Lecture 10: Vector Algebra: Orthogonal Basis

• Orthogonal Basis of a subspace
• Computing an orthogonal basis for a subspace using Gram-Schmidt Orthogonalization Process
Orthogonal Set

• Any set of vectors that are mutually orthogonal, is a an orthogonal set.

**EXAMPLE 1** Show that \( \{ \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \} \) is an orthogonal set, where

\[
\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}
\]

**SOLUTION** Consider the three possible pairs of distinct vectors, namely, \( \{ \mathbf{u}_1, \mathbf{u}_2 \} \), \( \{ \mathbf{u}_1, \mathbf{u}_3 \} \), and \( \{ \mathbf{u}_2, \mathbf{u}_3 \} \).

\[
\mathbf{u}_1 \cdot \mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0
\]
\[
\mathbf{u}_1 \cdot \mathbf{u}_3 = 3 \left( -\frac{1}{2} \right) + 1(-2) + 1 \left( \frac{7}{2} \right) = 0
\]
\[
\mathbf{u}_2 \cdot \mathbf{u}_3 = -1 \left( -\frac{1}{2} \right) + 2(-2) + 1 \left( \frac{7}{2} \right) = 0
\]

Each pair of distinct vectors is orthogonal, and so \( \{ \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \} \) is an orthogonal set.
Orthonormal Set

• Any set of unit vectors that are mutually orthogonal, is a an **orthonormal** set.
• In other words, any orthogonal set is an orthonormal set if all the vectors in the set are unit vectors.
• Example: \(\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}\) is an orthonormal set, where,

\[
\mathbf{u}_1 = \begin{bmatrix} \frac{3}{\sqrt{11}} \\ 1 \\ \frac{1}{\sqrt{11}} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ 2 \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{66}} \\ 4 \\ -\frac{1}{\sqrt{66}} \end{bmatrix}
\]
An orthogonal set is Linearly Independent

If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then $S$ is linearly independent and hence is a basis for the subspace spanned by $S$.

**PROOF** If $0 = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \ldots, c_p$, then

$$0 = 0 \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1$$
$$= (c_1\mathbf{u}_1) \cdot \mathbf{u}_1 + (c_2\mathbf{u}_2) \cdot \mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p) \cdot \mathbf{u}_1$$
$$= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{u}_1)$$
$$= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$. Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1 \cdot \mathbf{u}_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \ldots, c_p$ must be zero. Thus $S$ is linearly independent.
Projection of vector \( \mathbf{b} \) on vector \( \mathbf{a} \)

- \( \mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta \)
- Vector \( \mathbf{c} \) is the image/perpendicular projection of \( \mathbf{b} \) on \( \mathbf{a} \)
- Direction of \( \mathbf{c} \) is the same as \( \mathbf{a} \)
- Magnitude of \( \mathbf{c} \) is \( \|\mathbf{c}\| = \|\mathbf{b}\| \cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|} \)

\[ \|\mathbf{c}\| = \hat{\mathbf{a}} \cdot \mathbf{b} \]

- If \( \hat{\mathbf{a}} \) is the unit vector of \( \mathbf{a} \), then
- vector \( \mathbf{c} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|} \hat{\mathbf{a}} = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\|} \frac{\mathbf{a}}{\|\mathbf{a}\|} = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a} \)
Orthogonal Basis

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

• Example:
  \[
  \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}
  \]
  is basically the $x$, $y$, and $z$ axis. It is an orthogonal basis in $\mathbb{R}^3$, and it spans the whole $\mathbb{R}^3$ space. It is also an orthogonal set.
Orthogonal Basis

• We know that given a basis of a subspace, any vector in that subspace will be a linear combination of the basis vectors.

• For example, if \( \mathbf{u} \) and \( \mathbf{v} \) are linearly independent and form the basis for a subspace \( S \), then any vector \( \mathbf{y} \) in \( S \) can be expressed as:

\[
\mathbf{y} = c_1 \mathbf{u} + c_2 \mathbf{v}
\]

• But computing \( c_1 \) and \( c_2 \) is not straight forward.

