrix}-1\\5\\-3\end{bmatrix} = -1+10-9=0, \quad \begin{bmatrix}1\\2\\3\end{bmatrix} \cdot \begin{bmatrix}3\\0\\-1\end{bmatrix} = 3+0-3=0, \quad \begin{bmatrix}-1\\5\\-3\end{bmatrix} \cdot \begin{bmatrix}3\\0\\-1\end{bmatrix} = -3+0+3=0.$$

To obtain an orthonormal set, we normalise each vector in the set:

$$\left\{ \begin{bmatrix}1/\sqrt{14}\\2/\sqrt{14}\\3/\sqrt{14}\end{bmatrix}, \begin{bmatrix}-1/\sqrt{35}\\5/\sqrt{35}\\-3/\sqrt{35}\end{bmatrix}, \begin{bmatrix}3/\sqrt{10}\\0\\-1/\sqrt{10}\end{bmatrix} \right\} \text{ is an orthonormal set.}$$

EXAMPLE: In $\mathbb{R}^6$, the set $\{\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_5, \mathbf{e}_6, \mathbf{0}\}$ is an orthogonal set, because $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ for all $i \neq j$, and $\mathbf{e}_i \cdot \mathbf{0} = 0$.

So an orthogonal set **may contain the zero vector**. But when it doesn't:

THEOREM 4 If $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal set of nonzero vectors, then it is linearly independent.

PROOF We need to show that ________________ is the only solution to

$$(*)$$

Take the dot product of both sides with $\mathbf{v}_1$:

$$c_1\_\_\_\_ + c_2\_\_\_\_ + \cdots + c_p\_\_\_\_ = \_\_\_\_$$

If $j \neq 1$, then $\mathbf{v}_j \cdot \mathbf{v}_1 =$ _____, so

$$c_1\_\_\_\_ + c_2\_\_\_\_ + \cdots + c_p\_\_\_\_ = \_\_\_\_$$

Because ______________, we have $\mathbf{v}_1 \cdot \mathbf{v}_1$ is nonzero, so it must be that $c_1 = 0$.

By taking the dot product of $(*)$ with each of the other $\mathbf{v}_i$s and using this argument, each c_i must be 0.

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal set of nonzero vectors, as before, and use the same idea with

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p. \qquad (*)$$

Take the dot product of both sides with $\mathbf{v}_1$:

$$\mathbf{y}\cdot\mathbf{v}_1 = (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p)\cdot\mathbf{v}_1$$

$$\mathbf{y}\cdot\mathbf{v}_1 = c_1\mathbf{v}_1\cdot\mathbf{v}_1 + c_2\mathbf{v}_2\cdot\mathbf{v}_1 + \dots + c_p\mathbf{v}_p\cdot\mathbf{v}_1.$$

Using that $\mathbf{v}_j\cdot\mathbf{v}_1 = 0$ whenever $j \neq 1$:

$$\mathbf{y}\cdot\mathbf{v}_1 = c_1\mathbf{v}_1\cdot\mathbf{v}_1 + c_2 0 \qquad + \dots + c_p 0$$

Since $\mathbf{v}_1$ is nonzero, $\mathbf{v}_1\cdot\mathbf{v}_1$ is nonzero, we can divide both sides by $\mathbf{v}_1\cdot\mathbf{v}_1$:

$$\frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1} = c_1$$

By taking the dot product of $(*)$ with each of the other $\mathbf{v}_j$s and using this argument, we obtain $c_j = \dfrac{\mathbf{y}\cdot\mathbf{v}_j}{\mathbf{v}_j\cdot\mathbf{v}_j}$.

So finding the weights in a linear combination of orthogonal vectors is much easier than for arbitrary vectors (see the example on p6).

Definition:
- A set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an *orthogonal basis* for a subspace W if it is both an orthogonal set and a basis for W.
- A set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an *orthonormal basis* for a subspace W if it is both an orthonormal set and a basis for W.

Example: The standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

By the previous theorem, any orthogonal set of nonzero vectors is an orthogonal basis for its span.

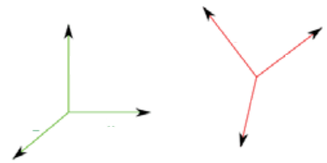
As proved on the previous page, a big advantage of orthogonal bases is:

Theorem 5: Weights for Orthogonal Bases: If $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W, then, for each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{v}_1 + \dots + c_p\mathbf{v}_p$$

are given by

$$c_j = \frac{\mathbf{y}\cdot\mathbf{v}_j}{\mathbf{v}_j\cdot\mathbf{v}_j}.$$



In particular, if $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis, then the weights are $c_j = \mathbf{y}\cdot\mathbf{u}_j$.

