Orthogonality
Orthonormal Bases, Orthogonal Matrices
The Major Ideas from Last Lecture

- Vector Span
The Major Ideas from Last Lecture

- Vector Span
- Subspace
The Major Ideas from Last Lecture

- Vector Span
- Subspace
- Basis Vectors
The Major Ideas from Last Lecture

- Vector Span
- Subspace
- Basis Vectors
- Coordinates in different bases
The Major Ideas from Last Lecture

- Vector Span
- Subspace
- Basis Vectors
- Coordinates in different bases
- Matrix Factorization (Basics)
The span of basis vectors forms a vector space or a subspace.
The span of basis vectors forms a vector space or a subspace.

Initially, we think about elementary basis vectors.
The span of basis vectors forms a vector space or a subspace.

Initially, we think about elementary basis vectors.

For example, in 3-space (think x-, y-, z-axis), The span of any two of the elementary basis vectors forms a 2-dimensional subspace.
We don’t always have to restrict ourselves to the elementary basis vectors.
The Major Ideas from Last Lecture

• We don’t always have to restrict ourselves to the elementary basis vectors.

• We’ll want to derive new basis vectors (new variables) that somehow reveal hidden patterns in the data or give us a better low dimensional approximation.
The Major Ideas from Last Lecture

- We don’t always have to restrict ourselves to the elementary basis vectors.
- We’ll want to derive new basis vectors (new variables) that somehow reveal hidden patterns in the data or give us a better low dimensional approximation.
- However, these new basis vectors (new variables) are just linear combinations of our elementary basis vectors (original variables).
The next concept we need is **orthonormality**. A collection of vectors is orthonormal if they are mutually orthogonal (orthogonal means perpendicular) and if every vector in the collection is a unit vector (i.e. has length 1: $\|x\|_2 = 1$).
The next concept we need is **orthonormality**. A collection of vectors is orthonormal if they are mutually orthogonal (orthogonal means perpendicular) and if every vector in the collection is a unit vector (i.e. has length 1: $\|x\|_2 = 1$).

Easiest example?
The next concept we need is orthonormality. A collection of vectors is orthonormal if they are mutually orthogonal (orthogonal means perpendicular) and if every vector in the collection is a unit vector (i.e. has length 1: $\|x\|_2 = 1$).

Easiest example?

The elementary basis vectors!
Angle between vectors

One way to measure how close vectors are to each other is to consider the angle between them. Do they point in the same direction? Opposite direction?

The cosine of the angle between two vectors is given by the inner product of their unit vectors:

$$\cos(\theta) = \frac{x^T y}{\|x\|_2 \|y\|_2}$$
One way to measure how close vectors are to each other is to consider the angle between them. Do they point in the same direction? Opposite direction?

The **cosine** of the angle between two vectors is given by the inner product of their unit vectors:

\[
\cos(\theta) = \frac{x^T y}{\|x\|_2 \|y\|_2}
\]
You may remember the formula for Pearson’s correlation coefficient:

\[ r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \]
Pearson’s Correlation

You may remember the formula for Pearson’s correlation coefficient:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

This is just the cosine between the *centered* variable vectors!

$$\cos(\theta) = \frac{x^T y}{\|x\|_2 \|y\|_2}$$
The cosine of the angle between two vectors is an extremely common measure of similarity in text mining. Hopefully we’ll get to talk more about that later!
So how do we know when two vectors are orthogonal (perpendicular)? When the angle between them is 90°. You may recall that $\cos(90°) = 0$.

$$
cos(\theta) = 0 \iff \frac{x^T y}{\|x\|_2 \|y\|_2} = 0 \iff x^T y = 0
$$
So how do we know when two vectors are orthogonal (perpendicular)? When the angle between them is $90^\circ$. You may recall that $\cos(90^\circ) = 0$.

$$\cos(\theta) = 0 \iff \frac{x^T y}{\|x\|_2 \|y\|_2} = 0 \iff x^T y = 0$$

Two vectors are orthogonal if and only if their inner product is zero.
So how do we know when two vectors are orthogonal (perpendicular)? When the angle between them is $90^\circ$. You may recall that $\cos(90^\circ) = 0$.

$$\cos(\theta) = 0 \iff \frac{x^T y}{\|x\|_2 \|y\|_2} = 0 \iff x^T y = 0$$

Two vectors are orthogonal if and only if their inner product is zero.

