Section 6.1 : Inner Product, Length, and Orthogonality

Chapter 6: Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

1. Dot product of vectors
2. Magnitude of vectors, and distances in $\mathbb{R}^n$
3. Orthogonal vectors and complements
4. Angles between vectors

Learning Objectives

1. Compute (a) dot product of two vectors, (b) length (or magnitude) of a vector, (c) distance between two points in $\mathbb{R}^n$, and (d) angles between vectors.
2. Apply theorems related to orthogonal complements, and their relationships to Row and Null space, to characterize vectors and linear systems.

Motivating Question

For a matrix A, which vectors are orthogonal to all the rows of A? To the columns of A?

The Dot Product

The dot product between two vectors, $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$, is defined as

$$\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Example 1: For what values of k is $\vec{u} \cdot \vec{v} = 0$?

$$\vec{u} = \begin{pmatrix} -1 \\ 3 \\ k \\ 2 \end{pmatrix}, \qquad \vec{v} = \begin{pmatrix} 4 \\ 2 \\ 1 \\ -3 \end{pmatrix}$$

Properties of the Dot Product

The dot product is a special form of matrix multiplication, so it inherits linear properties.

Theorem (Basic Identities of Dot Product)

Let $\vec{u}, \vec{v}, \vec{w}$ be three vectors in $\mathbb{R}^n$, and $c \in \mathbb{R}$.

1. (Symmetry) $\vec{u} \cdot \vec{w} =$ ____________
2. (Linear in each vector) $(\vec{v} + \vec{w}) \cdot \vec{u} =$ ____________
3. (Scalars) $(c\vec{u}) \cdot \vec{w} =$ ____________
4. (Positivity) $\vec{u} \cdot \vec{u} \geq 0$, and the dot product equals ____________

The Length of a Vector

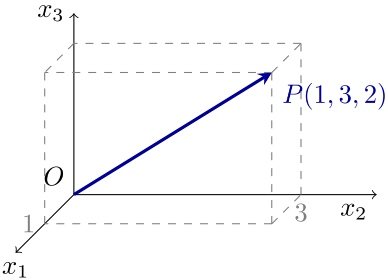
Definition

The **length** of a vector $\vec{u} \in \mathbb{R}^n$ is

$$\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

Example: the length of the vector $\overrightarrow{OP}$ is

$$\sqrt{1^2 + 3^2 + 2^2} = \sqrt{14}$$



Example

Let $\vec{u}, \vec{v}$ be two vectors in $\mathbb{R}^n$ with $\|\vec{u}\| = 5$, $\|\vec{v}\| = \sqrt{3}$, and $\vec{u} \cdot \vec{v} = -1$. Compute the value of $\|\vec{u} + \vec{v}\|$.

Length of Vectors and Unit Vectors

Note: for any vector $\vec{v}$ and scalar c, the length of $c\vec{v}$ is

$$||c\vec{v}|| = |c| \, ||\vec{v}||$$

Definition

If $\vec{v} \in \mathbb{R}^n$ has length one, we say that it is a **unit vector.**

For example, each of the following vectors are unit vectors.

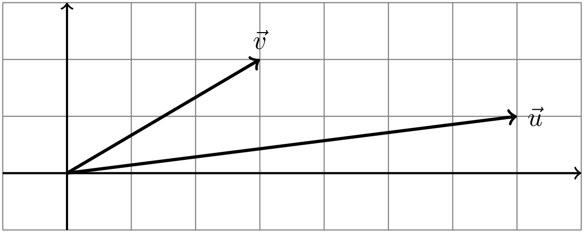
$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \vec{y} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \vec{v} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

Distance in $\mathbb{R}^n$

Definition

For $\vec{u}, \vec{v} \in \mathbb{R}^n$, the **distance** between $\vec{u}$ and $\vec{v}$ is given by the formula

Example: Compute the distance from $\vec{u} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}$ and $\vec{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.



Orthogonality

Definition (Orthogonal Vectors)

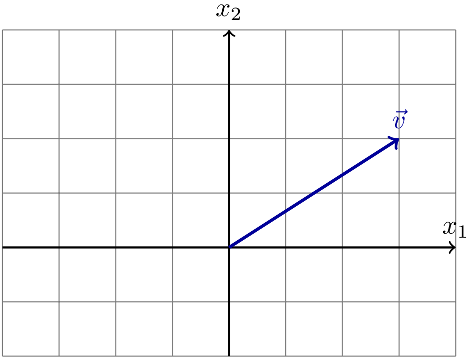
Two vectors $\vec{u}$ and $\vec{w}$ are **orthogonal** if $\vec{u} \cdot \vec{w} = 0$. This is equivalent to:

$$\|\vec{u} + \vec{w}\|^2 =$$

Note: The zero vector in $\mathbb{R}^n$ is orthogonal to every vector in $\mathbb{R}^n$. But we usually only mean non-zero vectors.