Example: Express $\begin{bmatrix}10\\9\\0\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\5\\-3\end{bmatrix}, \begin{bmatrix}3\\0\\-1\end{bmatrix}$.

Slow Answer: (works for any basis)

$$\left[\begin{array}{ccc|c}1 & -1 & 3 & 10\\ 2 & 5 & 0 & 9\\ 3 & -3 & -1 & 0\end{array}\right]$$

$$\begin{array}{l}R_2 - 2R_1\\ R_3 - 3R_1\end{array} \left[\begin{array}{ccc|c}1 & -1 & 3 & 10\\ 0 & 7 & -6 & -11\\ 0 & 0 & -10 & -30\end{array}\right]$$

$$R_3/-10 \left[\begin{array}{ccc|c}1 & -1 & 3 & 10\\ 0 & 7 & -6 & -11\\ 0 & 0 & 1 & 3\end{array}\right]$$

$$\begin{array}{l}R_1 - 3R_3\\ R_2 + 6R_3\end{array} \left[\begin{array}{ccc|c}1 & -1 & 0 & 1\\ 0 & 7 & 0 & 7\\ 0 & 0 & 1 & 3\end{array}\right]$$

$$R_2/7 \left[\begin{array}{ccc|c}1 & -1 & 0 & 1\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 3\end{array}\right]$$

$$R_1 + R_2 \left[\begin{array}{ccc|c}1 & 0 & 0 & 2\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 3\end{array}\right]$$

So $\begin{bmatrix}10\\9\\0\end{bmatrix} = 2\begin{bmatrix}1\\2\\3\end{bmatrix} + 1\begin{bmatrix}-1\\5\\-3\end{bmatrix} + 3\begin{bmatrix}3\\0\\-1\end{bmatrix}$.

Example: Express $\begin{bmatrix}10\\9\\0\end{bmatrix}$ as a linear combination of $\begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}-1\\5\\-3\end{bmatrix}, \begin{bmatrix}3\\0\\-1\end{bmatrix}$.

Fast Answer: (for an orthogonal basis) We showed on p2 that these three vectors form an orthogonal set. Since the vectors are nonzero, the set is linearly independent, and is therefore a basis for $\mathbb{R}^3$. Now use the formula $c_j = \dfrac{\mathbf{y}\cdot\mathbf{v}_j}{\mathbf{v}_j\cdot\mathbf{v}_j}$:

$$c_1 = \frac{\begin{bmatrix}10\\9\\0\end{bmatrix}\cdot\begin{bmatrix}1\\2\\3\end{bmatrix}}{\begin{bmatrix}1\\2\\3\end{bmatrix}\cdot\begin{bmatrix}1\\2\\3\end{bmatrix}} = \frac{10+18+0}{1^2+2^2+3^2} = 2, \quad c_2 = \frac{\begin{bmatrix}10\\9\\0\end{bmatrix}\cdot\begin{bmatrix}-1\\5\\3\end{bmatrix}}{\begin{bmatrix}-1\\5\\3\end{bmatrix}\cdot\begin{bmatrix}-1\\5\\3\end{bmatrix}} = \frac{-10+45+0}{(-1)^2+5^2+(-3)^2} = 1,$$

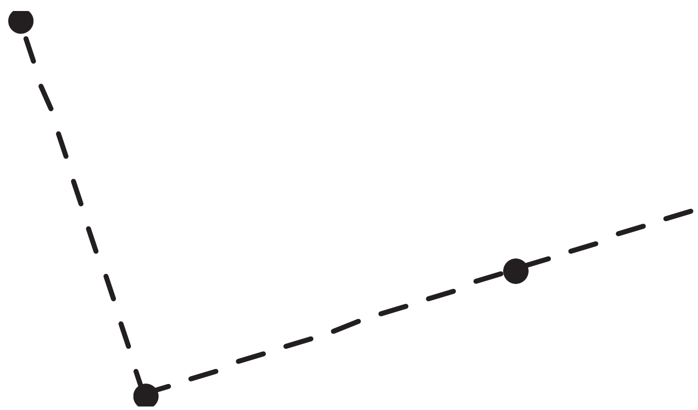
$$c_3 = \frac{\begin{bmatrix}10\\9\\0\end{bmatrix}\cdot\begin{bmatrix}3\\0\\-1\end{bmatrix}}{\begin{bmatrix}3\\0\\-1\end{bmatrix}\cdot\begin{bmatrix}3\\0\\-1\end{bmatrix}} = \frac{30+0+0}{3^2+0+(-1)^2} = 3,$$

so $\begin{bmatrix}10\\9\\0\end{bmatrix} = 2\begin{bmatrix}1\\2\\3\end{bmatrix} + 1\begin{bmatrix}-1\\5\\-3\end{bmatrix} + 3\begin{bmatrix}3\\0\\-1\end{bmatrix}$.

