Orthogonal Projections and Orthonormal Bases

1. Student Learning Objectives
   - SLOs: Orthogonal Projections and Orthonormal Bases

2. Orthogonality and Least Squares
   - Orthogonal Projections and Orthonormal Bases

3. Why Orthonormality Matters
   - Application: The (Fast) Fourier Transform

4. Suggested Problems
   - Suggested Problems 5.1
   - Lecture–Book Roadmap

Definition (Orthogonality, Length, Unit Vectors) (old)

- Two vectors \( \vec{v} \) and \( \vec{w} \in \mathbb{R}^n \) are called orthogonal (or perpendicular) if \( \vec{v} \cdot \vec{w} = 0 \).
- The length (or norm, or magnitude) of a vector \( \vec{v} \in \mathbb{R}^n \) is \( \| \vec{v} \| = \sqrt{\vec{v} \cdot \vec{v}} \).
- A vector \( \vec{u} \in \mathbb{R}^n \) is called a unit vector if \( \| \vec{u} \| = 1 \).

Definition (Orthonormal Vectors) (new)

The vectors \( \vec{u}_1, \ldots, \vec{u}_m \in \mathbb{R}^n \) are called orthonormal if they are all unit vectors and orthogonal to one another:

\[
\vec{u}_i \cdot \vec{u}_j = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{if } i \neq j 
\end{cases}
\]
Orthogonal Projections and Orthonormal Bases

**Example (The “Standard Basis Vectors”)**
The vectors \( \vec{e}_1, \ldots, \vec{e}_n \in \mathbb{R}^n \) are orthonormal. — They form an orthonormal basis for \( \mathbb{R}^n \).

**Example (Rotated Standard Vectors in \( \mathbb{R}^2 \))**
Consider \( \vec{e}_1 \) and \( \vec{e}_2 \) in \( \mathbb{R}^2 \); and their rotated versions:

\[
\vec{r}_1(\theta) = \begin{bmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}, \\
\vec{r}_2(\theta) = \begin{bmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix}.
\]

In \( \mathbb{R}^4 \) they are orthonormal.

**Unit Length:** \( \sqrt{4 \times \frac{1}{2^2}} = 1 \).

**Orthogonality:** For each pair of vectors, two of the products \( \vec{u}_i(k)\vec{u}_j(k) \) will be positive and two negative; hence the sum \( \sum_{k=1}^{4} \vec{u}_i(k)\vec{u}_j(k) \) will be zero.

---

**Properties**

1 of 2

**Theorem (Properties of Orthonormal Vectors)**

a. **Orthonormal Vectors are linearly independent.**

b. **Orthonormal Vectors \( \vec{u}_1, \ldots, \vec{u}_n \in \mathbb{R}^n \) form a basis of \( \mathbb{R}^n \).**

a. Clearly, there is no way to (linearly) combine perpendicular vectors to describe each other (for example, think of \( \vec{e}_1, \vec{e}_2 \), and \( \vec{e}_3 \) in \( \mathbb{R}^3 \)).

b. By previous theorems, \( n \) linearly independent vectors in \( \mathbb{R}^n \) necessarily form a basis of \( \mathbb{R}^n \). (Think of the standard basis \( \vec{e}_1, \ldots, \vec{e}_n \in \mathbb{R}^n \); and rotations / reflections of it...)

---

**Theorem (Orthogonal Projection)** (old)

Consider a vector \( \vec{x} \in \mathbb{R}^n \) and a subspace \( V \) of \( \mathbb{R}^n \). Then we can write

\[
\vec{x} = \vec{x}^\| + \vec{x}^\perp,
\]

where \( \vec{x}^\| \in V \), and \( \vec{x}^\perp \perp V \). This representation is unique. The vector \( \vec{x}^\| \) is called the orthogonal projection of \( \vec{x} \) onto \( V \), sometimes denoted \( \text{proj}_V(\vec{x}) \); the transformation \( T(\vec{x}) = \text{proj}_V(\vec{x}) = \vec{x}^\| \) from \( \mathbb{R}^n \to \mathbb{R}^n \) is linear.

