MT-1004
Linear Algebra

Fall 2023

Week # 10-11

National University of Computer and Emerging Sciences

November 2, 2023

Chapter 6

Orthogonality and Least Squares

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect.

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

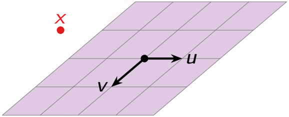
Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.

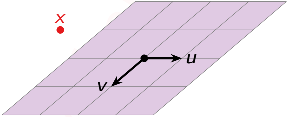


Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



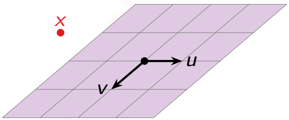
Due to measurement error, though, the measured x is not actually in $\text{Span}\{u, v\}$.

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



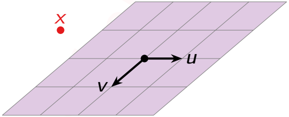
Due to measurement error, though, the measured x is not actually in $\text{Span}\{u, v\}$. In other words, the equation $au + bv = x$ has no solution ($x \in$ Span of u and v).

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



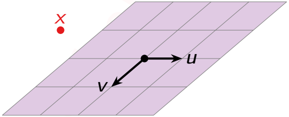
Due to measurement error, though, the measured x is not actually in Span$\{u, v\}$. In other words, the equation $au + bv = x$ has no solution ($x \in$ Span of u and v). What do you do?

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



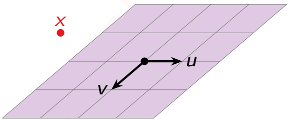
Due to measurement error, though, the measured x is not actually in $\text{Span}\{u, v\}$. In other words, the equation $au + bv = x$ has no solution ($x \in$ Span of u and v). What do you do? The real value is probably the *closest* point to x on $\text{Span}\{u, v\}$.

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



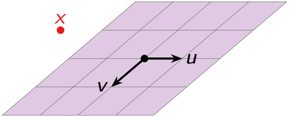
Due to measurement error, though, the measured x is not actually in $\text{Span}\{u, v\}$. In other words, the equation $au + bv = x$ has no solution ($x \in$ Span of u and v). What do you do? The real value is probably the *closest* point to x on $\text{Span}\{u, v\}$. Which point is that?

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



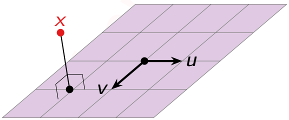
Due to measurement error, though, the measured x is not actually in $\text{Span}\{u, v\}$. In other words, the equation $au + bv = x$ has no solution ($x \in$ Span of u and v). What do you do? The real value is probably the *closest* point to x on $\text{Span}\{u, v\}$. Which point is that?

Motivation

Recall: This course is about learning to:

- Solve the matrix equation $Ax = b$ (Echelon Form, Reduced Echelon Form, Column Space, Null Space)
- Solve the matrix equation $Ax = \lambda x$ (Eigenvalues, eigenvectors and their applications)
- Almost solve the equation $Ax = b$ (We are going to study)

Idea 1: In the real world, data is imperfect. Suppose you measure a data point x which you know for theoretical reasons must lie on a plane spanned by two vectors u and v.



Due to measurement error, though, the measured x is not actually in $\text{Span}\{u, v\}$. In other words, the equation $au + bv = x$ has no solution ($x \in$ Span of u and v). What do you do? The real value is probably the *closest* point to x on $\text{Span}\{u, v\}$. Which point is that?

The Dot Product

Purpose To find *length* and *angle* between two vectors, and in particular, to understand the notion of *orthogonality* (i.e. perpendicularity).

The Dot Product

Purpose To find *length* and *angle* between two vectors, and in particular, to understand the notion of *orthogonality* (i.e. perpendicularity).

Definition

The **dot product** of two vectors x, y in R^n is

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

The Dot Product

Purpose To find *length* and *angle* between two vectors, and in particular, to understand the notion of *orthogonality* (i.e. perpendicularity).

Definition

The **dot product** of two vectors x, y in R^n is

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = y^T x.$$

The Dot Product

Purpose To find *length* and *angle* between two vectors, and in particular, to understand the notion of *orthogonality* (i.e. perpendicularity).