Orthogonal $\Rightarrow$ Completely Uncorrelated
So how do we know when two vectors are orthogonal (perpendicular)? When the angle between them is 90°. You may recall that $\cos(90°) = 0$.

$$\cos(\theta) = 0 \iff \frac{x^T y}{\|x\|_2 \|y\|_2} = 0 \iff x^T y = 0$$

Two vectors are orthogonal if and only if their inner product is zero.

Orthogonal $\Rightarrow$ Completely Uncorrelated

$\not\Rightarrow$ Independent
What is the cosine of the angle between \( x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \) and \( y = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \)?
Let’s Practice

1. What is the cosine of the angle between \( x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \) and \( y = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \)?

2. Are the vectors \( v_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \) and \( v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \) orthogonal? Are they orthonormal?
Let’s Practice

1. What is the cosine of the angle between \( x = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \) and \( y = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \)?

2. Are the vectors \( v_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \) and \( v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \) orthogonal? Are they orthonormal?

3. What are the two conditions necessary for a collection of vectors to be orthonormal?
Orthonormal Basis

If a set of basis vectors \( \{v_1, v_2, \ldots, v_p\} \) forms an orthonormal basis then it must be that:

1. \( v_i^T v_j = 0 \) when \( i \neq j \). (i.e. mutually orthogonal)
2. \( v_i^T v_i = 1 \) (i.e. each vector is a unit vector)
If a set of basis vectors \( \{v_1, v_2, \ldots, v_p\} \) forms an orthonormal basis then it must be that:

1. \( v_i^T v_j = 0 \) when \( i \neq j \). (i.e. mutually orthogonal)
Orthonormal Basis

If a set of basis vectors \( \{v_1, v_2, \ldots, v_p\} \) forms an orthonormal basis then it must be that:

1. \( v_i^T v_j = 0 \) when \( i \neq j \). (i.e. mutually orthogonal)
2. \( v_i^T v_i = 1 \) (i.e. each vector is a unit vector)
Suppose the columns of a matrix are mutually orthonormal:

\[ V = (v_1 | v_2 | \ldots | v_p) \]

Consider the matrix product \( V^T V \):
Suppose the columns of a matrix are mutually orthonormal:

\[ V = (v_1 | v_2 | \ldots | v_p) \]

Consider the matrix product \( V^T V \):

\[
V^T V = \begin{pmatrix}
    v_1^T \\
v_2^T \\
v_3^T \\
    \vdots \\
v_p^T
\end{pmatrix}
\begin{pmatrix}
    v_1 & v_2 & v_3 & \ldots & v_p
\end{pmatrix} =
\]
Suppose the columns of a matrix are mutually orthonormal:

\[ V = (v_1 \mid v_2 \mid \ldots \mid v_p) \]

Consider the matrix product \( V^T V \):

\[
V^T V = \begin{pmatrix}
  v_1^T \\
  v_2^T \\
  v_3^T \\
  \vdots \\
  v_p^T
\end{pmatrix}
\begin{pmatrix}
  v_1 \\
  v_2 \\
  v_3 \\
  \vdots \\
  v_p
\end{pmatrix}
= \begin{pmatrix}
  v_1^T v_1 & v_1^T v_2 & v_1^T v_3 & \cdots & v_1^T v_p \\
  v_2^T v_1 & v_2^T v_2 & v_2^T v_3 & \cdots & v_2^T v_p \\
  v_3^T v_1 & v_3^T v_2 & v_3^T v_3 & \cdots & v_3^T v_p \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  v_p^T v_1 & v_p^T v_2 & v_p^T v_3 & \cdots & v_p^T v_p
\end{pmatrix}
\]
Suppose the columns of a matrix are mutually orthonormal:

\[ V = (v_1 | v_2 | \ldots | v_p) \]

Consider the matrix product \( V^T V \):

\[
V^T V = \begin{pmatrix}
  v_1^T \\
  v_2^T \\
  v_3^T \\
  \vdots \\
  v_p^T
\end{pmatrix}
\begin{pmatrix}
  v_1 \\
  v_2 \\
  v_3 \\
  \vdots \\
  v_p
\end{pmatrix} =
\begin{pmatrix}
  1 & 0 & 0 & \ldots & 0 \\
  0 & 1 & 0 & \ldots & 0 \\
  0 & 0 & 1 & \ldots & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & 0 & \ldots & 1
\end{pmatrix} = I_p
\]
Suppose the columns of a matrix are mutually orthonormal:

\[ V = (v_1 | v_2 | \ldots | v_p) \]