Example

Sketch the subspace spanned by the set of all vectors $\vec{u}$ that are orthogonal to $\vec{v} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.



Orthogonal Compliments

Definitions

Let W be a subspace of $\mathbb{R}^n$. Vector $\vec{z} \in \mathbb{R}^n$ is **orthogonal** to W if $\vec{z}$ is orthogonal to every vector in W.

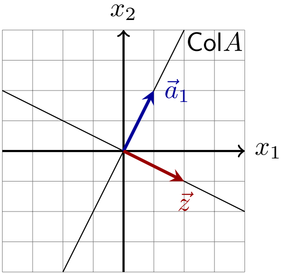
The set of all vectors orthogonal to W is a subspace, the **orthogonal compliment** of W, or $W^\perp$ or 'W perp.'

$$W^\perp = \{\vec{z} \in \mathbb{R}^n \ : \ \vec{z} \cdot \vec{w} = 0 \text{ for all } \vec{w} \in W\}$$

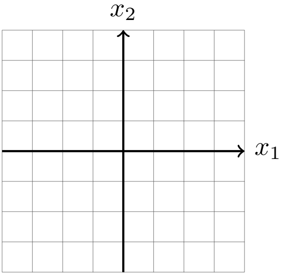
Example

Example: suppose $A = \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$.

- Col$A$ is the span of $\vec{a}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$
- Col$A^\perp$ is the span of $\vec{z} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$

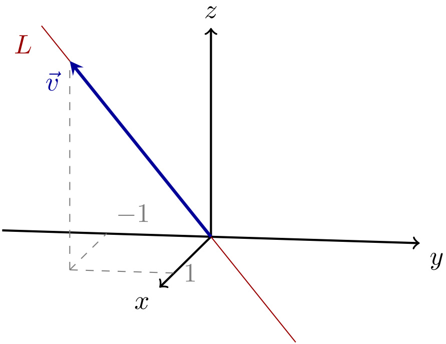


Sketch Null$A$ and Null$A^\perp$ on the grid below.



Example

Line L is a subspace of $\mathbb{R}^3$ spanned by $\vec{v} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$. Then the space $L^\perp$ is a plane. Construct an equation of the plane $L^\perp$.



Can also visualise line and plane with CalcPlot3D: web.monroecc.edu/calcNSF

Row$A$

Definition

Row$A$ is the space spanned by the rows of matrix A.

We can show that

- dim(Row(A)) $=$ dim(Col(A))
- a basis for Row$A$ is the pivot rows of A

Note that Row(A) $=$ Col(A^T), but in general Row$A$ and Col$A$ are not related to each other

Example 3

Describe the Null(A) in terms of an orthogonal subspace.

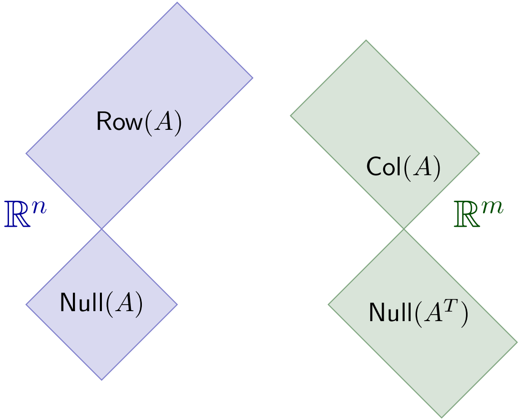
A vector $\vec{x}$ is in $\mathrm{Null}\, A$ if and only if

1. $A\vec{x} =$
2. This means that $\vec{x}$ is [ ] to each row of A.
3. $\mathrm{Row}\, A$ is [ ] to $\mathrm{Null}\, A$.
4. The dimension of $\mathrm{Row}\, A$ plus the dimension of $\mathrm{Null}\, A$ equals [ ]

Theorem (The Four Subspaces)

For any $A \in \mathbb{R}^{m\times n}$, the orthogonal complement of $\mathrm{Row}\, A$ is $\mathrm{Null}\, A$, and the orthogonal complement of $\mathrm{Col}\, A$ is $\mathrm{Null}\, A^T$.

The idea behind this theorem is described in the diagram below.