Definition

The **dot product** of two vectors x, y in R^n is

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = y^T x.$$

Example

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32.$$

Properties of the Dot Product

- $x \cdot x \geq 0$

Properties of the Dot Product

- $x \cdot x \geq 0$
- $x \cdot x = 0$ if and only if $x = 0$.
- $x \cdot y = y \cdot x$

Properties of the Dot Product

- $x \cdot x \geq 0$
- $x \cdot x = 0$ if and only if $x = 0$.
- $x \cdot y = y \cdot x$
- $(x + y) \cdot z = x \cdot z + y \cdot z$

Properties of the Dot Product

- $x \cdot x \geq 0$
- $x \cdot x = 0$ if and only if $x = 0$.
- $x \cdot y = y \cdot x$
- $(x + y) \cdot z = x \cdot z + y \cdot z$
- $(cx) \cdot y = c(x \cdot y)$

The Dot Product and Length

Definition

The **length** or **norm** of a vector x in R^n is

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

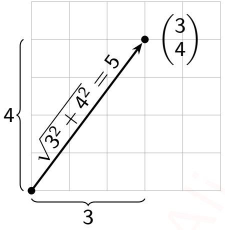
The Dot Product and Length

Definition

The **length** or **norm** of a vector x in R^n is

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

The Pythagorean theorem!



$$\left\|\begin{pmatrix}3\\4\end{pmatrix}\right\| = \sqrt{3^2+4^2} = 5$$

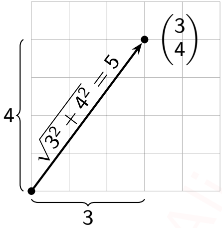
The Dot Product and Length

Definition

The **length** or **norm** of a vector x in R^n is

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

The Pythagorean theorem!



$$\left\|\begin{pmatrix}3\\4\end{pmatrix}\right\| = \sqrt{3^2+4^2} = 5$$

Fact

If x is a vector and c is a scalar, then $\|cx\| = |c| \cdot \|x\|$.

The Dot Product and Distance

Definition

The **distance** between two points x, y in R^n is

$$\text{dist}(x, y) =$$

The Dot Product and Distance

Definition

The **distance** between two points x, y in R^n is

$$\mathrm{dist}(x, y) = \|y - x\|.$$

The Dot Product and Distance

Definition

The **distance** between two points x, y in R^n is

$$\text{dist}(x, y) = \|y - x\|.$$

This is just the length of the vector from x to y.

The Dot Product and Distance

Definition

The **distance** between two points x, y in R^n is

$$\mathrm{dist}(x, y) = \|y - x\|.$$

This is just the length of the vector from x to y.

Example

Let $x = (1, 2)$ and $y = (4, 4)$. Then

$$\mathrm{dist}(x, y) = \|y - x\| = \left\| \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right\| = \sqrt{3^2 + 2^2} = \sqrt{13}.$$

The Dot Product and Distance

Definition

The **distance** between two points x, y in $\mathbb{R}^n$ is

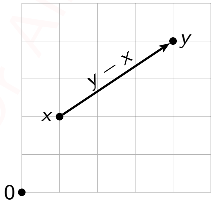
$$\operatorname{dist}(x, y) = \|y - x\|.$$

This is just the length of the vector from x to y.

Example

Let $x = (1, 2)$ and $y = (4, 4)$. Then

$$\operatorname{dist}(x, y) = \|y - x\| = \left\| \begin{pmatrix} 3 \\ 2 \end{pmatrix} \right\| = \sqrt{3^2 + 2^2} = \sqrt{13}.$$



Unit Vectors

Definition

A **unit vector** is a vector v with length $\|v\| = 1$.

Unit Vectors

Definition

A **unit vector** is a vector v with length $\|v\| = 1$.

Example

The unit coordinate vectors are unit vectors:

$$\|e_1\| = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\| = \sqrt{1^2 + 0^2 + 0^2} = 1$$

Unit Vectors

Definition

A **unit vector** is a vector v with length $\|v\| = 1$.

Example

The unit coordinate vectors are unit vectors:

$$\|e_1\| = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\| = \sqrt{1^2 + 0^2 + 0^2} = 1$$

Definition

Let x be a nonzero vector in R^n. The **unit vector in the direction of** x is the vector $\dfrac{x}{\|x\|}$.

Unit Vectors

Definition

A **unit vector** is a vector v with length $\|v\| = 1$.

Example

The unit coordinate vectors are unit vectors:

$$\|e_1\| = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\| = \sqrt{1^2 + 0^2 + 0^2} = 1$$

Definition

Let x be a nonzero vector in R^n. The **unit vector in the direction of** x is the vector $\dfrac{x}{\|x\|}$.