• On the other hand, if \( \mathbf{u} \) and \( \mathbf{v} \) form an orthogonal basis, then

\[
c_1 = \frac{\mathbf{y} \cdot \mathbf{u}}{\|\mathbf{u}\|} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \quad \text{and} \quad c_2 = \frac{\mathbf{y} \cdot \mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{y} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}
\]
Does not work if it is not Orthogonal basis

• \( y = c_1 u + c_2 v \)

• But computing \( c_1 \) and \( c_2 \) is not straight forward (yet).

What is computed is,

• \( d_1 = \frac{y \cdot u}{\|u\|} = \frac{y \cdot u}{u \cdot u} \)

• \( d_2 = \frac{y \cdot v}{\|v\|} = \frac{y \cdot v}{v \cdot v} \)
Orthogonal Decomposition Theorem

Given a vector \( u \) in \( \mathbb{R}^n \), consider the problem of decomposing a vector \( y \) in \( \mathbb{R}^n \) into two components:

\[ y = \hat{y} + z \]

where \( \hat{y} \) is in \( \text{span}\{u\} \) and \( z \) is orthogonal to \( u \). \( \hat{y} \) is called the orthogonal projection of \( y \) onto \( u \).
Orthogonal Decomposition Theorem

- Decompose $y = \hat{y} + z$
- Let $\hat{y} = \alpha u$ for some scalar $\alpha$. Then
  \[
z = y - \hat{y} = y - \alpha u\]
- Since $z \cdot u = 0$, then
  \[
u^T(y - \alpha u) = 0\]
- so we have $u^Ty = \alpha u^Tu$, and
  \[
  \alpha = \frac{y^Tu}{u^Tu}\]
- $\text{Proj}_u(y) = \hat{y} = \alpha u = \frac{y^Tu}{u^Tu}u$
Orthogonal Decomposition Theorem

Let $W = \text{span}\{u_1, \ldots, u_p\}$ is a subspace of $R^n$, where $\{u_1, \ldots, u_p\}$ is an orthogonal set. Decompose $y$ into two components:

$$y = \hat{y} + z$$

where $\hat{y}$ is a vector in $W$ and $z$ is orthogonal to $W$. $\hat{y}$ is called the orthogonal projection of $y$ onto $W$.

- Since $\hat{y}$ is in $W$, write

$$\hat{y} = c_1 u_1 + c_2 u_2 + \ldots + c_p u_p$$

- $z = y - \hat{y}$ is orthogonal to $W$, implying that $u_i \cdot z = 0$ for every $i$.

- From $u_i^T (y - \hat{y}) = 0 \implies c_i = \frac{u_i^T y}{u_i^T u_i}$

- So the orthogonal project of $y$ onto $W$ is

$$\text{Proj}_W(y) = \hat{y} = \frac{u_1^T y}{u_1^T u_1} u_1 + \ldots + \frac{u_p^T y}{u_p^T u_p} u_p$$
Orthogonal Decomposition Theorem

Example

Let $u_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $u_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Find the orthogonal projection of $y$ onto $W = \text{Span}\{u_1, u_2\}$. 
Projection of a vector on a Subspace

- \( \mathbf{u} \) and \( \mathbf{v} \) are orthogonal 3D vectors.
- They span a plane (green plane) in 3D
- \( \mathbf{y} \) is an arbitrary 3D vector out of the plane.
- \( \mathbf{y}' \) is the projection of \( \mathbf{y} \) onto the plane.
- \( \mathbf{y}' = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} + \frac{\mathbf{y} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v} \)
- The “point” \( \mathbf{y}' \) is also the closest point to \( \mathbf{y} \) on the plane.
- \( \mathbf{y} - \mathbf{y}' \) is perpendicular to \( \mathbf{y}' \), \( \text{Span}\{\mathbf{u}, \mathbf{v}\} \), and hence \( \mathbf{u} \) and \( \mathbf{v} \)
Closest point of a vector to a span does not depend on the basis of that span

\( \mathbf{y} \) is an arbitrary 3D vector out of the plane.
• \( \mathbf{y}' \) is the projection of \( \mathbf{y} \) onto the plane.
• \( \mathbf{y}' = x_1 \mathbf{u}_1 + x_2 \mathbf{u}_2 \)
• \( \mathbf{y}' = x_3 \mathbf{u}_3 + x_4 \mathbf{u}_4 \)
• \( \mathbf{y}' = x_5 \mathbf{u}_1 + x_6 \mathbf{u}_3 \)
• ...
• Coordinates of \( \mathbf{y}' \) change if the basis changes.
• But the vector \( \mathbf{y}' \) itself does not change.
• Hence the **closest point** to \( \mathbf{y} \) on the plane does not change.