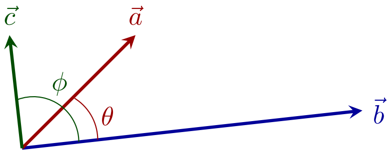
Angles

Theorem

$\vec{a} \cdot \vec{b} = |\vec{a}|\,|\vec{b}| \cos\theta$. Thus, if $\vec{a} \cdot \vec{b} = 0$, then:

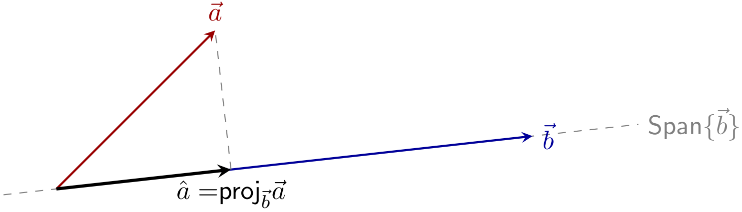
- $\vec{a}$ and/or $\vec{b}$ are ________ vectors, or
- $\vec{a}$ and $\vec{b}$ are ________.

For example, consider the vectors below.



Looking Ahead - Projections

Suppose we want to find the closed vector in Span$\{\vec{b}\}$ to $\vec{a}$.



- Later in this Chapter, we will make connections between dot products and **projections**.
- Projections are also used throughout multivariable calculus courses.

Section 6.2 : Orthogonal Sets

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra

Topics and Objectives

Topics

1. Orthogonal Sets of Vectors
2. Orthogonal Bases and Projections.

Learning Objectives

1. Apply the concepts of orthogonality to
   a) compute orthogonal projections and distances,
   b) express a vector as a linear combination of orthogonal vectors,
   c) characterize bases for subspaces of $\mathbb{R}^n$, and
   d) construct orthonormal bases.

Motivating Question

What are the special properties of this basis for $\mathbb{R}^3$?

$$\begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} / \sqrt{11}, \quad \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} / \sqrt{6}, \quad \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix} / \sqrt{66}$$

Orthogonal Vector Sets

Definition

A set of vectors $\{\vec{u}_1, \ldots, \vec{u}_p\}$ are an **orthogonal set** of vectors if for each $j \neq k$, $\vec{u}_j \perp \vec{u}_k$.

Example: Fill in the missing entries to make $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ an orthogonal set of vectors.

$$\vec{u}_1 = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} \quad -2 \quad \\ 0 \\ \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} \quad 0 \quad \\ \\ \end{bmatrix}$$

Linear Independence

Theorem (Linear Independence for Orthogonal Sets)

Let $\{\vec{u}_1, \dots, \vec{u}_p\}$ be an orthogonal set of vectors. Then, for scalars $c_1, \dots, c_p$,

$$\left\|c_1\vec{u}_1 + \cdots + c_p\vec{u}_p\right\|^2 = c_1^2\|\vec{u}_1\|^2 + \cdots + c_p^2\|\vec{u}_p\|^2.$$

In particular, if all the vectors $\vec{u}_r$ are non-zero, the set of vectors $\{\vec{u}_1, \dots, \vec{u}_p\}$ are linearly independent.

Orthogonal Bases

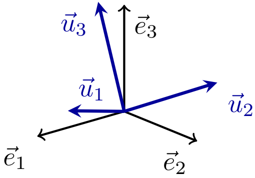
Theorem (Expansion in Orthogonal Basis)

Let $\{\vec{u}_1, \ldots, \vec{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Then, for any vector $\vec{w} \in W$,

$$\vec{w} = c_1\vec{u}_1 + \cdots + c_p\vec{u}_p.$$

Above, the scalars are $c_q = \dfrac{\vec{w} \cdot \vec{u}_q}{\vec{u}_q \cdot \vec{u}_q}$.

For example, any vector $\vec{w} \in \mathbb{R}^3$ can be written as a linear combination of $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, or some other orthogonal basis $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$.



Example

$$\vec{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{u} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{s} = \begin{pmatrix} 3 \\ -4 \\ 1 \end{pmatrix}$$

Let W be the subspace of $\mathbb{R}^3$ that is orthogonal to $\vec{x}$.

a) Check that an orthogonal basis for W is given by $\vec{u}$ and $\vec{v}$.

b) Compute the expansion of $\vec{s}$ in basis W.