This is in fact a unit vector:

scalar

$$\left\| \frac{x}{\|x\|} \right\| = \frac{1}{\|x\|}\|x\| = 1.$$

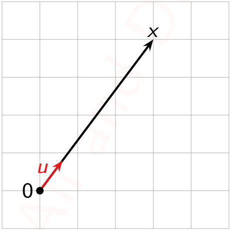
Unit Vectors

Example

Example

What is the unit vector in the direction of $x = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$?

$$u = \frac{x}{\|x\|} = \frac{1}{\sqrt{3^2 + 4^2}} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$



Orthogonality

Definition

Two vectors x, y are **orthogonal** or **perpendicular** if $x \cdot y = 0$.

Orthogonality

Definition

Two vectors x, y are **orthogonal** or **perpendicular** if $x \cdot y = 0$.
Notation: $x \perp y$ means $x \cdot y = 0$.

Orthogonality

Definition

Two vectors x, y are **orthogonal** or **perpendicular** if $x \cdot y = 0$.
Notation: $x \perp y$ means $x \cdot y = 0$.

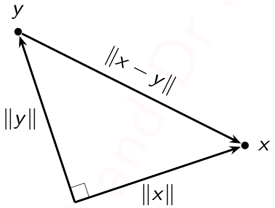
Why is this a good definition?

Orthogonality

Definition

Two vectors x, y are **orthogonal** or **perpendicular** if $x \cdot y = 0$.
Notation: $x \perp y$ means $x \cdot y = 0$.

Why is this a good definition? The Pythagorean theorem / law of cosines!



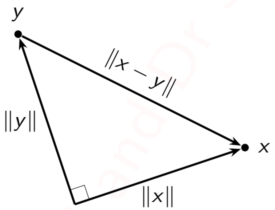
$$
\begin{aligned}
x \text{ and } y \text{ are perpendicular} &\iff \|x\|^2 + \|y\|^2 = \|x-y\|^2 \\
&\iff x\cdot x + y\cdot y = (x-y)\cdot(x-y) \\
&\iff x\cdot x + y\cdot y = x\cdot x + y\cdot y - 2x\cdot y \\
&\iff x\cdot y = 0
\end{aligned}
$$

Orthogonality

Definition

Two vectors x, y are **orthogonal** or **perpendicular** if $x \cdot y = 0$.
Notation: $x \perp y$ means $x \cdot y = 0$.

Why is this a good definition? The Pythagorean theorem / law of cosines!



$$\begin{aligned} x \text{ and } y \text{ are perpendicular} &\iff \|x\|^2 + \|y\|^2 = \|x - y\|^2 \\ &\iff x \cdot x + y \cdot y = (x - y) \cdot (x - y) \\ &\iff x \cdot x + y \cdot y = x \cdot x + y \cdot y - 2x \cdot y \\ &\iff x \cdot y = 0 \end{aligned}$$

Fact: $x \perp y \iff \|x - y\|^2 = \|x\|^2 + \|y\|^2$

Orthogonality

Definition

A set of vectors $\{x_1, x_2, ..., x_k\}$ in R^n is called an **orthogonal set** if $x_i \cdot x_j = 0$ whenever $i \neq j$ for $i, j = 1, 2, ..., k$.

Problem: Show that $\{x_1,\ x_2,\ x_3\}$ is an orthogonal set in R^3 if

$$x_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix},\ x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},\ x_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

Orthogonality

Definition

A set of vectors $\{x_1, x_2, ..., x_k\}$ in R^n is called an **orthogonal set** if $x_i \cdot x_j = 0$ whenever $i \neq j$ for $i, j = 1, 2, ..., k$.

Problem: Show that $\{x_1,\ x_2,\ x_3\}$ is an orthogonal set in R^3 if

$$x_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix},\ x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},\ x_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

$$x_1 \cdot x_2 = 2(0) + 1(1) + (-1)(1) = 0$$
$$x_2 \cdot x_3 = 0(1) + 1(-1) + (1)(1) = 0$$
$$x_1 \cdot x_3 = 2(1) + 1(-1) + (-1)(1) = 0$$

Orthogonality

Example

Problem: Find *all* vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

We have to find all vectors x such that $x \cdot v = 0$.