From the Weights for Orthogonal Bases Theorem: if $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W in $\mathbb{R}^n$, then each $\mathbf{y}$ in W is

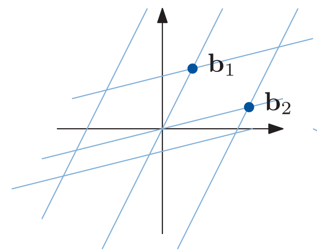
$$\mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p.$$

A geometric interpretation of this decomposition in $\mathbb{R}^2$:

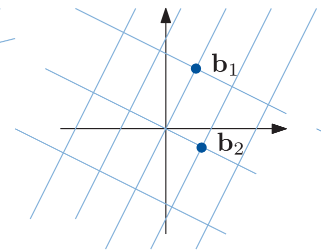


A geometric comparison of bases with different properties:

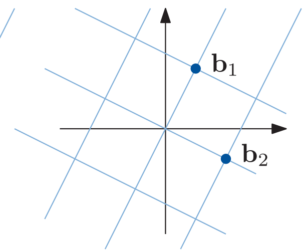
arbitrary basis - parallelogram grid



orthogonal basis - rectangular grid



orthonormal basis - square grid



§6.3: Orthogonal Projections

Recall that our motivation for defining orthogonal bases is to calculate the unique closest point in a subspace.
Let W be a subspace, and $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be an orthogonal basis for W. Let $\mathbf{y}$ be any vector, and $\hat{\mathbf{y}}$ be the vector in W that is closest to $\mathbf{y}$.
Since $\hat{\mathbf{y}}$ is in W, and $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a basis for W, we must have $\hat{\mathbf{y}} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ for some weights $c_1, \dots, c_p$.
We know from the Best Approximation Theorem that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^\perp$. By the properties of $W^\perp$, it's enough to show that $(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{v}_i = 0$ for each i. We can use this condition to solve for c_i:

$$\begin{aligned}
(\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{v}_1 &= 0\\
(\mathbf{y} - c_1\mathbf{v}_1 - c_2\mathbf{v}_2 - \cdots - c_p\mathbf{v}_p) \cdot \mathbf{v}_1 &= 0\\
\mathbf{y} \cdot \mathbf{v}_1 - c_1\mathbf{v}_1 \cdot \mathbf{v}_1 - c_2\mathbf{v}_2 \cdot \mathbf{v}_1 - \cdots - c_p\mathbf{v}_p \cdot \mathbf{v}_1 &= 0\\
\mathbf{y} \cdot \mathbf{v}_1 - c_1\mathbf{v}_1 \cdot \mathbf{v}_1 - c_2\, 0 \qquad - \cdots - c_p 0 \qquad &= 0
\end{aligned}$$

so $c_1 = \dfrac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}$. Similarly, $c_i = \dfrac{\mathbf{y} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$.

So we have proved (using the Best Approximation Theorem to deduce the uniqueness of $\hat{\mathbf{y}}$):
Theorem 8: Orthogonal Decomposition Theorem: Let W be a subspace of $\mathbb{R}^n$. Then every $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely as $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ with $\hat{\mathbf{y}}$ in W and $\mathbf{z}$ in $W^\perp$. In fact, if $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is any orthogonal basis for W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{v}_p}{\mathbf{v}_p \cdot \mathbf{v}_p}\mathbf{v}_p \quad \text{and} \quad \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}.$$

(Technically, to complete the proof, we need to show that every subspace has an orthogonal basis - see p20-23 for an explicit construction.)

Definition: The *orthogonal projection onto W* is the function $\text{proj}_W : \mathbb{R}^n \to \mathbb{R}^n$ such that $\text{proj}_W(\mathbf{y})$ is the unique $\hat{\mathbf{y}}$ in the above theorem. The image vector $\text{proj}_W(\mathbf{y})$ is the *orthogonal projection of $\mathbf{y}$ onto W*.
The uniqueness part of the theorem means that the $\text{proj}_W(\mathbf{y})$ does not depend on the orthogonal basis used to calculate it.

