Orthogonal Transformations and Orthogonal Matrices

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:

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

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 \( \forall \theta \).

Example (Rotations)

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

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

\[\square\]

### Products and Inverses of Orthogonal Matrices

#### 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

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

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

---

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

#### 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

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

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

\[
\| A^{-1}\vec{x} \| = \| AA^{-1}\vec{x} \| = \| \vec{x} \|.
\]

---

### A Warning

**WARNING!!!**

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

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

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

\[
\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \| \vec{x} \| = \sqrt{2}, \quad A\vec{x} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}, \quad \| A\vec{x} \| = \sqrt{50}
\]
**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 ($B$):

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 6 & 7 \end{bmatrix}. $$

We compute

$$BA = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 49 & 0 \\ 0 & 49 \\ 0 & 49 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}. $$

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.

**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

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Note: $\mathbb{R}^{2\times2}$ (the collection of all 2-by-2 matrices) is a linear space (formal definition on the next slide)...
**Example (Skew-Symmetric 2 × 2 Matrices)**

The symmetric 2 × 2 matrices are of the form

\[
\begin{bmatrix}
0 & b \\
-b & 0
\end{bmatrix}
\]

They form a 1-dimensional subspace of \( \mathbb{R}^{2 \times 2} \) with basis

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

Note: \( \dim(\mathbb{R}^{2 \times 2}) = 4; \) \{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \} \) is a basis.

**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:

\[
A = [\vec{v}_1 \ldots \vec{v}_n]
\]

then

\[
A^T A = \begin{bmatrix}
\vec{v}_1^T \\
\vdots \\
\vec{v}_n^T
\end{bmatrix}
\begin{bmatrix}
\vec{v}_1 & \ldots & \vec{v}_n
\end{bmatrix}
= \begin{bmatrix}
\vec{v}_1^T \vec{v}_1 & \ldots & \vec{v}_1^T \vec{v}_n \\
\vdots & \ddots & \vdots \\
\vec{v}_n^T \vec{v}_1 & \ldots & \vec{v}_n^T \vec{v}_n
\end{bmatrix}
\]

this is \( I_n \) if and only if \( A \) is orthogonal.
Properties of the Matrix Transpose

Theorem (Properties of the Transpose)

1. \((A + B)^T = A^T + B^T\) \(\forall A, B \in \mathbb{R}^{m \times n}\)
2. \((kA)^T = kA^T\) \(\forall A \in \mathbb{R}^{m \times n}, \forall k \in \mathbb{R}\)
3. \((AB)^T = (B^T A^T)\) \(\forall A \in \mathbb{R}^{m \times p}, \forall B \in \mathbb{R}^{p \times n}\)
4. \(\text{rank}(A) = \text{rank}(A^T)\) \(\forall \text{matrices } A\)
5. \((A^T)^{-1} = (A^{-1})^T\) \(\forall \text{invertible matrices } A\)

The Matrix of an Orthogonal Projection

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

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

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

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

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

Notation; ① Associative property for matrix multiplication; ② "Book-keeping"/interpretation.

Vector-Vector Products

New: Outer Product

\[a = \vec{u}\vec{v}^T \quad \text{is known as the outer product.}\]

Old: Inner Product / Dot Product

\[s = \vec{u}^T \vec{v} \quad \text{is known as the dot product.}\]

Upcoming: Cross Product

\[q = \vec{u} \times \vec{v} \quad (\text{or} \quad \vec{w} = \vec{u} \times \vec{v})\]
The Matrix of an Orthogonal Projection: Summary

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
\[
P = QQ^T, \quad \text{where} \quad Q = [\vec{q}_1 \ldots \vec{q}_m].
\]

- Note that it is \( QQ^T \) **not** \( Q^T Q \).
- \( P \) is symmetric — \( P^T = (QQ^T)^T = (Q^T)^TQ^T = QQT = P. \)

Let's define
\[
\vec{v}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}, \quad \vec{v}_3 = \begin{bmatrix} 18 \\ 0 \\ 0 \end{bmatrix},
\]
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,
\[
Q_1Q_1^T = \frac{1}{9} \begin{bmatrix} 4 & 4 & 2 \\ 4 & 4 & 2 \\ 2 & 2 & 1 \end{bmatrix}, \quad Q_1^TQ_1 = \begin{bmatrix} 1 \end{bmatrix}
\]
\[
Q_2Q_2^T = \frac{1}{9} \begin{bmatrix} 8 & 2 & -2 \\ 2 & 5 & 4 \\ -2 & 4 & 5 \end{bmatrix}, \quad Q_2^TQ_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]
\[
Q_3Q_3^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q_3^TQ_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]

Note, \( Q_1Q_1^T, Q_2Q_2^T, \) and \( Q_3Q_3^T \) 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 \( \mathbb{R}^3 \) \( \text{span}(\vec{q}_1, \vec{q}_2, \vec{q}_3) \).
### Lecture – Book Roadmap

<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>

### Suggested Problems 5.3

(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?

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

(5.3.5) If the \(n \times n\) matrices \(A\) and \(B\) are orthogonal, are the following matrices orthogonal as well?

\[
C = 3A
\]

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

\[ (5.3.13) \quad C = 3A \]
\[ (5.3.15) \quad D = AB \]
\[ (5.3.17) \quad F = B^{-1} \]
\[ (5.3.19) \quad G = 2I_n + 3A - 4A^2 \]

\[ (5.3.32), (5.3.33) \]

**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.

\[ (5.3.36) \]

**5.3.36** Find and orthogonal matrix of the form

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

\[
\vec{1}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}
\]

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 \([\text{Notes 5.2: Supplement}]\).)

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.”

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

\[
A^T = \begin{bmatrix} \vec{v}_1^T \\ \vdots \\ \vec{v}_m^T \end{bmatrix}.
\]

Theorem (The Orthogonal Complement of the Image)

For any matrix \( A \),

\[
(\text{im}(A))^\perp = \ker \left( A^T \right)
\]
A Line in \( \mathbb{R}^3 \)

**Example (A Line in \( \mathbb{R}^3 \))**

Consider the line

\[
V = \text{im} \left( \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right)
\]

Then

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

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)

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

**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

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

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].

**The Error, or Residual**

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.
Least-Squares Solution

Definition (Least-Squares Solution)
Consider a linear system
\[ A\tilde{x} = \tilde{b}, \]
where \( A \) is an \( m \times n \) matrix. A vector \( \tilde{x}^* \in \mathbb{R}^n \) is called a least-squares solution of this system if
\[ \|\tilde{b} - A\tilde{x}^*\| \leq \|\tilde{b} - A\tilde{x}\|, \forall \tilde{x} \in \mathbb{R}^n. \]
The name least-squares solution comes from the fact that we are minimizing the sum-of-squares length of the error vector \( \tilde{e} = \tilde{b} - A\tilde{x} \).

If/When the system \( A\tilde{x} = \tilde{b} \) is consistent the least-squares solution is the exact solution, and \( \|\tilde{b} - A\tilde{x}\| = 0 \).

Finding Least-Squares Solutions

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

The case where \( \text{ker}(A) = \{0\} \) 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:
\[ \tilde{x}^* = (A^T A)^{-1} A^T \tilde{b}, \]
\[ A\tilde{x}^* = \text{proj}_{\text{im}(A)}(\tilde{b}) = A(A^T A)^{-1} A^T \tilde{b}, \]
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

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

respectively.

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