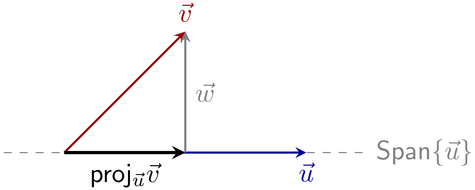
Projections

Let $\vec{u}$ be a non-zero vector, and let $\vec{v}$ be some other vector. The **orthogonal projection of $\vec{v}$ onto the direction of $\vec{u}$** is the vector in the span of $\vec{u}$ that is closest to $\vec{v}$.

$$\text{proj}_{\vec{u}}\vec{v} = \frac{\vec{v}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\vec{u}.$$

The vector $\vec{w} = \vec{v} - \text{proj}_{\vec{u}}\vec{v}$ is orthogonal to $\vec{u}$, so that

$$\vec{v} = \text{proj}_{\vec{u}}\vec{v} + \vec{w}$$
$$\|\vec{v}\|^2 = \|\text{proj}_{\vec{u}}\vec{v}\|^2 + \|\vec{w}\|^2$$



Example

Let L be spanned by $\vec{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

1. Calculate the projection of $\vec{y} = (-3, 5, 6, -4)$ onto line L.
2. How close is $\vec{y}$ to the line L?

Definition

Definition (Orthonormal Basis)

An **orthonormal basis** for a subspace W is an orthogonal basis $\{\vec{u}_1, \ldots, \vec{u}_p\}$ in which every vector $\vec{u}_q$ has unit length. In this case, for each $\vec{w} \in W$,

$$\vec{w} = (\vec{w} \cdot \vec{u}_1)\vec{u}_1 + \cdots + (\vec{w} \cdot \vec{u}_p)\vec{u}_p$$

$$\|\vec{w}\| = \sqrt{(\vec{w} \cdot \vec{u}_1)^2 + \cdots + (\vec{w} \cdot \vec{u}_p)^2}$$

Example

The subspace W is a subspace of $\mathbb{R}^3$ perpendicular to $x = (1, 1, 1)$. Calculate the missing coefficients in the orthonormal basis for W.

$$u = \frac{1}{\sqrt{}} \begin{bmatrix} 1 \\ 0 \\ \end{bmatrix} \qquad v = \frac{1}{\sqrt{}} \begin{bmatrix} \\ \\ \end{bmatrix}$$

Orthogonal Matrices

An **orthogonal matrix** is a square matrix whose columns are orthonormal.

Theorem

An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I_n$.

Can U have orthonormal columns if $n > m$?

Theorem

Theorem (Mapping Properties of Orthogonal Matrices)

Assume $m \times m$ matrix U has orthonormal columns. Then

1. (Preserves length) $\|U\vec{x}\| = \boxed{}$
2. (Preserves angles) $(U\vec{x}) \cdot (U\vec{y}) = \boxed{}$
3. (Preserves orthogonality)

Example

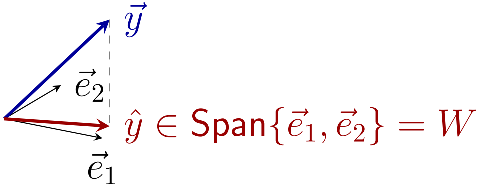
Compute the length of the vector below.

$$\begin{bmatrix} 1/2 & 2/\sqrt{14} \\ 1/2 & 1/\sqrt{14} \\ 1/2 & -3/\sqrt{14} \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ -3 \end{bmatrix}$$

Section 6.3 : Orthogonal Projections

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



Vectors $\vec{e}_1$ and $\vec{e}_2$ form an orthonormal basis for subspace W.
Vector $\vec{y}$ is not in W.
The orthogonal projection of $\vec{y}$ onto W =Span$\{\vec{e}_1, \vec{e}_2\}$ is $\hat{y}$.

Topics and Objectives

Topics

1. Orthogonal projections and their basic properties
2. Best approximations

Learning Objectives

1. Apply concepts of orthogonality and projections to
   a) compute orthogonal projections and distances,
   b) express a vector as a linear combination of orthogonal vectors,
   c) construct vector approximations using projections,
   d) characterize bases for subspaces of $\mathbb{R}^n$, and
   e) construct orthonormal bases.

Motivating Question For the matrix A and vector $\vec{b}$, which vector $\widehat{b}$ in column space of A, is closest to $\vec{b}$?

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -4 & -2 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Example 1

Let $\vec{u}_1, \ldots, \vec{u}_5$ be an orthonormal basis for $\mathbb{R}^5$. Let $W = \text{Span}\{\vec{u}_1, \vec{u}_2\}$. For a vector $\vec{y} \in \mathbb{R}^5$, write $\vec{y} = \widehat{y} + w^\perp$, where $\widehat{y} \in W$ and $w^\perp \in W^\perp$.