Example: Let $\mathbf{y} = \begin{bmatrix} 6 \\ 7 \\ -2 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$ and let $W = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$.
Find the point in W closest to $\mathbf{y}$ and the distance from $\mathbf{y}$ to W.
Answer: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, so $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for W. So the point in W closest to $\mathbf{y}$ is

$$\text{Proj}_W(\mathbf{y}) = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{y} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \frac{\begin{bmatrix} 6 \\ 7 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \frac{\begin{bmatrix} 6 \\ 7 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 5 \\ 3 \end{bmatrix}}{\begin{bmatrix} -1 \\ 5 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 5 \\ 3 \end{bmatrix}} \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$$

$$= \frac{9 + 14 - 9}{1^2 + 2^2 + 3^2} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \frac{-9 + 35 + 9}{(-1)^2 + 5^2 + (-3)^2} \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 7 \\ 0 \end{bmatrix}.$$

So the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \text{Proj}_W(\mathbf{y})\| = \left\| \begin{bmatrix} 6 \\ 7 \\ -2 \end{bmatrix} - \begin{bmatrix} 0 \\ 7 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 6 \\ 0 \\ -2 \end{bmatrix} \right\| = \sqrt{40}.$

Theorem 8: Orthogonal Decomposition Theorem: Let W be a subspace of $\mathbb{R}^n$. Then every $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely as $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ with $\hat{\mathbf{y}}$ in W and $\mathbf{z}$ in $W^\perp$. In fact, if $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is any orthogonal basis for W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{v}_p}{\mathbf{v}_p \cdot \mathbf{v}_p}\mathbf{v}_p \quad \text{and} \quad \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}.$$

The Best Approximation Theorem tells us that $\hat{\mathbf{y}}$ and $\mathbf{z}$ are unique, but here is an alternative proof that does not use the distance between $\hat{\mathbf{y}}$ and $\mathbf{y}$.
Suppose $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ and $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ are two such decompositions, so $\hat{\mathbf{y}}, \hat{\mathbf{y}}_1$ are in W, and $\mathbf{z}, \mathbf{z}_1$ are in $W^\perp$, and

$$\begin{aligned}
\hat{\mathbf{y}} + \mathbf{z} &= \hat{\mathbf{y}}_1 + \mathbf{z}_1\\
\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 &= \mathbf{z}_1 - \mathbf{z}.
\end{aligned}$$

LHS: Because $\hat{\mathbf{y}}, \hat{\mathbf{y}}_1$ are in W and W is a subspace, the difference $\hat{\mathbf{y}} - \hat{\mathbf{y}}_1$ is in W.
RHS: Because $\mathbf{z}, \mathbf{z}_1$ are in $W^\perp$ and $W^\perp$ is a subspace, the difference $\mathbf{z}_1 - \mathbf{z}$ is in $W^\perp$.
So the vector $\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$ is in both W and $W^\perp$, this vector is the zero vector (property 1 on week 11, p10). So $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$ and $\mathbf{z}_1 = \mathbf{z}$.

Theorem 8: Orthogonal Decomposition Theorem: Let W be a subspace of $\mathbb{R}^n$. Then every $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely as $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$ with $\hat{\mathbf{y}}$ in W and $\mathbf{z}$ in $W^\perp$. In fact, if $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is any orthogonal basis for W, then

$$\hat{\mathbf{y}} = \text{Proj}_W(\mathbf{y}) = \frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \dots + \frac{\mathbf{y}\cdot\mathbf{v}_p}{\mathbf{v}_p\cdot\mathbf{v}_p}\mathbf{v}_p \quad \text{and} \quad \mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}.$$

The formula for $\text{Proj}_W(\mathbf{y})$ above is similar to the Weights for Orthogonal Bases Theorem (p5). Let's look at how they are related.

For a vector $\mathbf{y}$ in W, the Weights for Orthogonal Bases Theorem says that $\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \dots + \frac{\mathbf{y}\cdot\mathbf{v}_p}{\mathbf{v}_p\cdot\mathbf{v}_p}\mathbf{v}_p = \text{Proj}_W(\mathbf{y})$. This makes sense because, if $\mathbf{y}$ is already in W, then the closest point in W to $\mathbf{y}$ must be $\mathbf{y}$ itself.

