Orthogonal Transforms and Orthogonal Matrices — (1/51)
1 Student Learning Objectives
   - SLOs: Orthogonal Transformations and Orthogonal Matrices

2 Orthogonal Transformations and Orthogonal Matrices
   - Examples, and Fundamental Theorems
   - Products, Inverses, and Transposes of Orthogonal Matrices
   - Matrix of Orthogonal Projection, using Orthonormal Basis

3 Suggested Problems
   - Suggested Problems 5.3
   - Lecture–Book Roadmap

4 Supplemental Material
   - Metacognitive Reflection
   - Problem Statements 5.3
   - Orthogonal Complements: Redux
   - Orthogonal Projections: Redux
   - Least Squares Data Fitting
After this lecture you should:

- Know what Orthogonal Transformations are; and their relation to Orthonormal Bases.
- Know the Properties of Orthogonal Matrices.
- Be able to perform an Orthogonal Projection using Orthonormal Basis you have constructed.
Orthogonal Transformations

For many reasons, we tend to “like” linear transformations that preserve the length of vectors; and angles between vectors:

**Definition (Orthogonal Transformations)**

A linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is called orthogonal if it preserves the length of vectors:

$$\| T(\vec{x}) \| = \| \vec{x} \|, \ \forall \vec{x} \in \mathbb{R}^n.$$

If $T(\vec{x}) = A\vec{x}$ is an orthogonal transformation, we say that $A$ is an orthogonal (or *unitary*, when it has complex entries) matrix.
Example: Rotations

The rotation

$$T(\vec{x'}) = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \vec{x}$$

is an orthogonal transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$, and $\forall \theta$. 

![Graphs showing rotations](image-url)
Example: Reflections

Consider a subspace $V$ of $\mathbb{R}^n$. For a vector $\vec{x} \in \mathbb{R}^n$, the vector \(\text{ref}_V(\vec{x}) = \vec{x}^\parallel - \vec{x}^\perp \equiv 2\text{proj}_V(\vec{x}) - \vec{x}\) is the reflection of $\vec{x}$ in $V$.

We show that reflections are orthogonal transformations:

By the Pythagorean theorem, we have

\[
\|\text{ref}_V(\vec{x})\|^2 = \|\vec{x}^\parallel - \vec{x}^\perp\|^2 = \|\vec{x}^\parallel\|^2 + \|\vec{x}^\perp\|^2 = \|\vec{x}\|^2
\]
Theorem (Preservation of Orthogonality)

Consider an orthogonal transformation \( T : \mathbb{R}^n \rightarrow \mathbb{R}^n \). If the vectors \( \vec{v}, \vec{w} \in \mathbb{R}^n \) are orthogonal, then so are \( T(\vec{v}) \) and \( T(\vec{w}) \).

Proof: \{Short: relies on fundamental properties/definitions\}.

By the theorem of Pythagoras, we have to show that

\[
\| T(\vec{v}) + T(\vec{w}) \|^2 = \| T(\vec{v}) \|^2 + \| T(\vec{w}) \|^2
\]

\[
\| T(\vec{v}) + T(\vec{w}) \|^2 = \| T(\vec{v} + \vec{w}) \|^2 \quad \text{[Linearity of } T]\n\]
\[
= \| \vec{v} + \vec{w} \|^2 \quad \text{[Orthogonality of } T]\n\]
\[
= \| \vec{v} \|^2 + \| \vec{w} \|^2 \quad [\vec{v} \perp \vec{w}]\n\]
\[
= \| T(\vec{v}) \|^2 + \| T(\vec{w}) \|^2 \quad \text{[Orthogonality of } T]\n\]
Theorem (Orthogonal Transformations and Orthonormal Bases)

a. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal if and only if the vectors $T(\vec{e}_1), \ldots, T(\vec{e}_n)$ form an orthonormal basis of $\mathbb{R}^n$.

b. An $n \times n$ matrix $A$ is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$.

Proof in supplemental slides.
Part (a).

⇒ If \( T \) is orthogonal, then, by definition, the \( T(\vec{e}_k) \) are unit vectors, and orthogonal by the previous theorem; hence a basis for \( \mathbb{R}^n \).