Orthogonal Decomposition Theorem

Theorem

Let W be a subspace of $\mathbb{R}^n$. Then, each vector $\vec{y} \in \mathbb{R}^n$ has the **unique** decomposition

$$\vec{y} = \widehat{y} + w^{\perp}, \quad \widehat{y} \in W, \quad w^{\perp} \in W^{\perp}.$$

And, if $\vec{u}_1, \ldots, \vec{u}_p$ is any orthogonal basis for W,

$$\hat{y} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \cdots + \frac{\vec{y} \cdot \vec{u}_p}{\vec{u}_p \cdot \vec{u}_p}\vec{u}_p.$$

We say that $\widehat{y}$ is the **orthogonal projection of $\vec{y}$ onto W.**

If time permits, we will explain some of this theorem on the next slide.

Explanation (if time permits)

We can write

$$\widehat{y} =$$

Then, $w^{\perp} = \vec{y} - \widehat{y}$ is in $W^{\perp}$ because

Example 2a

$$\vec{y} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Construct the decomposition $\vec{y} = \widehat{y} + w^{\perp}$, where $\widehat{y}$ is the orthogonal projection of $\vec{y}$ onto the subspace $W = \mathrm{Span}\{\vec{u}_1, \vec{u}_2\}$.

Best Approximation Theorem

Theorem

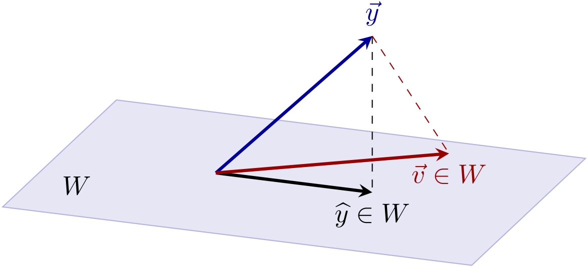
Let W be a subspace of $\mathbb{R}^n$, $\vec{y} \in \mathbb{R}^n$, and $\widehat{y}$ is the orthogonal projection of $\vec{y}$ onto W. Then for **any** $\vec{w} \neq \hat{y} \in W$, we have

$$\|\vec{y} - \widehat{y}\| < \|\vec{y} - \vec{w}\|$$

That is, $\widehat{y}$ is the unique vector in W that is closest to $\vec{y}$.

Proof (if time permits)

The orthogonal projection of $\vec{y}$ onto W is the closest point in W to $\vec{y}$.



Example 2b

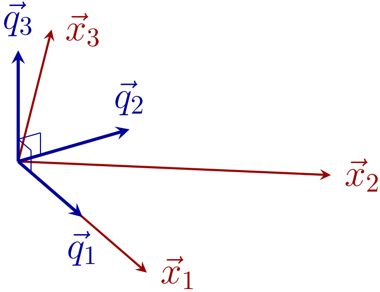
$$\vec{y} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

What is the distance between $\vec{y}$ and subspace $W = \mathrm{Span}\{\vec{u}_1, \vec{u}_2\}$? Note that these vectors are the same vectors that we used in Example 2a.

Section 6.4 : The Gram-Schmidt Process

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



Vectors $\vec{x}_1, \vec{x}_2, \vec{x}_3$ are given linearly independent vectors. We wish to construct an orthonormal basis $\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}$ for the space that they span.

Topics and Objectives

Topics

1. Gram Schmidt Process
2. The QR decomposition of matrices and its properties

Learning Objectives

1. Apply the iterative Gram Schmidt Process, and the QR decomposition, to construct an orthogonal basis.
2. Compute the QR factorization of a matrix.

Motivating Question The vectors below span a subspace W of $\mathbb{R}^4$. Identify an orthogonal basis for W.

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Example

The vectors below span a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

The Gram-Schmidt Process

Given a basis $\{\vec{x}_1, \dots, \vec{x}_p\}$ for a subspace W of $\mathbb{R}^n$, iteratively define