If $\mathbf{y}$ is not in W, then suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is part of a larger orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p, \mathbf{v}_{p+1}, \dots, \mathbf{v}_n\}$ for $\mathbb{R}^n$. So the Weights for Orthogonal Bases Theorem says that $\mathbf{y} = \underbrace{\frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \dots + \frac{\mathbf{y}\cdot\mathbf{v}_p}{\mathbf{v}_p\cdot\mathbf{v}_p}\mathbf{v}_p}_{\text{Proj}_W\mathbf{y}} + \underbrace{\frac{\mathbf{y}\cdot\mathbf{v}_{p+1}}{\mathbf{v}_{p+1}\cdot\mathbf{v}_{p+1}}\mathbf{v}_{p+1} + \dots + \frac{\mathbf{y}\cdot\mathbf{v}_n}{\mathbf{v}_n\cdot\mathbf{v}_n}\mathbf{v}_n}_{\mathbf{z}}.$

If an orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ for W is part of a larger orthogonal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_p, \mathbf{v}_{p+1}, \dots, \mathbf{v}_n\}$ for $\mathbb{R}^n$, then

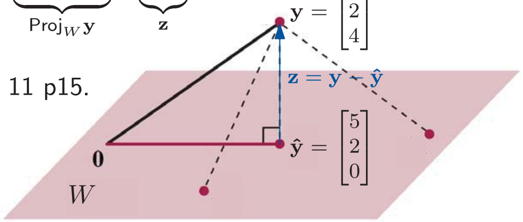
$$\mathbf{y} = \underbrace{\frac{\mathbf{y}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \dots + \frac{\mathbf{y}\cdot\mathbf{v}_p}{\mathbf{v}_p\cdot\mathbf{v}_p}\mathbf{v}_p}_{\text{Proj}_W\mathbf{y}} + \underbrace{\frac{\mathbf{y}\cdot\mathbf{v}_{p+1}}{\mathbf{v}_{p+1}\cdot\mathbf{v}_{p+1}}\mathbf{v}_{p+1} + \dots + \frac{\mathbf{y}\cdot\mathbf{v}_n}{\mathbf{v}_n\cdot\mathbf{v}_n}\mathbf{v}_n}_{\mathbf{z}}.$$

Example: Consider the orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for $\mathbb{R}^3$. Let

$W = \text{Span}\{\mathbf{e}_1, \mathbf{e}_2\}$, and $\mathbf{y} = \begin{bmatrix}5\\2\\4\end{bmatrix} = \underbrace{5\,\mathbf{e}_1 + 2\mathbf{e}_2}_{\text{Proj}_W\mathbf{y}} + \underbrace{4\mathbf{e}_3}_{\mathbf{z}}$.

So $\text{Proj}_W(\mathbf{y}) = \begin{bmatrix}5\\2\\0\end{bmatrix}$, as we saw week 11 p15.

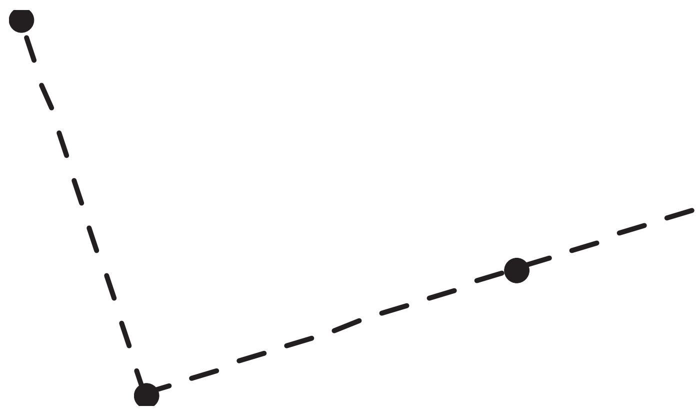
So, informally, the orthogonal projection "changes the coordinates outside W to 0".



Let W be a subspace of $\mathbb{R}^n$. If $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is an orthonormal basis for W, then, for every $\mathbf{y}$ in $\mathbb{R}^n$,

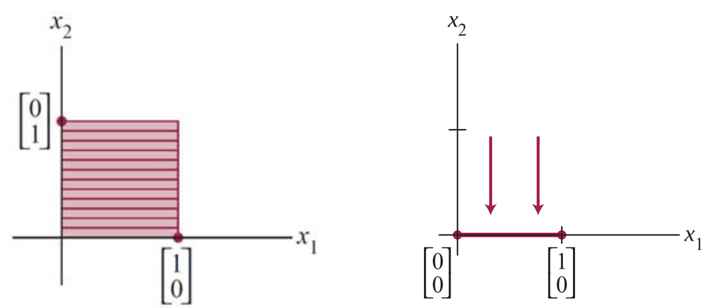
$$\text{proj}_W\mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + \dots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p.$$

Thinking about $\text{proj}_W : \mathbb{R}^n \to \mathbb{R}^n$ as a function:



We saw a special case in Week 4 §1.8-1.9:

Projection onto the x_1-axis



Properties of the function $\text{proj}_W : \mathbb{R}^n \to \mathbb{R}^n$:

a. proj_W is a linear transformation.
b. $\text{proj}_W(\mathbf{y}) = \mathbf{y}$ if and only if $\mathbf{y}$ is in W.
c. The range of proj_W is W.
d. The kernel of proj_W is $W^\perp$.
e. $\text{proj}_W^2 = \text{proj}_W$.
f. $\text{proj}_W + \text{proj}_{W^\perp}$ is the identity transformation.