⇐ Conversely, suppose \( T(\vec{e}_1), \ldots, T(\vec{e}_n) \) form an orthonormal basis. Consider a vector \( \vec{x} = x_1 \vec{e}_1 + \cdots + x_n \vec{e}_n \in \mathbb{R}^n \). Then

\[
\| T(\vec{x}) \|^2 = \| x_1 T(\vec{e}_1) + \cdots + x_n T(\vec{e}_n) \|^2 \quad \text{[Linearity]}
\]

\[
= \| x_1 T(\vec{e}_1) \|^2 + \cdots + \| x_n T(\vec{e}_n) \|^2 \quad \text{[Pythagoras]}
\]

\[
= x_1^2 \| T(\vec{e}_1) \|^2 + \cdots + x_n^2 \| T(\vec{e}_n) \|^2
\]

\[
= x_1^2 + \cdots + x_n^2
\]

\[
= \| \vec{x} \|^2.
\]
Part (b).

This follows from the result from \[\text{NOTES 2.1}\] restated below...

\[\text{Theorem (The Columns of the Matrix of a Linear Transformation)}\]

Consider a linear transformation \(T : \mathbb{R}^m \to \mathbb{R}^n\). Then, the matrix of \(T\) is

\[
A = \begin{bmatrix}
T(\vec{e}_1) & T(\vec{e}_2) & \ldots & T(\vec{e}_m)
\end{bmatrix},
\]

where \(\vec{e}_i \in \mathbb{R}^m\) is the vector of all zeros, except entry \(\#i\) which is 1.
A Warning

A matrix with orthogonal columns need not be an orthogonal matrix, e.g.

\[
A = \begin{bmatrix}
4 & -3 \\
3 & 4
\end{bmatrix}.
\]

Example \((A = \begin{bmatrix} 4 & -3 \\ 3 & 4 \end{bmatrix})\)

\[
\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \|\vec{x}\| = \sqrt{2}, \quad A\vec{x} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}, \quad \|A\vec{x}\| = \sqrt{50}
\]
Theorem (Products and Inverses of Orthogonal Matrices)

a. The product $AB$ of two orthogonal $n \times n$ matrices $A$ and $B$ is orthogonal.

b. The inverse $A^{-1}$ of an orthogonal $n \times n$ matrix $A$ is orthogonal.

Proof: \{Short: relies on fundamental properties/definitions\}.

(a), the linear transformation $T(\vec{x}) = AB\vec{x}$ preserves length, because

$$\|T(\vec{x})\| = \|A(B\vec{x})\| = \|B\vec{x}\| = \|\vec{x}\|.$$ 

(b), the linear transformation $T(\vec{x}) = A^{-1}\vec{x}$ preserves length, since

$$\|A^{-1}\vec{x}\| = \|AA^{-1}\vec{x}\| = \|\vec{x}\|.$$
Example: Properties of the Transpose of an Orthonormal Matrix

Consider the orthogonal matrix $A$, and the matrix where the $ij$ entry has been shifted to the $ji$ position ($B$):

$$
A = \frac{1}{7} \begin{bmatrix} 2 & 6 & 3 \\ 3 & 2 & -6 \\ 6 & 3 & 2 \end{bmatrix}, \quad
B = \frac{1}{7} \begin{bmatrix} 2 & 3 & 6 \\ 6 & 2 & -3 \\ 3 & -6 & 2 \end{bmatrix}.
$$

We compute

$$
BA = \frac{1}{49} \begin{bmatrix} 49 & 0 & 0 \\ 0 & 49 & 0 \\ 0 & 0 & 49 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.
$$

The $k, \ell$ entry in $BA$ is the dot product of the $k^{th}$ row of $B$, and the $\ell^{th}$ column of $A$; by construction this is the dot product of the $k^{th}$ and $\ell^{th}$ columns of $A$; since $A$ is orthogonal this gives 1 when $k = \ell$, and 0 otherwise.
Matrix Transpose, Symmetric and Skew-symmetric Matrices

**Definition (Matrix Transpose, Symmetric and Skew-symmetric Matrices)**

Consider an $m \times n$ matrix $A$.

- The *transpose* $A^T$ of $A$ is the $n \times m$ matrix whose $ij^{th}$ entry is the $ji^{th}$ entry of $A$: The roles of rows and columns are reversed.
- We say that a square matrix $A$ is *symmetric* if $A^T = A$, and
- $A$ is called *skew-symmetric* if $A^T = -A$.