We have discussed this previously, but only in the context of describing the image and kernel of the projection... We are now ready to start discussing HOW we can compute the projection in any dimension space/subspace.
Orthogonal Projections and Orthonormal Bases — (9/33)

Orthogonal Projection: Formula

Theorem (Formula for the Orthogonal Projection)
If $V$ is a subspace of $\mathbb{R}^n$ with an orthonormal basis $\vec{u}_1, \ldots, \vec{u}_m$, then
\[
\text{proj}_V(x) = \tilde{x}^\parallel = (\vec{u}_1 \cdot x)\vec{u}_1 + \cdots + (\vec{u}_m \cdot x)\vec{u}_m
\]
$\forall x \in \mathbb{R}^n$.

Having the orthonormal basis is the absolute key to this formula. Any other basis will produce strange (incorrect) results.

Orthogonal Projections and Orthonormal Bases — (10/33)

Orthogonal Projection: Example

Example (Orthogonal Projection (part 1))
Consider the subspace $V = \text{im}(A)$ of $\mathbb{R}^4$, find $\text{proj}_V(\vec{x})$; where
\[
A = \begin{bmatrix}
1 & 1 \\
1 & -1 \\
1 & 1
\end{bmatrix}, \quad \vec{x} = \begin{bmatrix}
1 \\
3 \\
1
\end{bmatrix}, \quad \left\| \begin{bmatrix}
\pm 1 \\
\mp 1 \\
\pm 1
\end{bmatrix} \right\| = \sqrt{4(\pm 1)^2} = \sqrt{4} = 2.
\]

Since the columns $\vec{a}_1$ and $\vec{a}_2$ are linearly independent, and orthogonal (zero dot-product), they form an orthogonal basis of $V$. Dividing each vector by its length gives us an orthonormal basis for $V = \text{span}(\vec{u}_1, \vec{u}_2)$, where
\[
\vec{u}_1 = \begin{bmatrix}
1/2 \\
1/2 \\
1/2
\end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix}
-1/2 \\
-1/2 \\
1/2
\end{bmatrix}.
\]

Continued on the next slide...

Orthogonal Projections and Orthonormal Bases — (11/33)

Orthogonal Projection: Example

Example (Orthogonal Projection (part 2))
Now we use the projection formula,
\[
\vec{x} \cdot \vec{u}_1 = \begin{bmatrix}
1/2 \\
1/2 \\
1/2
\end{bmatrix} \cdot \begin{bmatrix}
3 \\
1 \\
7
\end{bmatrix} = 6, \quad \vec{x} \cdot \vec{u}_2 = \begin{bmatrix}
1/2 \\
-1/2 \\
1/2
\end{bmatrix} \cdot \begin{bmatrix}
3 \\
1 \\
7
\end{bmatrix} = 2,
\]
therefore
\[
\text{proj}_V(\vec{x}) = 6\vec{u}_1 + 2\vec{u}_2 = \begin{bmatrix}
3 \\
3 \\
3
\end{bmatrix} + \begin{bmatrix}
-1 \\
-1 \\
1
\end{bmatrix} = \begin{bmatrix}
4 \\
2 \\
2
\end{bmatrix}.
\]

Orthogonal Projections and Orthonormal Bases — (12/33)

Orthogonal Projection: Example

Example (Orthogonal Projection (part 3))

It is worth noting:
\[
\vec{x} = \begin{bmatrix}
1 \\
3 \\
1 \\
7
\end{bmatrix}, \quad \text{proj}_V(\vec{x}) = \begin{bmatrix}
4 \\
2 \\
2 \\
4
\end{bmatrix}.
\]

\[
\|\vec{x}\| = \sqrt{1 + 9 + 1 + 49} = \sqrt{60}, \quad \|\text{proj}_V(\vec{x})\| = \sqrt{16 + 4 + 4 + 16} = \sqrt{40}
\]
so that $\|\text{proj}_V(\vec{x})\| \leq \|\vec{x}\|$.
(The projection should NOT be longer than the original vector! We will show this in a few slides...)
The Orthogonal Complement