It is easy to prove a,b,c,d,e using the formula, but we can also prove them from the existence and uniqueness of the orthogonal decomposition, e.g. to see a: if we have orthogonal decompositions $\mathbf{y}_1 = \text{proj}_W(\mathbf{y}_1) + \mathbf{z}_1$ and $\mathbf{y}_2 = \text{proj}_W(\mathbf{y}_2) + \mathbf{z}_2$, then

$$\begin{aligned} c\mathbf{y}_1 + d\mathbf{y}_2 &= c(\text{proj}_W(\mathbf{y}_1) + \mathbf{z}_1) + d(\text{proj}_W(\mathbf{y}_2) + \mathbf{z}_2) \\ &= \underbrace{c\,\text{proj}_W(\mathbf{y}_1) + d\,\text{proj}_W(\mathbf{y}_2)}_{\text{in } W} + \underbrace{c\mathbf{z}_1 + d\mathbf{z}_2}_{\text{in } W^\perp} \end{aligned}$$

Since the orthogonal decomposition is unique, this shows
$\text{proj}_W(c\mathbf{y}_1 + d\mathbf{y}_2) = c\,\text{proj}_W(\mathbf{y}_1) + d\,\text{proj}_W(\mathbf{y}_2)$.

Properties of the function $\text{proj}_W : \mathbb{R}^n \to \mathbb{R}^n$:

a. proj_W is a linear transformation.
b. $\text{proj}_W(\mathbf{y}) = \mathbf{y}$ if and only if $\mathbf{y}$ is in W.
c. The range of proj_W is W.
d. The kernel of proj_W is $W^\perp$.
e. $\text{proj}_W^2 = \text{proj}_W$.
f. $\text{proj}_W + \text{proj}_{W^\perp}$ is the identity transformation.

To see f: Write U for $W^\perp$. Then,

$$\mathbf{y} = \underbrace{\hat{\mathbf{y}}}_{\text{in } W = U^\perp} + \underbrace{\mathbf{z}}_{\text{in } W^\perp = U}.$$

By uniqueness of the orthogonal decomposition, $\mathbf{z} = \text{proj}_U(\mathbf{y})$. So $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} = \text{proj}_W(\mathbf{y}) + \text{proj}_{W^\perp}(\mathbf{y})$ for each $\mathbf{y}$ in $\mathbb{R}^n$, so $\text{proj}_W + \text{proj}_{W^\perp}$ is the identity transformation.

The orthogonal projection is a linear transformation, so we can ask for its standard matrix. (It is faster to compute orthogonal projections by taking dot products (formula on p9) than using the standard matrix, but this result is useful theoretically.)

Theorem 10: Matrix for Orthogonal Projection: Let $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ be an orthonormal basis for a subspace W, and U be the matrix $U = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_p \\ | & & | \end{bmatrix}$.

Then the standard matrix for proj_W is $[\text{proj}_W]_\mathcal{E} = UU^T$.

Proof:

$$UU^T\mathbf{y} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_p \\ | & & | \end{bmatrix} \begin{bmatrix} -- & \mathbf{u}_1 & -- \\ -- & \vdots & -- \\ -- & \mathbf{u}_p & -- \end{bmatrix} \mathbf{y} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_p \\ | & & | \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{y} \\ \vdots \\ \mathbf{u}_p \cdot \mathbf{y} \end{bmatrix}$$
$$= (\mathbf{u}_1 \cdot \mathbf{y})\mathbf{u}_1 + \dots + (\mathbf{u}_p \cdot \mathbf{y})\mathbf{u}_p.$$

Tip: to remember that $[\text{proj}_W]_\mathcal{E} = UU^T$ and not U^TU (which is important too, see p21), make sure this matrix is $n \times n$.

§6.4: The Gram-Schmidt Process

This is an algorithm to make an orthogonal basis out of an arbitrary basis.