**Example (Transpose)**

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$
Example (Symmetric $2 \times 2$ Matrices)

The symmetric $2 \times 2$ matrices are of the form

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

They form a 3-dimensional subspace of $\mathbb{R}^{2\times2}$ with basis

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Note: $\mathbb{R}^{2\times2}$ (the collection of all 2-by-2 matrices) is a linear space (formal definition on the next slide)...
Definition (Linear Space)

A Linear Space $V$ is a set with a definition (rule) for addition “+”, and a definition (rule) for scalar multiplication; and the following must hold ($\forall u, v, w \in V, \forall c, k \in \mathbb{R}$)

a. $v + w \in V$.

b. $kv \in V$.

c. $(u + v) + w = u + (v + w)$.

d. $u + v = v + u$.

e. $\exists n \in V: u + n = u$, [Neutral Element, denoted by 0]

f. $\exists \hat{u}: u + \hat{u} = 0$; $\hat{u}$ unique, and denoted by $-u$.

g. $k(u + v) = ku + kv$.

h. $(c + k)u = cu + ku$.

i. $c(ku) = (ck)u$.

j. $1u = u$.

In $\mathbb{R}^{2 \times 2}$, the neutral element is $n = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.
Example (Skew-Symmetric $2 \times 2$ Matrices)

The symmetric $2 \times 2$ matrices are of the form

$$\begin{bmatrix}
0 & b \\
-b & 0
\end{bmatrix}$$

They form a 1-dimensional subspace of $\mathbb{R}^{2\times2}$ with basis

$$\left\{ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$$

Note: $\text{dim}(\mathbb{R}^{2\times2}) = 4$; $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis.
### Example (Transpose of a Vector)

$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \Rightarrow \vec{v}^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

### Theorem

*If $\vec{v}$ and $\vec{w}$ are two (column) vectors $\in \mathbb{R}^n$, then*

$$\vec{v} \cdot \vec{w} \equiv \vec{v}^T \vec{w}$$

*Dot Product*  

*"Matrix" Product*
Orthogonal Matrices: $A^T$ and $A^{-1}$

Theorem

Consider an $n \times n$ matrix $A$. The matrix $A$ is orthogonal if and only if $A^T A = I_n$ or, equivalently, if $A^{-1} = A^T$.

Proof: \{Short: relies on fundamental properties/definitions\}.

Write $A$ in terms of its columns:

$$A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix}$$

then

$$A^T A = \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_n^T \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{v}_1^T \vec{v}_1 & \cdots & \vec{v}_1^T \vec{v}_n \\ \vdots & \ddots & \vdots \\ \vec{v}_n^T \vec{v}_1 & \cdots & \vec{v}_n^T \vec{v}_n \end{bmatrix}$$

this is $I_n$ if and only if $A$ is orthogonal.
Summary :: Orthogonal Matrices

Consider an $n \times n$ matrix $A$. The following statements are equivalent:

i. $A$ is an orthogonal matrix.

ii. The transformation $T(\vec{x}) = A\vec{x}$ preserves length, that is, $\|A\vec{x}\| = \|\vec{x}\| \ \forall \vec{x} \in \mathbb{R}^n$.

iii. The columns of $A$ form an orthonormal basis of $\mathbb{R}^n$.

iv. $A^TA = I_n$.

v. $A^{-1} = A^T$.

vi. $A$ preserves the dot product, meaning that $(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$ $\forall \vec{x}, \vec{y} \in \mathbb{R}^n$. 
Theorem (Properties of the Transpose)

a. \((A + B)^T = A^T + B^T\) \(\forall A, B \in \mathbb{R}^{m \times n}\)
b. \((kA)^T = kA^T\) \(\forall A \in \mathbb{R}^{m \times n}, \forall k \in \mathbb{R}\)
c. \((AB)^T = (B^T A^T)\) \(\forall A \in \mathbb{R}^{m \times p}, \forall B \in \mathbb{R}^{p \times n}\)
d. \(\text{rank}(A) = \text{rank}(A^T)\) \(\forall \text{matrices } A\)
e. \((A^T)^{-1} = (A^{-1})^T\) \(\forall \text{invertible matrices } A\)
The Matrix of an Orthogonal Projection