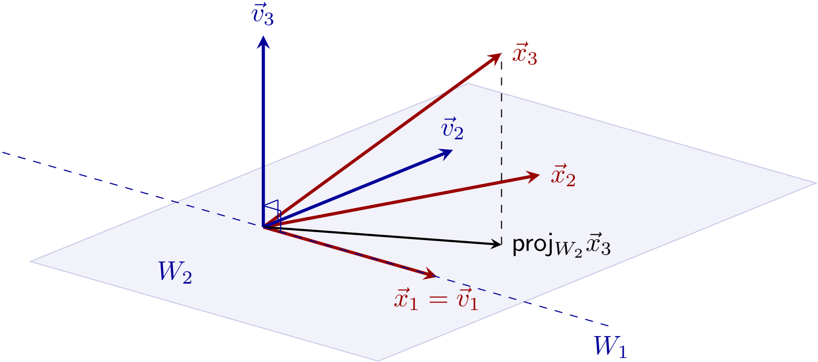
$$\begin{aligned}
\vec{v}_1 &= \vec{x}_1 \\
\vec{v}_2 &= \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 \\
\vec{v}_3 &= \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 \\
&\vdots \\
\vec{v}_p &= \vec{x}_p - \frac{\vec{x}_p \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \cdots - \frac{\vec{x}_p \cdot \vec{v}_{p-1}}{\vec{v}_{p-1} \cdot \vec{v}_{p-1}}\vec{v}_{p-1}
\end{aligned}$$

Then, $\{\vec{v}_1, \dots, \vec{v}_p\}$ is an orthogonal basis for W.

Proof

Geometric Interpretation

Suppose $\vec{x}_1, \vec{x}_2, \vec{x}_3$ are linearly independent vectors in $\mathbb{R}^3$. We wish to construct an orthogonal basis for the space that they span.



We construct vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$, which form our **orthogonal** basis.
$W_1 = \text{Span}\{\vec{v}_1\}$, $W_2 = \text{Span}\{\vec{v}_1, \vec{v}_2\}$.

Orthonormal Bases

Definition

A set of vectors form an **orthonormal basis** if the vectors are mutually orthogonal and have unit length.

Example

The two vectors below form an orthogonal basis for a subspace W. Obtain an orthonormal basis for W.

$$\vec{v}_1 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}.$$

QR Factorization

Theorem

Any $m \times n$ matrix A with linearly independent columns has the **QR factorization**

$$A = QR$$

where

1. Q is $m \times n$, its columns are an orthonormal basis for $\operatorname{Col} A$.
2. R is $n \times n$, upper triangular, with positive entries on its diagonal, and the length of the j^{th} column of R is equal to the length of the j^{th} column of A.

In the interest of time:

- we will not consider the case where A has linearly dependent columns
- students are not expected to know the conditions for which A has a QR factorization

Proof

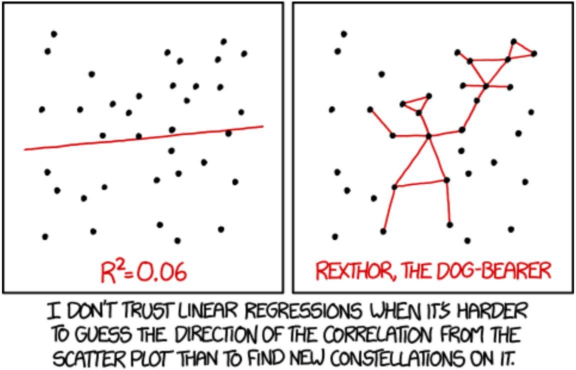
Example

Construct the QR decomposition for $A = \begin{bmatrix} 3 & -2 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$.

Section 6.5 : Least-Squares Problems

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



https://xkcd.com/1725

Topics and Objectives

Topics

1. Least Squares Problems
2. Different methods to solve Least Squares Problems

Learning Objectives

1. Compute general solutions, and least squares errors, to least squares problems using the normal equations and the QR decomposition.

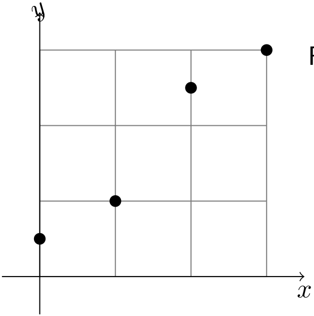
Motivating Question A series of measurements are corrupted by random errors. How can the dominant trend be extracted from the measurements with random error?

Inconsistent Systems

Suppose we want to construct a line of the form

$$y = mx + b$$

that best fits the data below.



From the data, we can construct the system:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 2.5 \\ 3 \end{bmatrix}$$

Can we ‘solve’ this inconsistent system?