Orthogonality

Example

Problem: Find *all* vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

We have to find all vectors x such that $x \cdot v = 0$. This means solving the equation

$$0 = x \cdot v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3.$$

Problem: Find *all* vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

We have to find all vectors x such that $x \cdot v = 0$. This means solving the equation

$$0 = x \cdot v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3.$$

The parametric form for the solution is $x_1 = -x_2 + x_3$, so the parametric vector form of the general solution is

Orthogonality

Example

Problem: Find *all* vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

We have to find all vectors x such that $x \cdot v = 0$. This means solving the equation

$$0 = x \cdot v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3.$$

The parametric form for the solution is $x_1 = -x_2 + x_3$, so the parametric vector form of the general solution is

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Orthogonality

Example

Problem: Find *all* vectors orthogonal to $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

We have to find all vectors x such that $x \cdot v = 0$. This means solving the equation

$$0 = x \cdot v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3.$$

The parametric form for the solution is $x_1 = -x_2 + x_3$, so the parametric vector form of the general solution is

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

For instance, $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \perp \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ because $\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = 0.$

Problem: Find *all* vectors orthogonal to both $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Now we have to solve the system of two homogeneous equations

$$0 = x \cdot v = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = x_1 + x_2 - x_3$$
$$0 = x \cdot w = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = x_1 + x_2 + x_3.$$

In matrix form:

$$\text{The rows are } v \text{ and } w \longrightarrow \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The parametric vector form of the solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Orthogonal Basis

Theorem

If $\{u_1, u_2, ..., u_k\}$ is an orthogonal set of nonzero vectors in R^n, then these vectors are linearly independent.

Orthogonal Basis

Theorem

If $\{u_1, u_2, ..., u_k\}$ is an orthogonal set of nonzero vectors in R^n, then these vectors are linearly independent.

Suppose $\{u_1, u_2, \ldots, u_m\}$ is orthogonal. We need to show that the equation

$$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m = 0$$

has only the trivial solution $c_1 = c_2 = \cdots = c_m = 0$.

$$0 = u_1 \cdot (c_1 u_1 + c_2 u_2 + \cdots + c_m u_m) = c_1(u_1 \cdot u_1) + 0 + 0 + \cdots + 0.$$

Hence $c_1 = 0$. Similarly for the other c_i.

Orthogonal Basis

Theorem
If $\{u_1, u_2, ..., u_k\}$ is an orthogonal set of nonzero vectors in R^n, then these vectors are linearly independent.

Suppose $\{u_1, u_2, \ldots, u_m\}$ is orthogonal. We need to show that the equation

$$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m = 0$$

has only the trivial solution $c_1 = c_2 = \cdots = c_m = 0$.

$$0 = u_1 \cdot (c_1 u_1 + c_2 u_2 + \cdots + c_m u_m) = c_1(u_1 \cdot u_1) + 0 + 0 + \cdots + 0.$$

Hence $c_1 = 0$. Similarly for the other c_i.

Definition
An **orthogonal basis** for a subspace W of R^n is a basis of W that is an orthogonal set.

Orthogonal Basis

Theorem

If $\{u_1, u_2, ..., u_k\}$ is an orthogonal set of nonzero vectors in R^n, then these vectors are linearly independent.

Suppose $\{u_1, u_2, \ldots, u_m\}$ is orthogonal. We need to show that the equation

$$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m = 0$$

has only the trivial solution $c_1 = c_2 = \cdots = c_m = 0$.

$$0 = u_1 \cdot (c_1 u_1 + c_2 u_2 + \cdots + c_m u_m) = c_1(u_1 \cdot u_1) + 0 + 0 + \cdots + 0.$$

Hence $c_1 = 0$. Similarly for the other c_i.

Definition

An **orthogonal basis** for a subspace W of R^n is a basis of W that is an orthogonal set.

Problem Show that $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ forms orthogonal basis of R^2.

Orthogonal Basis
Example

Problem Find an orthogonal basis for the subspace W of R^3 given by

$$W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x - y + 2z = 0 \right\}.$$

Orthogonal Basis

Example

Problem Find an orthogonal basis for the subspace W of R^3 given by

$$W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x - y + 2z = 0 \right\}.$$

We have already studied how to calculate the basis of W.

$$\begin{bmatrix} y - 2z \\ y \\ z \end{bmatrix} = y \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix},$$

so, $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ forms basis of W

Orthogonal Basis

Example

Problem Find an orthogonal basis for the subspace W of $\mathbb{R}^3$ given by

$$W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x - y + 2z = 0 \right\}.$$

We have already studied how to calculate the basis of W.

$$\begin{bmatrix} y - 2z \\ y \\ z \end{bmatrix} = y \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix},$$

so, $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ forms basis of W

They are not orthogonal.