We can use our expanded matrix-notation-language to express orthogonal projections.... First consider

\[ \text{proj}_L(\vec{x}) = (\vec{u}_1 \cdot \vec{x})\vec{u}_1 \]

onto a line \( L \) in \( \mathbb{R}^n \); where \( \vec{u}_1 \) is a unit vector in \( L \). Think of this vector as an \( n \times 1 \) matrix, and the scalar \( (\vec{u}_1 \cdot \vec{x}) \) as an \( 1 \times 1 \) matrix; we can rearrange

\[ \text{proj}_L(\vec{x}) = \vec{u}_1(\vec{u}_1 \cdot \vec{x}) \overset{1}{=} \vec{u}_1(\vec{u}_1^\top \vec{x}) \overset{2}{=} \vec{u}_1 \vec{u}_1^\top \vec{x} \overset{3}{=} (\vec{u}_1 \vec{u}_1^\top)\vec{x} \overset{4}{=} A\vec{x} \]

where \( A = \vec{u}_1 \vec{u}_1^\top \).

---

**Orthogonal Transforms and Orthogonal Matrices**

**Suggested Problems**

**Examples, and Fundamental Theorems**

**Products, Inverses, and Transposes of Orthogonal Matrices**

**The Matrix of an Ortho. Projection, using an Ortho. Basis**

---

**Vector-Vector Products**

**New: Outer Product**

\[ A = \begin{bmatrix} \mathbf{u} \mathbf{v}^T \end{bmatrix} \quad \text{is known as the outer product.} \]

\[ [n \times n] \times [n \times 1] \times [1 \times n] \]

**Old: Inner Product / Dot Product**

\[ s = \begin{bmatrix} \mathbf{u}^T \mathbf{v} \end{bmatrix} \]

\[ [1 \times 1] \times [1 \times n] \times [n \times 1] \]

**Upcoming: Cross Product**

\[ q = \begin{bmatrix} \mathbf{u} \mathbf{v} \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{u} \mathbf{v} \end{bmatrix} \]

\[ [3 \times 1] \times [3 \times 1] \times [3 \times 1], \quad [7 \times 1] \times [7 \times 1] \times [7 \times 1] \]

---

Peter Blomgren, ⟨blomgren.peter@gmail.com⟩

Orthogonal Transforms and Orthogonal Matrices— (23/51)
We can apply the same idea to the general projection formula

\[ \text{proj}_V(\vec{x}) = (\vec{u}_1 \cdot \vec{x}) \vec{u}_1 + \cdots + (\vec{u}_n \cdot \vec{x}) \vec{u}_n \]
\[ = \vec{u}_1 \vec{u}_1^T \vec{x} + \cdots + \vec{u}_n \vec{u}_n^T \vec{x} \]
\[ = \left( \vec{u}_1 \vec{u}_1^T + \cdots + \vec{u}_n \vec{u}_n^T \right) \vec{x} \]

and we can also write

\[ A = \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix} \]

We summarize on the next slide...
Theorem (The Matrix of an Orthogonal Projection: Summary)

Consider a subspace \( V \) of \( \mathbb{R}^n \) with orthonormal basis \( \vec{q}_1, \ldots, \vec{q}_m \). The matrix \( P \) of the orthogonal projection onto \( V \) is

\[
P = QQ^T, \quad \text{where } Q = \begin{bmatrix} \vec{q}_1 & \cdots & \vec{q}_m \end{bmatrix}.
\]

Note that it is \( QQ^T \) not \( Q^TQ \)

\( P \) is symmetric — \( P^T = (QQ^T)^T = (Q^T)^TQ^T = QQ^T = P \).
In (5.2.7) [SEE LEARNING GLASS] we orthogonalized the vectors

\[ \vec{v}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 18 \\ 0 \\ 0 \end{bmatrix}, \]

using the Gram-Schmidt method, and got

\[ \vec{q}_1 = \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix}, \quad \vec{q}_2 = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}, \quad \vec{q}_3 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}, \]

Let’s define \( \{Q_1 \in \mathbb{R}^{3 \times 1}, Q_2 \in \mathbb{R}^{3 \times 2}, Q_3 \in \mathbb{R}^{3 \times 3}\} \)

\[ Q_1 = [\vec{q}_1], \quad Q_2 = [\vec{q}_1 \ \vec{q}_2], \quad Q_3 = [\vec{q}_1 \ \vec{q}_2 \ \vec{q}_3]. \]
Now,

\[ Q_1 Q_1^T = \frac{1}{9} \begin{bmatrix} 4 & 4 & 2 \\ 4 & 4 & 2 \\ 2 & 2 & 1 \end{bmatrix}, \quad Q_1^T Q_1 = \begin{bmatrix} 1 \end{bmatrix} \]

\[ Q_2 Q_2^T = \frac{1}{9} \begin{bmatrix} 8 & 2 & -2 \\ 2 & 5 & 4 \\ -2 & 4 & 5 \end{bmatrix}, \quad Q_2^T Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \]

\[ Q_3 Q_3^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q_3^T Q_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \]

