Math 254: Introduction to Linear Algebra

Notes #5.3 — Orthogonal Transformations and Orthogonal Matrices

Peter Blomgren
⟨blomgren@sdsu.edu⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

http://terminus.sdsu.edu/

Spring 2022
(Revised: April 4, 2022)

Outline

SLOs 5.3 Orthogonal Transformations and Orthogonal Matrices

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 norm (length) of vectors; and angles between vectors:

Definition (Orthogonal Transformations)

A linear transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ is called orthogonal if it preserves the norm (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.

Related topic: "Isometries" in [MATH 524 (NOTES#7.2)].

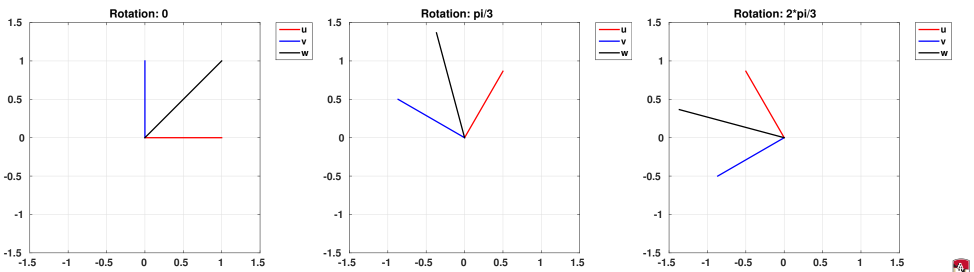
Example: Rotations

Example (Rotations)

The rotation

$$T(\vec{x}) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \vec{x}$$

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



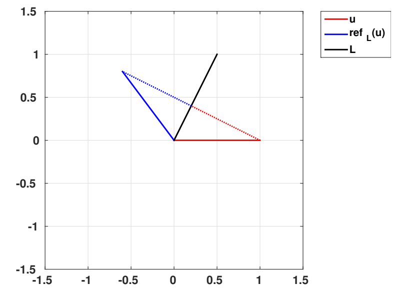
Example: Reflections

Example (Reflections)

Consider a subspace V of $\mathbb{R}^n$. For a vector $\vec{x} \in \mathbb{R}^n$, the vector $\mathrm{ref}_V(\vec{x}) = \vec{x}^{\parallel} - \vec{x}^{\perp} \equiv 2\mathrm{proj}_V(\vec{x}) - \vec{x}$ is the reflection of $\vec{x}$ in V. We show that reflections are orthogonal transformations:

By the [Pythagorean Theorem], we have

$$\|\mathrm{ref}_V(\vec{x})\|^2 = \left\|\vec{x}^{\parallel} - \vec{x}^{\perp}\right\|^2 = \left\|\vec{x}^{\parallel}\right\|^2 + \left\|-\vec{x}^{\perp}\right\|^2 = \left\|\vec{x}^{\parallel}\right\|^2 + \left\|\vec{x}^{\perp}\right\|^2 = \|\vec{x}\|^2$$



Preservation of Orthogonality

Theorem (Preservation of Orthogonality)

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

Proof

Proof (Preservation of Orthogonality {*Short: relies on fundamental properties/definitions*}**)**

By the theorem of Pythagoras, we have to show that

$$\|T(\vec{v}) + T(\vec{w})\|^2 = \|T(\vec{v})\|^2 + \|T(\vec{w})\|^2 :$$

$$\begin{aligned}
\|T(\vec{v}) + T(\vec{w})\|^2 &= \|T(\vec{v}+\vec{w})\|^2 && [\text{Linearity of } T] \\
&= \|\vec{v}+\vec{w}\|^2 && [\text{Orthogonality of } T] \\
&= \|\vec{v}\|^2 + \|\vec{w}\|^2 && [\vec{v} \perp \vec{w}] \\
&= \|T(\vec{v})\|^2 + \|T(\vec{w})\|^2 && [\text{Orthogonality of } T]
\end{aligned}$$

Orthogonal Transformations and Orthonormal Bases

Theorem (Orthogonal Transformations and Orthonormal Bases)

a. *A linear transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ is orthogonal if and only if the vectors $T(\vec{e}_1), \dots, T(\vec{e}_n)$ 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

[Proof in Supplemental Slides]

[Proof] Orthogonal Transformations and Orthonormal Bases [FOCUS :: MATH]

Proof (Part (a))

$\Rightarrow$ 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 $\mathbb{R}^n$.