Orthogonal Basis

We want to find a vector of W that is orthogonal to either $\begin{bmatrix}1\\1\\0\end{bmatrix}$ or $\begin{bmatrix}-2\\0\\1\end{bmatrix}$.

Orthogonal Basis

We want to find a vector of W that is orthogonal to either $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

Let $w = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a vector of W and is orthogonal to $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

Orthogonal Basis

We want to find a vector of W that is orthogonal to either $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

Let $w = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a vector of W and is orthogonal to $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

As $w \in W$, so we can write

$$x - y + 2z = 0.$$

Orthogonal Basis

We want to find a vector of W that is orthogonal to either $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

Let $w = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a vector of W and is orthogonal to $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

As $w \in W$, so we can write

$$x - y + 2z = 0.$$

Orthogonality condition implies

$$-2x + z = 0.$$

Orthogonal Basis

We want to find a vector of W that is orthogonal to either $\begin{bmatrix}1\\1\\0\end{bmatrix}$ or $\begin{bmatrix}-2\\0\\1\end{bmatrix}$.

Let $w = \begin{bmatrix}x\\y\\z\end{bmatrix}$ be a vector of W and is orthogonal to $\begin{bmatrix}-2\\0\\1\end{bmatrix}$.

As $w \in W$, so we can write

$$x - y + 2z = 0.$$

Orthogonality condition implies

$$-2x + z = 0.$$

Solving above two equations, we get

$$y = 5x, \; z = 2x.$$

So,

$$w = x\begin{bmatrix}1\\5\\2\end{bmatrix}.$$

Orthogonal Basis

We want to find a vector of W that is orthogonal to either $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

Let $w = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be a vector of W and is orthogonal to $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$.

As $w \in W$, so we can write

$$x - y + 2z = 0.$$

Orthogonality condition implies

$$-2x + z = 0.$$

Solving above two equations, we get

$$y = 5x, \; z = 2x.$$

So,

$$w = x \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}.$$

So, $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix}$ forms orthogonal basis of W.

Orthogonal Basis

Theorem

Let $\{x_1, x_2, ..., x_k\}$ be an orthogonal basis for a subspace W of R^n and let w be any vector in W. Then the unique scalars $c_1, c_2, ..., c_k$ such that

$$w = c_1x_1 + c_2x_2 + ... + c_kx_k$$

are given by

$$c_i = \frac{w \cdot x_i}{x_i \cdot x_i}, \quad i = 1, 2, ..., k.$$

Orthogonal Basis

Example

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

Orthogonal Basis

Example

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

As $v_1 \cdot v_2 = (4).(1) + (2)(-2) = 0,$

Orthogonal Basis

Example

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

As $v_1 \cdot v_2 = (4).(1) + (2)(-2) = 0$, so, v_1 and v_2 are orthogonal and hence forms basis of R^2.

Orthogonal Basis

Example

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

As $v_1 \cdot v_2 = (4).(1) + (2)(-2) = 0$, so, v_1 and v_2 are orthogonal and hence forms basis of R^2.
Now, for the coordinate vector

$$w = c_1 v_1 + c_2 v_2.$$

Orthogonal Basis
Example

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

As $v_1 \cdot v_2 = (4).(1) + (2)(-2) = 0$, so, v_1 and v_2 are orthogonal and hence forms basis of R^2.
Now, for the coordinate vector

$$w = c_1 v_1 + c_2 v_2.$$

Taking dot product with v_1,

Orthogonal Basis

Example

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

As $v_1 \cdot v_2 = (4).(1) + (2)(-2) = 0$, so, v_1 and v_2 are orthogonal and hence forms basis of R^2.
Now, for the coordinate vector

$$w = c_1 v_1 + c_2 v_2.$$

Taking dot product with v_1, we get

$$\begin{aligned} w \cdot v_1 &= c_1 v_1 \cdot v_1 \\ c_1 &= \frac{w \cdot v_1}{v_1 \cdot v_1} \end{aligned}$$

Orthogonal Basis

Example

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

As $v_1 \cdot v_2 = (4).(1) + (2)(-2) = 0$, so, v_1 and v_2 are orthogonal and hence forms basis of R^2.
Now, for the coordinate vector

$$w = c_1 v_1 + c_2 v_2.$$

Taking dot product with v_1, we get

$$w \cdot v_1 = c_1 v_1 \cdot v_1$$
$$c_1 = \frac{w \cdot v_1}{v_1 \cdot v_1}$$

Similarly, $c_2 = \dfrac{w \cdot v_2}{v_2 \cdot v_2}$.