Using our recently acquired vocabulary, we realize that
\[ \text{im}(\text{proj}_V(\vec{x})) = V \]
Also, we know we can write
\[ \vec{x} = \vec{x}^\perp + \vec{x}^\parallel \]
and \( \text{proj}_V(\vec{x}) = \vec{x}^\parallel \), so if we are looking for the kernel, we want all \( \vec{x} \) without a \( \vec{x}^\parallel \) part, i.e. \( \{ \vec{x} \in \mathbb{R}^n : \vec{x} = \vec{x}^\perp \} \), the collection of all vectors orthogonal to the subspace \( V \)

A formal definition follows on the next slide...

 Orthogonal Projections and Orthonormal Bases — (13/33)  

Orthogonal Projection onto a Basis

Theorem (Orthogonal Projection onto a Basis)
Consider an orthonormal basis \( \vec{u}_1, \ldots, \vec{u}_n \) of \( \mathbb{R}^n \). Then
\[ \vec{x} = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_n \cdot \vec{x})\vec{u}_n \]
\[ \forall \vec{x} \in \mathbb{R}^n. \]
Since the basis spans \( \mathbb{R}^n \), we can “rebuild” \( \vec{x} \) completely by adding up all the projected pieces.
We have \( \vec{x} \) as a unique linear combination
\[ \vec{x} = c_1\vec{u}_1 + \cdots + c_n\vec{u}_n \]
where \( c_k = \vec{u}_k \cdot \vec{x} \).

Orthogonality allows us to compute the coordinates one-at-a-time, i.e. they are independent from each other.

 Image, and Kernel...

Orthogonal Projections and Orthonormal Bases — (14/33)  

The Orthogonal Complement

Definition (Orthogonal Complement)
Consider a subspace \( V \) of \( \mathbb{R}^n \). The Orthogonal Complement \( V^\perp \) of \( V \) is the set of those vectors \( \vec{x} \in \mathbb{R}^n \) that are orthogonal to all vectors in \( V \):
\[ V^\perp = \{ \vec{x} \in \mathbb{R}^n : \vec{x} \cdot \vec{v} = 0, \forall \vec{v} \in V \} \]
Note that \( V^\perp \) is the kernel of \( \text{proj}_V(\vec{x}) \).

Theorem (Properties of the Orthogonal Complement)
Consider a subspace \( V \) of \( \mathbb{R}^n \).
\[ \text{a. The Orthogonal Complement } V^\perp \text{ of } V \text{ is a subspace of } \mathbb{R}^n. \]
\[ \text{b. The intersection (common elements) of } V^\perp \text{ and } V \text{ consists of the zero vector: } V^\perp \cap V = \{ \vec{0} \}, \{ \vec{x} \in V^\perp \cap V : \vec{x} \cdot \vec{x} = 0 \Rightarrow \vec{x} = \vec{0} \} \]
\[ \text{c. dim}(V) + \text{dim}(V^\perp) = n. [\text{By Rank-Nullity Theorem: } T(\vec{x}) = \text{proj}_V(\vec{x})] \]
\[ \text{d. } (V^\perp)^\perp = V. \]
From Pythagoras to Cauchy-Schwarz

**Example (Really Old Stuff in New Notation)**

Consider a line $L$, and a vector $\vec{x}$ in $\mathbb{R}^n$. If we project $\vec{x}$ onto $L$ and write $\vec{x} = \vec{x}^\parallel + \vec{x}^\perp$ good ole' Pythagoras says

$$||\vec{x}||^2 = ||\vec{x}^\parallel + \vec{x}^\perp||^2 = ||\vec{x}^\parallel||^2 + ||\vec{x}^\perp||^2.$$  

Checking:

$$||\vec{x}||^2 = ||\vec{x}^\parallel + \vec{x}^\perp||^2 = (\vec{x}^\parallel + \vec{x}^\perp) \cdot (\vec{x}^\parallel + \vec{x}^\perp) = \vec{x}^\parallel \cdot \vec{x}^\parallel + \vec{x}^\parallel \cdot \vec{x}^\perp + \vec{x}^\perp \cdot \vec{x}^\parallel + \vec{x}^\perp \cdot \vec{x}^\perp = ||\vec{x}^\parallel||^2 + 0 + 0 + ||\vec{x}^\perp||^2$$

... and there it is!