$\Leftarrow$ Conversely, suppose $T(\vec{e}_1), \dots, T(\vec{e}_n)$ 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

$$\begin{aligned} \|T(\vec{x})\|^2 &= \|x_1 T(\vec{e}_1) + \cdots + x_n T(\vec{e}_n)\|^2 & \text{[Linearity]} \\ &= \|x_1 T(\vec{e}_1)\|^2 + \cdots + \|x_n T(\vec{e}_n)\|^2 & \text{[Pythagoras]} \\ &= x_1^2\, \|T(\vec{e}_1)\|^2 + \cdots + x_n^2\, \|T(\vec{e}_n)\|^2 \\ &= x_1^2 + \cdots + x_n^2 \\ &= \|\vec{x}\|^2. \end{aligned}$$

[Proof] Orthogonal Transformations and Orthonormal Bases [FOCUS :: MATH]

Proof (Part (b))

This follows from the result from [NOTES#2.1] restated below...

Theorem (The Columns of the Matrix of a Linear Transformation)

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

$$A = \begin{bmatrix} | & | & & | \\ T(\vec{e}_1) & T(\vec{e}_2) & \dots & T(\vec{e}_m) \\ | & | & & | \end{bmatrix},$$

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

A Warning

WARNING 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 ($A = \begin{bmatrix} 4 & -3 \\ 3 & 4 \end{bmatrix}$ has Orthogonal Columns, but is Not Orthogonal)

$$\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \|\vec{x}\| = \sqrt{2}, \quad A\vec{x} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}, \quad \|A\vec{x}\| = \sqrt{50}$$

Products and Inverses 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 norm (length), since $\|T(\vec{x})\| = \|A(B\vec{x})\| = \|B\vec{x}\| = \|\vec{x}\|$.

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

Example: Properties of the Transpose of an Orthonormal Matrix

Example

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 ℓ^{th} column of A; by construction this is the dot product of the k^{th} and ℓ^{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)

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

[Focus :: Math] Symmetric (2×2) Matrices [Linear Spaces]

Example (Symmetric (2×2) Matrices)

The symmetric (2×2) matrices are of the form

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

They form a **3**-dimensional subspace of $\mathbb{R}^{2\times 2}$ 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: $\mathbb{R}^{2\times 2}$ (the collection of all 2-by-2 matrices) is a linear space (formal definition on the next slide)...

[Focus :: Math] Linear Space Definition [Linear Spaces]

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. $\exists \widehat{u}$: $u + \widehat{u} = 0$; $\widehat{u}$ 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 $n = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

[Focus :: Math] Skew-Symmetric (2 × 2) Matrices [Linear Spaces]

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

Transpose of a Vector

Example (Transpose of a Vector)

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

We use this all the time:

Theorem

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

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

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:

$$A = [\vec{v}_1 \quad \dots \quad \vec{v}_n,]$$

then

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

this is I_n if and only if A is orthogonal.

Orthogonal Matrices: Summary

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 norm (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.** $A^T A = I_n$.
- **v.** $A^{-1} = A^T$.
- **vi.** A preserves the dot product, meaning that $(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$ $\forall \vec{x}, \vec{y} \in \mathbb{R}^n$.

Properties of the Matrix Transpose

Theorem (Properties of the Transpose)

a. $(A+B)^T = A^T + B^T \quad \forall A, B \in \mathbb{R}^{m\times n}$

b. $(kA)^T = kA^T \quad \forall A \in \mathbb{R}^{m\times n}, \forall k \in \mathbb{R}$

c. $(AB)^T = (B^T A^T) \quad \forall A \in \mathbb{R}^{m\times p}, \forall B \in \mathbb{R}^{p\times n}$

d. $\text{rank}(A) = \text{rank}(A^T) \quad \forall \textit{matrices } A$

e. $(A^T)^{-1} = (A^{-1})^T \quad \forall \textit{invertible matrices } A$

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 (1×1) matrix; we can rearrange

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

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

(!) We derived an expression for A (for action in $\mathbb{R}^2$) back in [Notes#2.2].

①Notation; ②Associative property for matrix multiplication; ③Associative property for matrix multiplication; ④"Book-keeping"/interpretation.

Vector-Vector Products

New: Outer Product $\quad \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$

$\underbrace{A}_{[n\times n]} = \underbrace{\vec{u}\vec{v}^T}_{[n\times 1]\times[1\times n]}$ is known as the **outer product**.