The Least Squares Solution to a Linear System

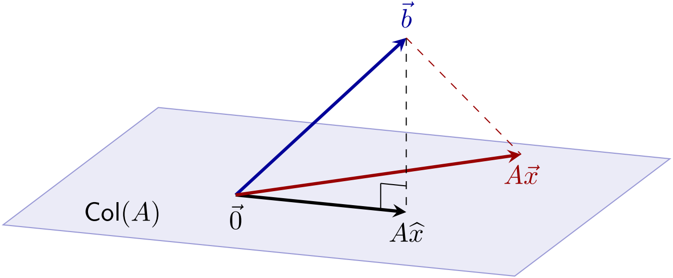
Definition: Least Squares Solution

Let A be a $m \times n$ matrix. A **least squares solution to** $A\vec{x} = \vec{b}$ is the solution $\widehat{x}$ for which

$$\| \vec{b} - A\widehat{x} \| \leq \| \vec{b} - A\vec{x} \|$$

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

A Geometric Interpretation

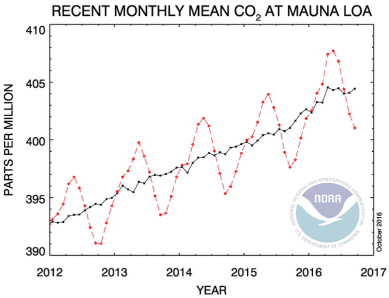


The vector $\vec{b}$ is closer to $A\hat{x}$ than to $A\vec{x}$ for all other $\vec{x} \in$ Col$A$.

1. If $\vec{b} \in \mathrm{Col}\, A$, then $\widehat{x}$ is ...
2. Seek $\widehat{x}$ so that $A\widehat{x}$ is as close to $\vec{b}$ as possible. That is, $\widehat{x}$ should solve $A\widehat{x} = \widehat{b}$ where $\widehat{b}$ is ...

Important Examples: Overdetermined Systems (Tall/Thin Matrices)

A variety of factors impact the measured quantity.



In the above figure, the dashed red line with diamond symbols represents the monthly mean values, centered on the middle of each month. The black line with the square symbols represents the same, after correction for the average seasonal cycle. (NOAA graph.)

Previous data is the important time series of mean CO_2 in the atmosphere. The data is collected at the Mauna Loa observatory on the island of Hawaii (The Big Island). One of the most important observatories in the world, it is located at the top of the Mauna Kea volcano, 4,205 meters altitude.

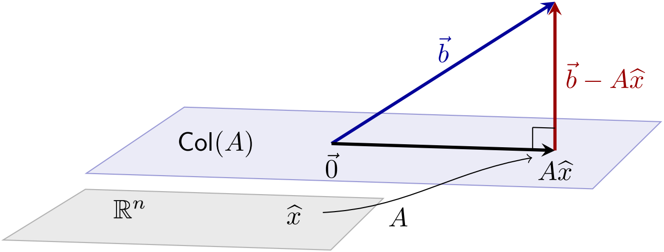
The Normal Equations

Theorem (Normal Equations for Least Squares)

The least squares solutions to $A\vec{x} = \vec{b}$ coincide with the solutions to

$$\underbrace{A^T A\vec{x} = A^T\vec{b}}_{\text{Normal Equations}}$$

Derivation



The least-squares solution $\hat{x}$ is in $\mathbb{R}^n$.

1. $\widehat{x}$ is the least squares solution, is equivalent to $\vec{b} - A\widehat{x}$ is orthogonal to $\boxed{}A$.
2. A vector $\vec{v}$ is in $\mathrm{Null}\, A^T$ if and only if $\boxed{}\vec{v} = \vec{0}$.
3. So we obtain the Normal Equations:

Example

Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Solution:

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} =$$

$$A^T \vec{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} =$$

The normal equations $A^T A\vec{x} = A^T\vec{b}$ become:

Theorem

Theorem (Unique Solutions for Least Squares)

Let A be any $m \times n$ matrix. These statements are equivalent.

1. The equation $A\vec{x} = \vec{b}$ has a unique least-squares solution for each $\vec{b} \in \mathbb{R}^m$.
2. The columns of A are linearly independent.
3. The matrix $A^T A$ is invertible.

And, if these statements hold, the least square solution is

$$\widehat{x} = (A^T A)^{-1} A^T \vec{b}.$$

Useful heuristic: $A^T A$ plays the role of 'length-squared' of the matrix A. (See the sections on symmetric matrices and singular value decomposition.)

Example

Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Hint: the columns of A are orthogonal.

Theorem (Least Squares and QR)

Let $m \times n$ matrix A have a QR decomposition. Then for each $\vec{b} \in \mathbb{R}^m$ the equation $A\vec{x} = \vec{b}$ has the unique least squares solution

$$R\widehat{x} = Q^T\vec{b}.$$

(Remember, R is upper triangular, so the equation above is solved by back-substitution.)