Problem Show that $B = \left\{ v_1 = \begin{bmatrix} 4 \\ -2 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms basis of R^2 and write coordinate vector of $w = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ w.r.t B.

As $v_1 \cdot v_2 = (4).(1) + (2)(-2) = 0$, so, v_1 and v_2 are orthogonal and hence forms basis of R^2.
Now, for the coordinate vector

$$w = c_1 v_1 + c_2 v_2.$$

Taking dot product with v_1, we get

$$\begin{aligned} w \cdot v_1 &= c_1 v_1 \cdot v_1 \\ c_1 &= \frac{w \cdot v_1}{v_1 \cdot v_1} \end{aligned}$$

Similarly, $c_2 = \dfrac{w \cdot v_2}{v_2 \cdot v_2}$.
As $v_1 \cdot v_1 = 12,\ v_2 \cdot v_2 = 5,\ w \cdot v_1 = 10,\ w \cdot v_2 = -5$. So, $c_1 = \frac{5}{6},\ c_2 = -1$.
Hence, coordinate vector of w is

$$\begin{bmatrix} \frac{5}{6} \\ -1 \end{bmatrix}.$$

Orthogonality

General procedure

Problem: Find all vectors orthogonal to some number of vectors $v_1, v_2, \ldots, v_m$ in R^n.

Orthogonality

General procedure

Problem: Find all vectors orthogonal to some number of vectors $v_1, v_2, \ldots, v_m$ in R^n.

This is the same as finding all vectors x such that

$$0 = v_1^T x = v_2^T x = \cdots = v_m^T x.$$

Problem: Find all vectors orthogonal to some number of vectors $v_1, v_2, \dots, v_m$ in R^n.

This is the same as finding all vectors x such that

$$0 = v_1^T x = v_2^T x = \cdots = v_m^T x.$$

Putting the *row* vectors $v_1^T, v_2^T, \dots, v_m^T$ into a matrix, this is the same as finding all x such that

$$\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix} x = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_m \cdot x \end{pmatrix} = 0.$$

Problem: Find all vectors orthogonal to some number of vectors $v_1, v_2, \ldots, v_m$ in $\mathbf{R}^n$.

This is the same as finding all vectors x such that

$$0 = v_1^T x = v_2^T x = \cdots = v_m^T x.$$

Putting the *row* vectors $v_1^T, v_2^T, \ldots, v_m^T$ into a matrix, this is the same as finding all x such that

$$\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix} x = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_m \cdot x \end{pmatrix} = 0.$$

Important

The set of all vectors orthogonal to some vectors $v_1, v_2, \ldots, v_m$ in $\mathbf{R}^n$ is the *null space* of the $m \times n$ matrix

$$\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}.$$

Orthogonality

General procedure

Problem: Find all vectors orthogonal to some number of vectors $v_1, v_2, \ldots, v_m$ in R^n.

This is the same as finding all vectors x such that

$$0 = v_1^T x = v_2^T x = \cdots = v_m^T x.$$

Putting the *row* vectors $v_1^T, v_2^T, \ldots, v_m^T$ into a matrix, this is the same as finding all x such that

$$\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix} x = \begin{pmatrix} v_1 \cdot x \\ v_2 \cdot x \\ \vdots \\ v_m \cdot x \end{pmatrix} = 0.$$

Important

The set of all vectors orthogonal to some vectors $v_1, v_2, \ldots, v_m$ in R^n is the *null space* of the $m \times n$ matrix

$$\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}.$$

In particular, this set is a subspace!