Even here, \( \mathbf{y} - \mathbf{y}' \) is perpendicular to \( \mathbf{y}' \) and \( \text{Span}\{., .\} \)
Closest point of a vector to a span does not depend on the basis of that span

• FINDING THE CLOSEST POINT OF A VECTOR TO A SPAN means:

“Finding the coordinates of the projection of the vector”

• So, if you want to compute the closest point of a vector to a span, then find an appropriate basis with which you can compute the coordinates of the projection easily.

What would be that basis?
Answer: An orthogonal basis!
Projection on a span of non-orthogonal vectors

• How to find projection of any arbitrary 3D vector onto the span of two non-orthogonal, linearly independent vectors?
• $\mathbf{u}_1$ and $\mathbf{u}_2$ are not orthogonal, but linearly independent vectors in 3D.
• $\mathbf{y}$ is an arbitrary 3D vector.
• Find the projection of $\mathbf{y}$ in the space spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$.
  • a) First, find the orthogonal set of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ that span the same subspace as $\mathbf{u}_1$ and $\mathbf{u}_2$. In other words, find an orthogonal basis.
  • b) Project $\mathbf{y}$ onto the space spanned by orthogonal $\mathbf{v}_1$ and $\mathbf{v}_2$ vectors, as we earlier.
How to find an orthogonal basis?

Given a basis \( \{x_1, \ldots, x_p\} \) for a subspace \( W \) of \( \mathbb{R}^n \), find an orthogonal basis \( \{v_1, \ldots, v_p\} \) for \( W \) such that for any \( i = 1, \ldots, p \)

\[
\text{span}\{v_1, \ldots, v_i\} = \text{span}\{x_1, \ldots, x_i\}
\]
How to find an orthogonal basis?

• Assume that the first vector $u_1$ is in the orthogonal basis. Other vector(s) of the basis are computed that are perpendicular to $u_1$

• Let $v_1 = u_1$

• Let $v_2 = u_2 - \frac{u_2v_1}{v_1v_1} v_1$

• We know that $v_2$ is perpendicular to $v_1$.

• $v_2$ is in the $Span\{u_1, u_2\}$ (Why?)

• So $Span\{u_1, u_2\} = Span\{v_1, v_2\}$

• And $\{v_1, v_2\}$ is an orthogonal basis

• Projection of $y$ on to the $Span\{u_1, u_2\}$

\[ y' = \frac{y \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{y \cdot v_2}{v_2 \cdot v_2} v_2 \]
Gram-Schmidt Orthogonalization Process

Given a basis \( \{x_1, \ldots, x_p\} \) for a subspace \( W \) of \( \mathbb{R}^n \), find an orthogonal basis \( \{v_1, \ldots, v_p\} \) for \( W \) such that for any \( i = 1, \ldots, p \)

\[
\text{span}\{v_1, \ldots, v_i\} = \text{span}\{x_1, \ldots, x_i\}
\]

1. \( v_1 = x_1 \)
2. \( v_2 = x_2 - \text{Proj}_{v_1}(x_2) = x_2 - \frac{x_2^T v_1}{v_1^T v_1} v_1 \)
3. \( v_3 = x_3 - \text{Proj}_{\text{span}\{v_1, v_2\}}(x_3) = x_3 - \frac{x_3^T v_1}{v_1^T v_1} v_1 - \frac{x_3^T v_2}{v_2^T v_2} v_2 \)
4. \( \vdots \)
5. \( v_p = x_p - \text{Proj}_{\text{span}\{v_1, \ldots, v_{p-1}\}}(x_p) \)
   \[
   = x_p - \frac{x_p^T v_1}{v_1^T v_1} v_1 - \frac{x_p^T v_2}{v_2^T v_2} v_2 - \ldots - \frac{x_p^T v_{p-1}}{v_{p-1}^T v_{p-1}} v_{p-1}
   \]
Gram-Schmidt Orthogonalization Process