From Pythagoras to Cauchy-Schwarz

Proof.

Since

$$||\vec{x}^\parallel||^2 + ||\vec{x}^\perp||^2 = ||\vec{x}||^2$$

(and all lengths are non-negative), we must have

$$||\vec{x}^\parallel||^2 \leq ||\vec{x}||^2.$$ 

\[ \square \]

From Pythagoras to Cauchy-Schwarz

**Theorem (Pythagorean Theorem)**

Consider two vectors $\vec{x}, \vec{y} \in \mathbb{R}^n$. The equation

$$||\vec{x} + \vec{y}||^2 = ||\vec{x}||^2 + ||\vec{y}||^2$$

holds if and only if $\vec{x} \perp \vec{y}$.

**Theorem (proj$_V(\vec{x})$ is no longer than $\vec{x}$)**

Consider a subspace $V$ of $\mathbb{R}^n$, and a vector $\vec{x} \in \mathbb{R}^n$. Then

$$||\text{proj}_V(\vec{x})|| \leq ||\vec{x}||$$

where equality is achieved if and only if $\vec{x} \in V$.

Now, ponder a one-dimensional subspace (a line through the origin) $V$ of $\mathbb{R}^n$, and let $\vec{y}$ be a vector in that subspace; let

$$\vec{u} = \frac{1}{||\vec{y}||} \vec{y}$$

be a unit vector spanning $V$.

We can now write the projection using $\vec{u}$:

$$\text{proj}_V(\vec{x}) = (\vec{x} \cdot \vec{u}) \vec{u}$$

It follows that

$$||\vec{x}|| \geq ||\text{proj}_V(\vec{x})|| = ||(\vec{x} \cdot \vec{u}) \vec{u}|| = ||\vec{x} \cdot \vec{u}|| ||\vec{u}|| = ||\vec{x} \cdot \vec{u}|| = \left| \frac{1}{||\vec{y}||} \vec{x} \cdot \vec{y} \right| = \frac{1}{||\vec{y}||} ||\vec{x} \cdot \vec{y}||$$

so that

$$|\vec{x} \cdot \vec{y}| \leq ||\vec{x}|| ||\vec{y}||$$
From Pythagoras to Cauchy-Schwarz

What we showed on the previous slide is known as:

**Theorem (The Cauchy-Schwarz Inequality)**

If \( \vec{x} \) and \( \vec{y} \in \mathbb{R}^n \), then

\[
|\vec{x} \cdot \vec{y}| \leq ||\vec{x}|| \cdot ||\vec{y}||.
\]

The statement is an equality if and only if \( \vec{x} \) and \( \vec{y} \) are parallel.

---

**Example: Angle Between Vectors**

Find the angle between

\[
\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\]

We have

\[
||\vec{x}|| = 1, \quad ||\vec{y}|| = \sqrt{4} = 2, \quad \vec{x} \cdot \vec{y} = 1
\]

so that

\[
\cos(\theta) = \frac{1}{2}, \quad \theta = \arccos\left(\frac{1}{2}\right) = \frac{\pi}{3}
\]

---

The Dot Product, \( \cos(\theta) \), and Cauchy-Schwarz

Consider two vectors \( \vec{x}, \vec{y} \in \mathbb{R}^n \). We have previously expressed the dot product in terms of the angle between the two vectors.

\[
\vec{x} \cdot \vec{y} = \cos(\theta) ||\vec{x}|| \cdot ||\vec{y}||
\]

The real use of the formula is to find

\[
\cos(\theta) = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| \cdot ||\vec{y}||}, \quad \theta = \arccos\left(\frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| \cdot ||\vec{y}||}\right)
\]

Cauchy-Schwarz, \( |\vec{x} \cdot \vec{y}| \leq ||\vec{x}|| \cdot ||\vec{y}|| \), guarantees that the argument to \( \arccos \), and the value of \( \cos(\theta) \) (as defined here) make sense.