Old: Inner Product / Dot Product $\quad \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^1$

$\underbrace{s}_{[1\times 1]} = \underbrace{\vec{u}^T\vec{v}}_{[1\times n]\times[n\times 1]}$

Upcoming: Cross Product $\quad \mathbb{R}^3 \times \mathbb{R}^3 \mapsto \mathbb{R}^3$, (or $\mathbb{R}^7 \times \mathbb{R}^7 \mapsto \mathbb{R}^7$)

$\underbrace{\vec{q}}_{[3\times 1]} = \underbrace{\vec{u}}_{[3\times 1]} \times \underbrace{\vec{v}}_{[3\times 1]}, \quad \left(\text{or} \quad \underbrace{\vec{w}}_{[7\times 1]} = \underbrace{\vec{u}}_{[7\times 1]} \times \underbrace{\vec{v}}_{[7\times 1]}\right)$

The Matrix of an Orthogonal Projection

We can apply the same idea to the general projection formula

$$\begin{aligned} \text{proj}_V(\vec{x}) &= (\vec{u}_1 \cdot \vec{x})\vec{u}_1 + \cdots + (\vec{u}_n \cdot \vec{x})\vec{u}_n \\ &= \vec{u}_1\vec{u}_1^T\vec{x} + \cdots + \vec{u}_n\vec{u}_n^T\vec{x} \\ &= \underbrace{\left(\vec{u}_1\vec{u}_1^T + \cdots + \vec{u}_n\vec{u}_n^T\right)}_{A}\vec{x} \end{aligned}$$

and we can also write

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

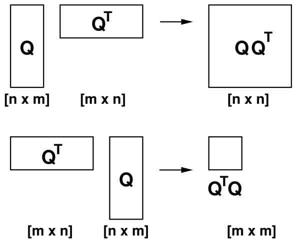
We summarize on the next slide...

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

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



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

Example

In (5.2.7) [SEE LEARNING GLASS] we orthogonaliz(ed) 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 get(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 $\{Q_1 \in \mathbb{R}^{3\times 1},\ Q_2 \in \mathbb{R}^{3\times 2},\ Q_3 \in \mathbb{R}^{3\times 3}\}$

$$Q_1 = \begin{bmatrix} \vec{q}_1 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 \end{bmatrix}, \quad Q_3 = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \vec{q}_3 \end{bmatrix}.$$

Example

Projection Matrices	Orthonormality Confirmation
$P_1 = Q_1Q_1^T = \frac{1}{9}\begin{bmatrix} 4 & 4 & 2 \\ 4 & 4 & 2 \\ 2 & 2 & 1 \end{bmatrix}$	$Q_1^TQ_1 = \begin{bmatrix} 1 \end{bmatrix}$
$P_2 = Q_2Q_2^T = \frac{1}{9}\begin{bmatrix} 8 & 2 & -2 \\ 2 & 5 & 4 \\ -2 & 4 & 5 \end{bmatrix}$	$Q_2^TQ_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$P_3 = Q_3Q_3^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$Q_3^TQ_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Note, $Q_1Q_1^T$, $Q_2Q_2^T$, and $Q_3Q_3^T$ are the matrices of orthogonal projections onto a line $L = \text{span}(\vec{q}_1)$, a plane $V = \text{span}(\vec{q}_1, \vec{q}_2)$, and $\mathbb{R}^3 = \text{span}(\vec{q}_1, \vec{q}_2, \vec{q}_3)$.

Suggested Problems 5.3

Available on Learning Glass videos:

5.3 — 1, 2, 5, 6, 13, 15, 17, 19, 28, 32, 33, <u>36</u>, 41

Lecture – Book Roadmap

Lecture	Book, [GS5–]
5.1	§4.1, §4.2, **§4.4**
5.2	§4.1, §4.2, **§4.4**
5.3	§4.1, §4.2, **§4.4**

Metacognitive Exercise — Thinking About Thinking & Learning

I know / learned	Almost there	Huh?!?
Right After Lecture		
After Thinking / Office Hours / SI-session		
After Reviewing for Quiz/Midterm/Final		

(5.3.1), (5.3.2)

(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), (5.3.6)

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) $D = -B$

(5.3.13), (5.3.15), (5.3.17), (5.3.19)

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) $G = 2I_n + 3A - 4A^2$

(5.3.28)

(5.3.28) 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 norm (length)-preserving.

(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 is 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 is necessarily true that $AA^T = I_n$? (Explain!)

(5.3.33) Find all orthogonal (2×2) matrices.

(5.3.36)

(5.3.36) Find an orthogonal matrix of the form

$$A = \begin{bmatrix} 2/3 & 1\sqrt{2} & a \\ 2/3 & -1/\sqrt{2} & b \\ 1/3 & 0 & c \end{bmatrix}$$

(5.3.41)

(5.3.41) Find the matrix A of the orthogonal projection onto the line in $\mathbb{R}^n$ spanned by the vector

$$\vec{1}_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$$

Revisiting the "Why?!?"