Example:

Let \( x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \) and \( x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \). Construct an orthogonal basis for \( W = \text{Span}\{x_1, x_2, x_3\} \).
Solving Inconsistent Systems

- Suppose $Ax = b$ has no solutions. Can we still find a solution $x$ such that $Ax$ is “closest” to $b$?

- Most common cases: $A$ is an $m \times n$ matrix with $m > n$. The system $Ax = b$ has more equations than variables. So in general there is no solution.

- “Best solution” in the following sense: Find $\hat{x}$ such that $A\hat{x}$ is the closest point to $b$. That is,

  $$\|A\hat{x} - b\| \leq \|Ax - b\|$$

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

- $\hat{x}$ is called the least square solution.
Solving Inconsistent Systems

- Problem: Find $\hat{x}$ such that $A\hat{x}$ is closest to $b$.
- The problem is equivalent to finding a point $\hat{b}$ in $\text{Col } A$ that is closest to $b$.
- From the best approximation theorem, the point in $\text{Col } A$ closest to $b$ is the orthogonal projection of $b$ onto $\text{Col } A$:

$$A\hat{x} = \hat{b} = \text{proj}_{\text{Col } A} b$$
Solving Inconsistent Systems

Example 1:
A trader buys and/or sells tomatoes and potatoes. (Negative number means buys, positive number means sells.) In the process, he either makes profit (positive number) or loss (negative number). A week’s transaction is shown; find the approximate cost of tomatoes and potatoes.

<table>
<thead>
<tr>
<th>Tomatoes (tons)</th>
<th>Potatoes (tons)</th>
<th>Profit/Loss (in thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-6</td>
<td>-1</td>
</tr>
<tr>
<td>1</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>6</td>
</tr>
</tbody>
</table>
Solving Inconsistent Systems

• $1t - 6p = -1$
• $1t - 2p = 2$
• $1t + 1p = 1$
• $1t + 7p = 6$

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} t + \begin{bmatrix} -6 \\ -2 \\ 1 \end{bmatrix} p = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

The above equation might not have a solution (values of t and p that would satisfy that equation). So the best we can do is to find the values of t and p that would result in a vector on the right hand side that is as close as possible to the desired right hand side vector.
Solving Inconsistent Systems

Example 2:

**EXAMPLE 4** Find a least-squares solution of \( Ax = b \) for

\[
A = \begin{bmatrix}
1 & -6 \\
1 & -2 \\
1 & 1 \\
1 & 7
\end{bmatrix}, \quad b = \begin{bmatrix}
-1 \\
2 \\
1 \\
6
\end{bmatrix}
\]

**SOLUTION** Because the columns \( a_1 \) and \( a_2 \) of \( A \) are orthogonal, the orthogonal projection of \( b \) onto \( \text{Col} \, A \) is given by

\[
\hat{b} = \frac{b \cdot a_1}{a_1 \cdot a_1} a_1 + \frac{b \cdot a_2}{a_2 \cdot a_2} a_2 = \frac{8}{4} a_1 + \frac{45}{90} a_2
\]

\[
= \begin{bmatrix}
2 \\
2 \\
2 \\
2
\end{bmatrix} + \begin{bmatrix}
-3 \\
-1 \\
1/2 \\
7/2
\end{bmatrix} = \begin{bmatrix}
-1 \\
1 \\
5/2 \\
11/2
\end{bmatrix}
\]

Now that \( \hat{b} \) is known, we can solve \( A\hat{x} = \hat{b} \). But this is trivial, since we already know what weights to place on the columns of \( A \) to produce \( \hat{b} \). It is clear from (5) that

\[
\hat{x} = \begin{bmatrix}
8/4 \\
45/90
\end{bmatrix} = \begin{bmatrix}
2 \\
1/2
\end{bmatrix}
\]