Note, \( Q_1 Q_1^T \), \( Q_2 Q_2^T \), and \( Q_3 Q_3^T \) are the matrices of orthogonal projections onto a line \( \text{span}(\vec{q}_1) \), a plane \( \text{span}(\vec{q}_1, \vec{q}_2) \), and \( \mathbb{R}^3 \) \( \text{span}(\vec{q}_1, \vec{q}_2, \vec{q}_3) \).
Available on Learning Glass videos:
5.3 — 1, 2, 5, 6, 13, 15, 17, 19, 28, 32, 33, 36, 41
<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>
# Metacognitive Exercise — Thinking About Thinking & Learning

<table>
<thead>
<tr>
<th>I know / learned</th>
<th>Almost there</th>
<th>Huh?!?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Right After Lecture</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>After Thinking / Office Hours / SI-session</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>After Reviewing for Midterm/Final</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(5.3.1) Is the given matrix Orthogonal?

\[ A = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & 0.6 \end{bmatrix} \]

(5.3.2) Is the given matrix Orthogonal?

\[ A = \begin{bmatrix} -0.8 & 0.6 \\ 0.6 & 0.8 \end{bmatrix} \]
If the $n \times n$ matrices $A$ and $B$ are orthogonal, are the following matrices orthogonal as well?

(5.3.5) $C = 3A$

(5.3.6) $D = -B$
If the $n \times n$ matrices $A$ and $B$ are symmetric, and $B$ is invertible; are the following matrices symmetric as well?

(5.3.13) $C = 3A$

(5.3.15) $D = AB$

(5.3.17) $F = B^{-1}$

(5.3.19) $G = 2I_n + 3A - 4A^2$
Consider an $n \times n$ matrix $A$. Show that $A$ is orthogonal if-and-only-if: $A$ preserves the dot product; i.e.

$$(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$$

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

Hint, show:

1. $A^T A = I_n \Rightarrow (A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$
2. $(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y} \Rightarrow L(\vec{x}) = A\vec{x}$ is length-preserving.
(5.3.32–a) Consider an $n \times m$ matrix $A$ such that $A^T A = I_m$. Is it necessarily true that $AA^T = I_n$? (Explain!)

(5.3.32–b) Consider an $n \times n$ matrix $A$ such that $A^T A = I_n$. Is it necessarily true that $AA^T = I_n$? (Explain!)

(5.3.33) Find all orthogonal $2 \times 2$ matrices.
(5.3.36) Find and orthogonal matrix of the form

\[ A = \begin{bmatrix} \frac{2}{3} & \frac{1}{\sqrt{2}} & a \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} & b \\ \frac{1}{3} & 0 & c \end{bmatrix} \]
(5.3.41) Find the matrix $A$ of the orthogonal projection on the line in $\mathbb{R}^n$ spanned by the vector

$$\tilde{1}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$$
This section provides one important answer to “why?!” we should care about orthogonality, orthogonal complements, and orthogonal projections.

We will talk about *Least Squares Solutions* to non-consistent linear systems. (From a slightly different point of view than [NOTES 5.2: SUPPLEMENT].)

The least squares formulation is useful for fitting model parameters to data and has applications in a wide range of fields: chemistry, physics, engineering, finance, economics, etc.