Consider the matrix product \( V^T V \):

\[
V^T V = \begin{pmatrix}
    v_1^T \\
v_2^T \\
v_3^T \\
\vdots \\
v_p^T
\end{pmatrix}
\begin{pmatrix}
v_1 \\
v_2 \\
v_3 \\
\vdots \\
v_p
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1
\end{pmatrix} = I_p
\]

However, we can’t say anything about \( VV^T \) unless the matrix is square.
When a **square** matrix $U$ has orthonormal columns, it also has orthonormal rows. Such a matrix is called an **orthogonal matrix**. Its inverse is equal to its transpose:

$$U^T U = UU^T = I \quad \implies U^{-1} = U^T$$

So to solve the system $Ux = b$, simply multiply by $U^T$.

Nice and Easy!

**Orthogonality**
When a square matrix $U$ has orthonormal columns, it also has orthonormal rows. Such a matrix is called an orthogonal matrix. Its inverse is equal to its transpose:

$$U^T U = U U^T = I \quad \implies U^{-1} = U^T$$

So to solve the system $Ux = b$, simply multiply by $U^T$.

$$x = U^T b$$
When a square matrix $U$ has orthonormal columns, it also has orthonormal rows. Such a matrix is called an orthogonal matrix. Its inverse is equal to its transpose:

$$U^T U = U U^T = I \quad \implies U^{-1} = U^T$$

So to solve the system $Ux = b$, simply multiply by $U^T$.

$$x = U^T b$$

Nice and Easy!
Why an Orthonormal Basis?

Two Conditions $\Rightarrow$ Two Reasons

- The basis vectors are mutually perpendicular. They are just rotations of the elementary basis vectors. Even in the new basis, we can still plot our data coordinates and examine them in a familiar way. Anything else would just be weird.

- The basis vectors have length 1. So when we look at the coordinates associated with each basis vectors, they tell us how many units to go in each direction. This way, we can essentially ignore the basis vectors and focus on the coordinates alone. We truly just rotated the axes.
Why an Orthonormal Basis?

Two Conditions ⇒ Two Reasons

- The basis vectors are mutually perpendicular. They are just rotations of the elementary basis vectors. Even in the new basis, we can still plot our data coordinates and examine them in a familiar way. Anything else would just be weird.

- The basis vectors have length 1. So when we look at the coordinates associated with each basis vector, they tell us how many *units* to go in each direction. This way, we can essentially ignore the basis vectors and focus on the coordinates alone. We truly just rotated the axes.
The basis vectors have length 1. So when we look at the coordinates associated with each basis vectors, they tell us how many *units* to go in each *direction*.
Example: Orthogonal but not orthonormal basis.

\[
\begin{align*}
\mathbf{v}_1 &= \begin{pmatrix} 10 \\ 10 \end{pmatrix} \\
\mathbf{v}_2 &= \begin{pmatrix} -1 \\ 1 \end{pmatrix}
\end{align*}
\]
Why an Orthonormal Basis?

What is the problem with the new coordinates?

\[
\begin{pmatrix}
2 \\
3
\end{pmatrix} = 0.25v_1 + 0.5v_2
\]
Why an Orthonormal Basis?

What is the problem with the new coordinates? Can I ignore the basis vectors and draw the point on a coordinate grid?
Why an Orthonormal Basis?

What is the problem with the new coordinates? Can I ignore the basis vectors and draw the point on a coordinate grid?

Is the point the same distance from the origin as it was before?
Why an Orthonormal Basis?

What is the problem with the new coordinates? Can I ignore the basis vectors and draw the point on a coordinate grid?

Is the point the same distance from the origin as it was before? Is it sensible that the coordinate along the direction $v_2$ is twice the coordinate along the direction $v_1$?
A basis which is not orthonormal will distort the data, while an orthonormal basis will merely rotate the data.
A basis which is not orthonormal will distort the data, while an orthonormal basis will merely rotate the data.

95% of dimension reduction methods create a new orthonormal basis for the data.