Note that the angle \( \theta \) is in the plane spanned by \( \vec{x}, \vec{y} \).

---

**The Super-Slow, Slow, and Fast Fourier Transform**

A lot of signal analysis and processing is frequency-based; meaning that it is highly useful to express a signal using basis functions that are determined by various frequencies.

Consider the functions:

\[
\begin{align*}
\Phi_0(x) &= \frac{1}{2} \\
\Phi_k(x) &= \cos(kx), \quad k = 1, \ldots, n \\
\Phi_{n+k}(x) &= \sin(kx), \quad k = 1, \ldots, n - 1
\end{align*}
\]

and let each one define a vector \( \vec{v}_j \in \mathbb{R}^{2n} \), by evaluating the function in the points \( x_j = -\pi + (j\pi/n), \quad j = 0, 1, \ldots, (2m - 1) \).
The Super-Slow, Slow, and Fast Fourier Transform

- Let those vectors be the columns in a matrix, $M \in \mathbb{R}^{2n \times 2n}$.
- It turns out that the vectors are linearly independent, which makes them a basis, $\mathcal{B}$, for $\mathbb{R}^{2n}$ and the matrix $M$ invertible;
- further, the vectors are orthogonal (which will help us save some work).

- Now, if we have sampled a signal in $2n$ locations / timepoints; then we can collect those samples in $\vec{f} \in \mathbb{R}^{2n}$.
- If we can to express the signal as a linear combination of the cos/sin-vectors, all we have to do is solve the linear system
  
  $M[\vec{f}]_\mathcal{B} = \vec{f}$

  which in general requires roughly $\frac{8}{3}n^3$ operations ($\times/+$). This is the Super-Slow Fourier Transform.

Much of the analysis was done by Jean Baptiste Joseph Fourier in the early 1800s, but the use of the Fourier series representation was not practical until...

1965, when Cooley and Tukey* published a 4-page paper describing an algorithm which computes the coefficients using only $O(n \log_2 n)$ operations.

It is hard to overstate the importance of this paper!!!

The algorithm is now known as the “Fast Fourier Transform” or just the “FFT”. We sweep the details of the FFT under the rug; it comes down to some clever complex analysis, and the facts that $1 + 1 = 2$, and $1 - 1 = 0$.


Comparing Operation Counts

<table>
<thead>
<tr>
<th>$n$</th>
<th>$4n^2$</th>
<th>$3n + n \log_2 n$</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>1,024</td>
<td>112</td>
<td>9</td>
</tr>
<tr>
<td>64</td>
<td>16,384</td>
<td>576</td>
<td>28</td>
</tr>
<tr>
<td>256</td>
<td>262,144</td>
<td>2,816</td>
<td>93</td>
</tr>
<tr>
<td>1,024</td>
<td>4,194,304</td>
<td>13,312</td>
<td>315</td>
</tr>
<tr>
<td>4,096</td>
<td>67,108,864</td>
<td>61,440</td>
<td>1,092</td>
</tr>
<tr>
<td>16,384</td>
<td>1,073,741,824</td>
<td>278,528</td>
<td>3,855</td>
</tr>
<tr>
<td>262,144</td>
<td>274,877,906,944</td>
<td>505,024</td>
<td>49,932</td>
</tr>
<tr>
<td>1,048,576</td>
<td>4,398,046,511,104</td>
<td>24,117,248</td>
<td>182,361</td>
</tr>
<tr>
<td>4,194,304</td>
<td>70,368,744,177,664</td>
<td>104,857,600</td>
<td>671,088</td>
</tr>
<tr>
<td>8,388,608</td>
<td>281,474,976,710,656</td>
<td>218,103,808</td>
<td>1,290,555</td>
</tr>
<tr>
<td>16,777,216</td>
<td>1,125,899,906,842,624</td>
<td>452,984,832</td>
<td>2,485,513</td>
</tr>
<tr>
<td>33,554,432</td>
<td>4,503,599,627,370,496</td>
<td>939,524,096</td>
<td>4,793,490</td>
</tr>
</tbody>
</table>