It is sometimes (often?) referred to as “*Linear Regression*.”
Supplemental Examples, Revisited

The Orthogonal Complement of the Image

Example (The Orthogonal Complement of $\text{im}(A)$)

Consider a subspace $V = \text{im}(A)$ of $\mathbb{R}^n$, where

$$A = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{bmatrix}.$$ 

Then the orthogonal complement is,

$$V^\perp = \{ \vec{x} \in \mathbb{R}^n : \vec{v} \cdot \vec{x} = 0, \, \forall \vec{v} \in V \}$$

$$= \{ \vec{x} \in \mathbb{R}^n : \vec{v}_i \cdot \vec{x} = 0, \, i = 1, \ldots, m \}$$

$$= \{ \vec{x} \in \mathbb{R}^n : \vec{v}_i^T \vec{x} = 0, \, i = 1, \ldots, m \}.$$

In other words, $V^\perp$ is the kernel of the matrix

$$A^T = \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_m^T \end{bmatrix}.$$
The Orthogonal Complement of the Image

**Theorem (The Orthogonal Complement of the Image)**

*For any matrix $A$,*

$$(\text{im}(A))^\perp = \ker (A^T)$$
Consider the line

\[ V = \text{im} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \]

Then

\[ V^\perp = \ker \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \]

is the plane with equation \( x_1 + 2x_2 + 3x_3 = 0 \); as usual we can parameterize (to get a basis), and Gram-Schmidt Orthogonalize (to make it orthonormal)

\[
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = s \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} = \hat{s} \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + \hat{t} \frac{1}{\sqrt{70}} \begin{bmatrix} -3 \\ -6 \\ 5 \end{bmatrix}.
\]
**Theorem**

a. If $A$ is an $m \times n$ matrix, then $\ker(A) = \ker(A^T A)$.

b. If $A$ is an $m \times n$ matrix with $\ker(A) = \{\vec{0}\}$, then $A^T A$ is invertible.

**Proof.**

a. Clearly, the kernel of $A$ is contained in the kernel of $A^T A$. Conversely, consider a vector $\vec{x} \in \ker(A^T A)$, so that $A^T A \vec{x} = \vec{0}$. Then, $A \vec{x}$ is in the image of $A$ and in the kernel of $A^T$. Since $\ker(A^T)$ is the orthogonal complement of $\text{im}(A)$ by the previous theorem, the vector $A \vec{x}$ is $\vec{0}$, [Notes 5.1], that is, $\vec{x} \in \ker(A)$.

b. Note that $A^T A$ is an $n \times n$ matrix. By part (a), $\ker(A^T A) = \{\vec{0}\}$, and $A^T A$ is therefore invertible. [Notes 3.3]
Orthogonal Projections

**Theorem**

Consider a vector $\vec{x} \in \mathbb{R}^n$ and a subspace $V$ of $\mathbb{R}^n$. Then, the orthogonal projection $\text{proj}_V(\vec{x})$ is the vector in $V$ closest to $\vec{x}$, in that

$$\|\vec{x} - \text{proj}_V(\vec{x})\| < \|\vec{x} - \vec{v}\|, \; \forall \vec{v} \in V \setminus \text{proj}_V(\vec{x}).$$

As usual $\vec{x}^\parallel \equiv \text{proj}_V(\vec{x})$, and $\vec{x}^\perp = \vec{x} - \vec{x}^\parallel$ is the orthogonal “left-over” of $\vec{x}$ after the projection. The distance $\|\vec{x}^\perp\|$ is the shortest distance from $V$ to $\vec{x}$.

If we move, in $V$, a distance $\epsilon$ away from $\vec{x}^\parallel$, the distance from that point to $\vec{x}$ is $\sqrt{\epsilon^2 + \|\vec{x}^\perp\|^2}$. [Pythagorean Theorem].
Consider a linear system $A\vec{x} = \vec{b}$, which is inconsistent; meaning that $\vec{b} \not\in \text{im}(A)$.

An inconsistent linear system does not have a solution (in the traditional sense).

However, we can find the $\vec{x}^*$ which is the best candidate in that it minimizes the distance between $A\vec{x}^*$ and $\vec{b}$ (even though that distance is not zero).

We measure that distance

$$||A\vec{x} - \vec{b}|| \equiv ||\vec{b} - A\vec{x}||$$

and call it the error, or residual.
**Definition (Least-Squares Solution)**

Consider a linear system

\[ Ax = b, \]

where \( A \) is an \( m \times n \) matrix. A vector \( x^* \in \mathbb{R}^n \) is called a **least-squares solution** of this system if

\[ \| b - Ax^* \| \leq \| b - Ax \|, \ \forall x \in \mathbb{R}^n. \]

The name **least-squares solution** comes from the fact that we are minimizing the sum-of-squares length of the error vector \( e^* = b - Ax. \)

If/When the system \( Ax = b \) is consistent the least-squares solution is the exact solution, and \( \| b - Ax^* \| = 0. \)
Finding Least-Squares Solutions

How do we hunt down this wild beast?!