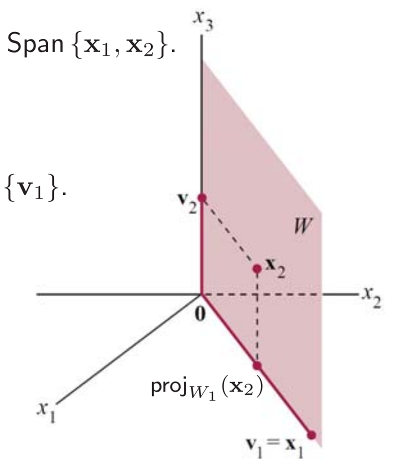
Example: Let $\mathbf{x}_1 = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$ and let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$.

Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

Answer: Let $\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$, and let $W_1 = \text{Span}\{\mathbf{v}_1\}$.

By the Orthogonal Decomposition Theorem,
$\mathbf{x}_2 - \text{proj}_{W_1}(\mathbf{x}_2)$ is orthogonal to W_1.
So let $\mathbf{v}_2 = \mathbf{x}_2 - \text{proj}_{W_1}(\mathbf{x}_2) = \mathbf{x}_2 - \dfrac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$

$$= \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \frac{8+2+0}{4^2+2^2+0} \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}.$$



For subspaces of dimension $p>2$, we repeat this idea p times, like this:

EXAMPLE Let $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 8 \\ 5 \\ -6 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} -6 \\ 7 \\ 2 \\ 1 \end{bmatrix}$, and suppose $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a basis for a subspace W of $\mathbf{R}^4$. Construct an orthogonal basis for W.

Solution:

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{W}_1 = \mathrm{Span}\{\mathbf{v}_1\}.$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{Proj}_{\mathbf{W}_1}(\mathbf{x}_2) = \underline{\qquad\qquad} = \begin{bmatrix} 8 \\ 5 \\ -6 \\ 0 \end{bmatrix} - \underline{\qquad\qquad} \begin{bmatrix} 3 \\ 0 \\ -1 \\ 0 \end{bmatrix} =$$

Check our answer so far:

Let $\mathbf{W}_2 = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ $\underline{\qquad\qquad}$

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathrm{Proj}_{\mathbf{W}_2}(\mathbf{x}_3) =$$

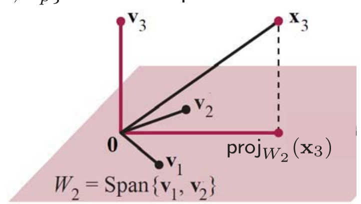
$$= \begin{bmatrix} -6 \\ 7 \\ 2 \\ 1 \end{bmatrix} - \underline{\qquad\qquad} \begin{bmatrix} 3 \\ 0 \\ -1 \\ 0 \end{bmatrix} - \underline{\qquad\qquad} \begin{bmatrix} 1 \\ -5 \\ 3 \\ 0 \end{bmatrix} =$$

Check our answer:

In general:

Theorem 11: Gram-Schmidt: Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_1 = \mathbf{x}_1$$
$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$
$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$
$$\vdots$$
$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$



Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W, and $\mathrm{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ for each k between 1 and p.

In fact, you can apply this algorithm to any spanning set (not necessarily linearly independent). Then some $\mathbf{v}_k$s might be zero, and you simply remove them.

pp361-362: Matrices with orthonormal columns

This is an important class of matrices.

Theorem 6: Matrices with Orthonormal Columns: A matrix U has orthonormal columns if and only if $U^TU = I$.

Proof: Let $\mathbf{u}_i$ denote the ith column of U. From the row-column rule of matrix multiplication (week 11 p14):

$$\begin{bmatrix} -- & \mathbf{u}_1 & -- \\ -- & \vdots & -- \\ -- & \mathbf{u}_p & -- \end{bmatrix} \begin{bmatrix} | & | & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_p \\ | & | & | \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \dots & \mathbf{u}_1 \cdot \mathbf{u}_p \\ \vdots & & \vdots \\ \mathbf{u}_p \cdot \mathbf{u}_1 & \dots & \mathbf{u}_p \cdot \mathbf{u}_p \end{bmatrix}.$$

so $U^TU = I$ if and only if $\mathbf{u}_i \cdot \mathbf{u}_i = 1$ for each i (diagonal entries), and $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for each pair $i \neq j$ (non-diagonal entries).

Theorem 7: Matrices with Orthonormal Columns represent Length-Preserving Linear Transformations: Let U be an $m \times n$ matrix with orthonormal columns. Then, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}.$$

In particular, $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$, and $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

Proof:

$$(U\mathbf{x}) \cdot (U\mathbf{y}) = (U\mathbf{x})^T(U\mathbf{y}) = \mathbf{x}^TU^TU\mathbf{y} = \mathbf{x}^T\mathbf{y} = \mathbf{x} \cdot \mathbf{y}.$$

because $U^TU = I_n$, by the previous theorem

Length-preserving linear transformations are sometimes called isometries.
Exercise: prove that an isometry also preserves angles; that is, if $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$, then $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y}$. (Hint: think about $\mathbf{x} + \mathbf{y}$.)