$33,554,432 = 2^{13} \times 2^{12} = 8,196 \times 4,096$

† The “8k Digital Video Format” has a resolution of $8,192 \times 4,320$ pixels; the tentative Ultra High Definition Television (UHDTV) specification calls for $7,680 \times 4,320$ pixels for $16:9$ aspect ratio (120 fps, 12 bits/channel (at least 3 − RGB) ≈ 4.0 × 10^9 36-bit pixels/sec). IMAX shot on 70 mm film has a theoretical pixel resolution of $12,000 \times 8,700$ (at 24 fps, for a total of $2.5 \times 10^9$ pixels/sec).
### Comparing Operation Counts 2 of 4

<table>
<thead>
<tr>
<th>$n$</th>
<th>SLOW-FT time</th>
<th>FFT time</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>9 s</td>
<td>1 s</td>
<td>9</td>
</tr>
<tr>
<td>64</td>
<td>29 s</td>
<td>1 s</td>
<td>28</td>
</tr>
<tr>
<td>256</td>
<td>1 m 33 s</td>
<td>1 s</td>
<td>93</td>
</tr>
<tr>
<td>1,024</td>
<td>5 m 15 s</td>
<td>1 s</td>
<td>315</td>
</tr>
<tr>
<td>4,096</td>
<td>18 m 12 s</td>
<td>1 s</td>
<td>1,092</td>
</tr>
<tr>
<td>16,384</td>
<td>1:04:26 s</td>
<td>1 s</td>
<td>3,855</td>
</tr>
<tr>
<td>65,536</td>
<td>3:49:57 s</td>
<td>1 s</td>
<td>13,797</td>
</tr>
<tr>
<td>262,144</td>
<td>13:52:12 s</td>
<td>1 s</td>
<td>49,932</td>
</tr>
<tr>
<td>1,048,576</td>
<td>2d 02:39:21 s</td>
<td>1 s</td>
<td>182,361</td>
</tr>
<tr>
<td>4,194,304</td>
<td>7d 18:24:48 s</td>
<td>1 s</td>
<td>671,088</td>
</tr>
<tr>
<td>8,388,608</td>
<td>14d 22:29:15 s</td>
<td>1 s</td>
<td>1,290,555</td>
</tr>
<tr>
<td>16,777,216</td>
<td>28d 18:25:13 s</td>
<td>1 s</td>
<td>2,485,513</td>
</tr>
<tr>
<td>33,554,432</td>
<td>55d 11:31:30 s</td>
<td>1 s</td>
<td>4,793,490</td>
</tr>
</tbody>
</table>

### Comparing Operation Counts 3 of 4

#### Fourier Transform: Computational Complexity

- **FFT**
- **Slow FT**
- **Super Slow FT**

### Comparing Operation Counts 4 of 4

#### Fourier Transform: Timing, assuming $10^{19}$ op/s

- **Age of Earth**
- **One Lifetime**
- **One Year**
- **One Month**
- **One Day**
- **One Minute**

### Suggested Problems 5.1

Available on Learning Glass videos:

5.1 — 5, 7, 10, 11, 15, 17, 27, 28
<table>
<thead>
<tr>
<th>Lecture</th>
<th>Book, [GS5–]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>§4.1, §4.2, §4.4</td>
</tr>
<tr>
<td>5.2</td>
<td>§4.1, §4.2, §4.4</td>
</tr>
<tr>
<td>5.3</td>
<td>§4.1, §4.2, §4.4</td>
</tr>
</tbody>
</table>