Example 3. Compute the least squares solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

Solution. The QR decomposition of A is

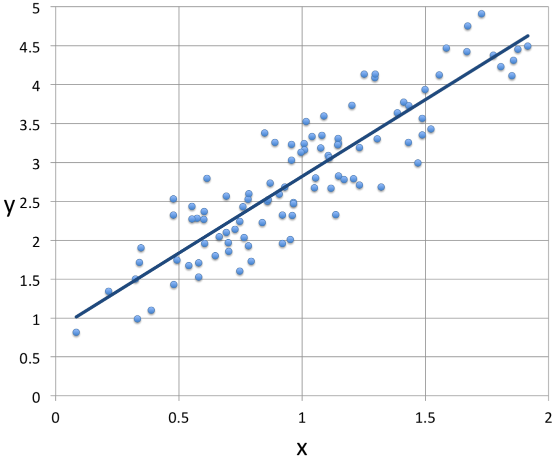
$$A = QR = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$$

$$Q^T\vec{b} = \frac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & -1 & -1 & 1\\ 1 & -1 & 1 & -1\end{bmatrix}\begin{bmatrix}3\\ 5\\ 7\\ -3\end{bmatrix} = \begin{bmatrix} \\ -6\\ 4\end{bmatrix}$$

And then we solve by backwards substitution $R\vec{x} = Q^T\vec{b}$

$$\begin{bmatrix}2 & 4 & 5\\ 0 & 2 & 3\\ 0 & 0 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} = \begin{bmatrix} \\ -6\\ 4\end{bmatrix}$$

Chapter 6 : Orthogonality and Least Squares
6.6 : Applications to Linear Models



Topics and Objectives

Topics

1. Least Squares Lines
2. Linear and more complicated models

Learning Objectives

For the topics covered in this section, students are expected to be able to do the following.

1. Apply least-squares and multiple regression to construct a linear model from a set of data points.
2. Apply least-squares to fit polynomials and other curves to data.

Motivating Question

Compute the equation of the line $y = \beta_0 + \beta_1 x$ that best fits the data

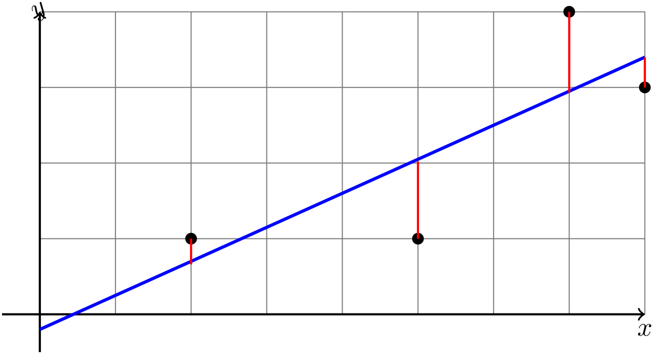
x	2	5	7	8
y	1	1	4	3

The Least Squares Line

Graph below gives an approximate linear relationship between x and y.

1. Black circles are data.
2. Blue line is the **least squares** line.
3. Lengths of red lines are the ___________.

The least squares line minimizes the sum of squares of the ___________.



Example 1 Compute the least squares line $y = \beta_0 + \beta_1 x$ that best fits the data

x	2	5	7	8
y	1	1	4	3

We want to solve

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

This is a least-squares problem : $X\vec{\beta} = \vec{y}$.

The normal equations are

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ & & & \end{bmatrix} \begin{bmatrix} 1 & \\ 1 & \\ 1 & \\ 1 & \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ & & & \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

So the least-squares solution is given by

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 59 \end{bmatrix}$$

$$y = \beta_0 + \beta_1 x = \frac{-5}{21} + \frac{19}{42} x$$

As we may have guessed, β_0 is negative, and β_1 is positive.

Least Squares Fitting for Other Curves

We can consider least squares fitting for the form

$$y = c_0 + c_1 f_1(x) + c_2 f_2(x) + \cdots + c_k f_k(x).$$

If functions f_i are known, this is a linear problem in the c_i variables.

Example

Consider the data in the table below.

x	-1	0	0	1
y	2	1	0	6

Determine the coefficients c_1 and c_2 for the curve $y = c_1 x + c_2 x^2$ that best fits the data.

WolframAlpha and Mathematica Syntax

Least squares problems can be computed with WolframAlpha, Mathematica, and many other software.

WolframAlpha

$$\texttt{linear fit } \{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_n, y_n\}\}$$

Mathematica

$$\texttt{LeastSquares}[\{\{x_1, x_1, y_1\}, \{x_2, x_2, y_2\}, \ldots, \{x_n, x_n, y_n\}\}]$$

Almost any spreadsheet program does this as a function as well.