- We want the least-squares solutions $\vec{x}^*$ to $A\vec{x} = \vec{b}$

- By definition we are looking for
  \[ \|\vec{b} - A\vec{x}^*\| \leq \|\vec{b} - A\vec{x}\|, \forall \vec{x} \in \mathbb{R}^n. \]

- Our projection theorem says:
  \[ A\vec{x}^* = \text{proj}_V(\vec{b}), \text{ where } V = \text{im}(A). \]

- So, the error is in the orthogonal complement of $\text{im}(A)$:
  \[ \vec{b} - A\vec{x}^* \in V^\perp = (\text{im}(A))^\perp = \text{ker}(A^T). \]

- Which means:
  \[ A^T(\vec{b} - A\vec{x}^*) = 0 \iff A^T A\vec{x} = A^T \vec{b}. \]
Theorem (The Normal Equations)

The least-squares solutions of the system $A \vec{x} = \vec{b}$, are the exact solutions of the (consistent) system $A^T A \vec{x} = A^T \vec{b}$. The system $A^T A \vec{x} = A^T \vec{b}$ is called the normal equations of $A \vec{x} = \vec{b}$.

The case where $\ker(A) = \{\vec{0}\}$ is of particular importance, since in that case the matrix $A^T A$ is invertible, and we can give a closed form expression for the solution:
Closed Form Least Square Solutions

**Theorem (Closed Form Expression for the Least Squares Solution)**

If \( \ker(A) = \{ \vec{0} \} \), the linear system \( A\vec{x} = \vec{b} \) has the unique least-squares solution

\[
\vec{x}^* = (A^T A)^{-1} A^T \vec{b},
\]

and

\[
A\vec{x}^* = \text{proj}_{\text{im}(A)}(\vec{b}) = \underbrace{A(A^T A)^{-1} A^T \vec{b}}_{P},
\]

where the matrix \( P = A(A^T A)^{-1} A^T \) is the matrix of the orthogonal projection onto \( \text{im}(A) \).

Note: Just because you can write down a mathematical expression, it does not mean using it for anything practical is a good idea.
A BIG Warning!

**WARNING**

Whereas the least-squares solution, and orthogonal projection CAN be expressed as

\[(A^T A)^{-1} A^T \vec{b}, \text{ and } A(A^T A)^{-1} A^T \vec{b}, \text{ respectively.}\]

Anyone using these expressions outside of small homework problems are likely to run into **Big Trouble!!!**

We do not have the tools (eigenvalues) to explain why yet, but the warning stands!
So... What Should One Do?

Well, recall the Gram-Schmidt Process, and the QR-factorization...

If we have computed $QR = A$, then the following is true:

**The Solution**

- $A\vec{x} = \vec{b}$
- $QR\vec{x} = \vec{b}$
- Multiply by $Q^T$
- $Q^T Q = I_n$
- $Q^T \vec{x} = Q^T \vec{b}$
- Solve
- $\vec{x}^* = R^{-1} Q^T \vec{b}$

**The Projection**

- $QR\vec{x}^* = QRR^{-1} Q^T \vec{b}$
- $QR\vec{x}^* = QQ^T \vec{b}$

<table>
<thead>
<tr>
<th>use</th>
<th>not</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\vec{x}^* = R^{-1} Q^T \vec{b}$</td>
<td>$(A^T A)^{-1} A^T \vec{b}$</td>
</tr>
<tr>
<td>$\text{proj}_{\text{im}(A)}(\vec{b}) = QQ^T \vec{b}$</td>
<td>$A(A^T A)^{-1} A^T \vec{b}$</td>
</tr>
</tbody>
</table>
It makes sense to return to the Least-Squares solutions with more tools (eigenvalues) in hand; but, alas, we will run out of time this semester.

Some additional examples and discussion can be found in [Available Online]:

<table>
<thead>
<tr>
<th>Class</th>
<th>Notes#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 541\textsuperscript{R.I.P.}</td>
<td>10, 11</td>
</tr>
<tr>
<td>Math 543</td>
<td>8, 14</td>
</tr>
<tr>
<td>Math 693a</td>
<td>22, 23, 24</td>
</tr>
</tbody>
</table>

Clearly, there’s a lot more to say...