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."*

The Orthogonal Complement of the Image

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

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

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

Then the *orthogonal complement* is,

$$\begin{aligned} V^\perp &= \{\vec{x} \in \mathbb{R}^n : \vec{v} \cdot \vec{x} = 0,\ \forall \vec{v} \in V\} \\ &= \{\vec{x} \in \mathbb{R}^n : \vec{v}_i \cdot \vec{x} = 0,\ i = 1, \dots, m\} \\ &= \left\{\vec{x} \in \mathbb{R}^n : \vec{v}_i^T \vec{x} = 0,\ i = 1, \dots, m\right\}. \end{aligned}$$

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,

$$(\mathrm{im}(A))^\perp = \ker\left(A^T\right)$$

A Line in $\mathbb{R}^3$

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

Consider the line

$$V = \mathrm{im}\left(\begin{bmatrix}1\\2\\3\end{bmatrix}\right)$$

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)

$$\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = s\begin{bmatrix}-2\\1\\0\end{bmatrix} + t\begin{bmatrix}-3\\0\\1\end{bmatrix} = \tilde{s}\frac{1}{\sqrt{5}}\begin{bmatrix}-2\\1\\0\end{bmatrix} + \tilde{t}\frac{1}{\sqrt{70}}\begin{bmatrix}-3\\-6\\5\end{bmatrix}.$$

[Focus :: Math] $\ker(A)$, $\ker(A^TA)$, and Invertibility of A^TA

Theorem

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

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

Proof (Proof)

a. Clearly, the kernel of A is contained in the kernel of A^TA. Conversely, consider a vector $\vec{x} \in \ker(A^TA)$, so that $A^TA\vec{x} = \vec{0}$. 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 $\mathrm{im}(A)$ 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^TA is an $(n \times n)$ matrix. By part (a), $\ker(A^TA) = \{\vec{0}\}$, and A^TA 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 $\mathrm{proj}_V(\vec{x})$ is the vector in V **closest** *to $\vec{x}$, in that*

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

As usual $\vec{x}^{\|} \equiv \mathrm{proj}_V(\vec{x})$, and $\vec{x}^\perp = \vec{x} - \vec{x}^{\|}$ 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 ϵ away from $\vec{x}^{\|}$, 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 $\vec{b} \notin \mathrm{im}(A)$.

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
$$A\vec{x} = \vec{b},$$
where A is an $(m \times n)$ matrix. A vector $\vec{x}^* \in \mathbb{R}^n$ is called a *least-squares solution* of this system if
$$\|\vec{b} - A\vec{x}^*\| \leq \|\vec{b} - A\vec{x}\|, \ \forall \vec{x} \in \mathbb{R}^n.$$

The name *least-squares solution* comes from the fact that we a minimizing the sum-of-squares norm (length) of the error vector $\vec{e} = \vec{b} - A\vec{x}$.

If/When the system $A\vec{x} = \vec{b}$ is consistent the least-squares solution is the exact solution, and $\|\vec{b} - A\vec{x}^*\| = 0$.

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:
  - $A\vec{x}^* = \text{proj}_V(\vec{b})$, where $V = \text{im}(A)$.
- So, the error is in the orthogonal complement of $\text{im}(A)$:
  - $\vec{b} - A\vec{x}^* \in V^\perp = (\text{im}(A))^\perp = \ker(A^T)$.
- Which means:
  - $A^T(\vec{b} - A\vec{x}^*) = 0 \Leftrightarrow A^TA\vec{x} = A^T\vec{b}$.

Finding Least-Squares Solutions — The Normal Equations

Theorem (The Normal Equations)

The least-squares solutions of the system $A\vec{x} = \vec{b}$, are the exact solutions of the (consistent) system $A^TA\vec{x} = A^T\vec{b}$. The system $A^TA\vec{x} = A^T\vec{b}$ 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^TA 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
$$\vec{x}^* = (A^TA)^{-1}A^T\vec{b},$$
and
$$A\vec{x}^* = \text{proj}_{\text{im}(A)}(\vec{b}) = \underbrace{A(A^TA)^{-1}A^T}_{P}\vec{b},$$
where the matrix $P = A(A^TA)^{-1}A^T$ 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	$Q^TQR\vec{x}$	$=$	$Q^T\vec{b}$
$Q^TQ = I_n$	$R\vec{x}$	$=$	$Q^T\vec{b}$
solve	$\vec{x}^*$	$=$	$\mathbf{R^{-1}Q^T\vec{b}}$
THE PROJECTION	$QR\vec{x}^*$	$=$	$QRR^{-1}Q^T\vec{b}$
	$QR\vec{x}^*$	$=$	$\mathbf{QQ^T\vec{b}}$

		use	not
$\vec{x}^*$	$=$	$R^{-1}Q^T\vec{b}$	$(A^TA)^{-1}A^T\vec{b}$
$\mathrm{proj}_{\mathrm{im}(A)}(\vec{b})$	$=$	$QQ^T\vec{b}$	$A(A^TA)^{-1}A^T\vec{b}$

More Examples and Discussion???

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]:

Class	Notes#
Math 541 R.I.P.	10, 11
Math 524	6
Math 543	8, 14
Math 693a	22, 23, 24

Clearly, there's a lot more to say...