- Principal Components Analysis
- Singular Value Decomposition
- Factor Analysis
- Correspondence Analysis
Let’s Practice

Let

\[
U = \frac{1}{3} \begin{pmatrix}
-1 & 2 & 0 & -2 \\
2 & 2 & 0 & 1 \\
0 & 0 & 3 & 0 \\
-2 & 1 & 0 & 2
\end{pmatrix}
\]

a. Show that \( U \) is an orthogonal matrix.
Let’s Practice

Let

\[ U = \frac{1}{3} \begin{pmatrix} -1 & 2 & 0 & -2 \\ 2 & 2 & 0 & 1 \\ 0 & 0 & 3 & 0 \\ -2 & 1 & 0 & 2 \end{pmatrix} \]

a. Show that \( U \) is an orthogonal matrix.

b. Let \( b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \). Solve the equation \( Ux = b \).
1. Let

\[ U = \frac{1}{3} \begin{pmatrix} -1 & 2 & 0 & -2 \\ 2 & 2 & 0 & 1 \\ 0 & 0 & 3 & 0 \\ -2 & 1 & 0 & 2 \end{pmatrix} \]

a. Show that \( U \) is an orthogonal matrix.

b. Let \( b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \). Solve the equation \( Ux = b \).

2. Find two vectors which are orthogonal to \( x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \).
We don’t always have to restrict ourselves to the elementary basis vectors.
We don’t always have to restrict ourselves to the elementary basis vectors.

We’ll want to derive new basis vectors (new variables) that somehow reveal hidden patterns in the data or give us a better low dimensional approximation.
We don’t always have to restrict ourselves to the elementary basis vectors.

We’ll want to derive new basis vectors (new variables) that somehow reveal hidden patterns in the data or give us a better low dimensional approximation.

However, these new basis vectors (new variables) are just linear combinations of our elementary basis vectors (original variables).
Our (elementary) basis vectors are essentially our variable units. Suppose we have a plot of household income in thousands on the x-axis and SAT scores on the y-axis.

**Original Data**

- **SAT Scores**
  - 500
  - 1000
  - 1500

- **Household Income in Thousands**
  - 200
  - 400
  - 600
  - 800
  - 1000
Our (elementary) basis vectors are essentially our variable units. Suppose we have a plot of household income in thousands on the x-axis and SAT scores on the y-axis.

Each point (vector) can be written as a linear combination of household income in thousands ($e_1$) and SAT score ($e_2$). The benefit of this “vector space model”? 
Connection to Data

Our (elementary) basis vectors are essentially our variable units. Suppose we have a plot of household income in thousands on the x-axis and SAT scores on the y-axis.

Each point \((vector)\) can be written as a linear combination of household income in thousands \((e_1)\) and SAT score \((e_2)\). The benefit of this “vector space model”? **Interpretability**.
Our (elementary) basis vectors are essentially our variable units. Suppose we have a plot of household income in thousands on the x-axis and SAT scores on the y-axis.

Are there any drawbacks? It Depends!
Think about taking a giant flashlight and shining all of the data onto a subspace at a right angle.
Orthogonal Projections

Think about taking a giant flashlight and shining all of the data onto a subspace at a right angle.

The resulting shadow would be the **orthogonal projection** of the data onto the subspace.

*(Orthogonal just means Perpendicular!)*
Orthogonal Projections/Components

Easy to think about projection onto the elementary axes!

$\text{span}(e_2)$

$\text{span}(e_1)$

orthogonal projection of $a$ onto $e_2$

orthogonal projection of $a$ onto $e_1$
When we “drop” a numeric variable, this is essentially what is happening. We had 2-dimensional data. We now have 1-dimensional data. We’ve projected the data onto one of the basis vectors.
When we “drop” a numeric variable, this is essentially what is happening. We had 2-dimensional data. We now have 1-dimensional data. We’ve projected the data onto one of the basis vectors (axes/variables).
When we “drop” a numeric variable, this is essentially what is happening. We had 2-dimensional data. We now have 1-dimensional data. We’ve projected the data onto one of the basis vectors (axes/variables).
Changing the Basis

What if we used a different set of axes? (i.e. new basis vectors...)
Rotation of same image, using new axes (variables/basis vectors)
Orthogonal Projections onto New Axis

Rotated Data: Projection onto NEW Horizontal axis
Let’s Practice

1. Draw the orthogonal projection of the points onto the subspace \( \text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \)
The Major Ideas from Today

Cosine / Angle between vectors
Orthonormality
Orthonormal basis
Orthogonal matrix
Orthogonal projections