An important special case:
Definition: A matrix U is *orthogonal* if it is a square matrix with orthonormal columns. Equivalently, $U^{-1} = U^T$.

Warning: An orthogonal matrix has orthonormal columns, not simply orthogonal columns.

Example: The standard matrix of a rotation in $\mathbb{R}^2$ is $U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, and this is an orthogonal matrix because

$$U^TU = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

It can be shown that every orthogonal 2×2 matrix U represents either a rotation (if $\det U = 1$) or a reflection (if $\det U = -1$). (Exercise: why are these the only possible values of $\det U$?) An orthogonal $n \times n$ matrix with determinant 1 is a high-dimensional generalisation of a rotation.

Recall (week 9 p7, §4.4) that, if $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$, then the change-of-coordinates matrix from $\mathcal{B}$-coordinates to standard coordinates is

$$\underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}} = \begin{bmatrix} | & | & | \\ \mathbf{u}_1 & \dots & \mathbf{u}_n \\ | & | & | \end{bmatrix}.$$

So an orthogonal matrix can also be viewed as a change-of-coordinates matrix from an orthonormal basis to the standard basis.
Now the change-of-coordinates matrix from the standard basis to the basis $\mathcal{B}$ is $\underset{\mathcal{B}\leftarrow\mathcal{E}}{\mathcal{P}} = \underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}}^{-1}$. So if $\underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}} = U$ is an orthogonal matrix, then $U^{-1} = U^T$ so, for an orthonormal basis $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathbb{R}^n$, we have

$$[\mathbf{x}]_{\mathcal{B}} = U^T\mathbf{x} = \begin{bmatrix} -- & \mathbf{u}_1 & -- \\ -- & \vdots & -- \\ -- & \mathbf{u}_n & -- \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{u}_n \cdot \mathbf{x} \end{bmatrix}.$$

Remembering the definition of coordinates, this says
$\mathbf{x} = (\mathbf{u}_1 \cdot \mathbf{x})\mathbf{u}_1 + \dots + (\mathbf{u}_n \cdot \mathbf{x})\mathbf{u}_n$, as in the Weights for Orthogonal Bases Theorem.

Non-examinable: distances for abstract vector spaces

On an abstract vector space, a function that takes two vectors to a scalar satisfying the symmetry, linearity and positivity properties (week 11 p5) is called an inner product. The inner product of $\mathbf{u}$ and $\mathbf{v}$ is often written $\langle \mathbf{u}, \mathbf{v} \rangle$ or $\langle \mathbf{u} | \mathbf{v} \rangle$. (So the dot product is one example of an inner product on $\mathbb{R}^n$, but other useful inner products exist; these can be used to compute weighted regression lines, see §6.8 of the textbook)

Many common inner products on $C([0,1])$, the vector space of continuous functions, have the form

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^1 f(t)g(t)w(t)\,dt$$

for some weight function $w(t)$. This inner product can be used to find polynomial approximations and Fourier approximations to functions, see §6.7-6.8 of the textbook.

Applying Gram-Schmidt to $\{1, t, t^2, \dots\}$ produces various families of orthogonal polynomials, which is a big field of study.

With the concept of inner product, we can understand what the transpose means for linear transformations:

First notice: if A is an $m \times n$ matrix, then, for all $\mathbf{v}$ in $\mathbb{R}^n$ and all $\mathbf{u}$ in $\mathbb{R}^m$:

$$\underbrace{(A^T\mathbf{u}) \cdot \mathbf{v}}_{\text{dot product in } \mathbb{R}^n} = (A^T\mathbf{u})^T\mathbf{v} = \mathbf{u}^T A\mathbf{v} = \underbrace{\mathbf{u} \cdot (A\mathbf{v})}_{\text{dot product in } \mathbb{R}^m}.$$

So, if A is the standard matrix of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, then A^T is the standard matrix of its adjoint $T^* : \mathbb{R}^m \to \mathbb{R}^n$, which satisfies

$$(T^*\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (T\mathbf{v}).$$

or, for abstract vector space with an inner product:

$$\langle T^*\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, T\mathbf{v} \rangle.$$

So symmetric matrices ($A^T = A$) represent self-adjoint linear transformations ($T^* = T$). For example, on $C([0,1])$ with any integral inner product, the multiplication-by-x function $\mathbf{f} \mapsto x\mathbf{f}$ is self-adjoint.