Orthogonal Complements

Definition

Let W be a subspace of R^n. Its **orthogonal complement** is

$$W^{\perp} = \{v \text{ in } \mathrm{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\} \qquad \text{read "}W\text{ perp".}$$

Orthogonal Complements

Definition

Let W be a subspace of $\mathbb{R}^n$. Its **orthogonal complement** is

$$W^{\perp} = \{v \text{ in } \mathbb{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\} \qquad \text{read "}W\text{ perp".}$$

$W^{\perp}$ is orthogonal complement
A^{T} is transpose

Orthogonal Complements

Definition

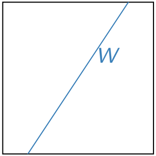
Let W be a subspace of $\mathbb{R}^n$. Its **orthogonal complement** is

$$W^{\perp} = \{v \text{ in } \mathbb{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\} \qquad \text{read “}W\text{ perp”}.$$

$W^{\perp}$ is orthogonal complement
A^T is transpose

Pictures:

The orthogonal complement of a line in $\mathbb{R}^2$ is



Orthogonal Complements

Definition

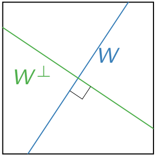
Let W be a subspace of R^n. Its **orthogonal complement** is

$$W^{\perp} = \{ v \text{ in } \mathrm{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W \} \qquad \text{read "}W\text{ perp".}$$

$W^{\perp}$ is orthogonal complement
A^T is transpose

Pictures:

The orthogonal complement of a line in R^2 is the perpendicular line.



Orthogonal Complements

Definition

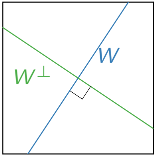
Let W be a subspace of R^n. Its **orthogonal complement** is

$$W^{\perp} = \{v \text{ in } \mathrm{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\} \qquad \text{read "}W\text{ perp".}$$

$W^{\perp}$ is orthogonal complement
A^{T} is transpose

Pictures:

The orthogonal complement of a line in R^2 is the perpendicular line.



The orthogonal complement of a line in R^3 is

Orthogonal Complements

Definition

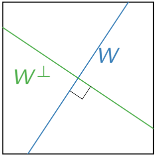
Let W be a subspace of R^n. Its **orthogonal complement** is

$$W^{\perp} = \{v \text{ in } \mathrm{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\} \qquad \text{read "}W\text{ perp".}$$

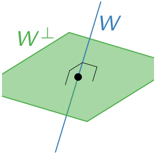
$W^{\perp}$ is orthogonal complement
A^T is transpose

Pictures:

The orthogonal complement of a line in R^2 is the perpendicular line.



The orthogonal complement of a line in R^3 is the perpendicular plane.



Orthogonal Complements

Definition

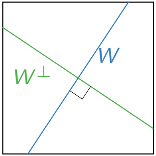
Let W be a subspace of $\mathbf{R}^n$. Its **orthogonal complement** is

$$W^{\perp} = \{v \text{ in } \mathbf{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\} \qquad \text{read “}W\text{ perp”}.$$

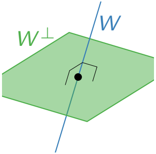
$W^{\perp}$ is orthogonal complement
A^T is transpose

Pictures:

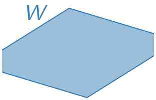
The orthogonal complement of a line in $\mathbf{R}^2$ is the perpendicular line.



The orthogonal complement of a line in $\mathbf{R}^3$ is the perpendicular plane.



The orthogonal complement of a plane in $\mathbf{R}^3$ is



Orthogonal Complements

Definition

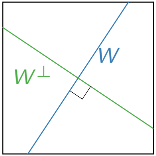
Let W be a subspace of $\mathbf{R}^n$. Its **orthogonal complement** is

$$W^{\perp} = \{ v \text{ in } \mathbf{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W \} \qquad \text{read "}W\text{ perp".}$$

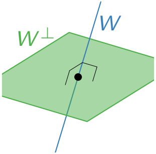
$W^{\perp}$ is orthogonal complement
A^T is transpose

Pictures:

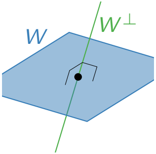
The orthogonal complement of a line in $\mathbf{R}^2$ is the perpendicular line.



The orthogonal complement of a line in $\mathbf{R}^3$ is the perpendicular plane.



The orthogonal complement of a plane in $\mathbf{R}^3$ is the perpendicular line.



Poll

Poll

Let W be a plane in $\mathbb{R}^4$. How would you describe $W^\perp$?

A. The zero space $\{0\}$.
B. A line in $\mathbb{R}^4$.
C. A plane in $\mathbb{R}^4$.
D. A 3-dimensional space in $\mathbb{R}^4$.
E. All of $\mathbb{R}^4$.

Orthogonal Complements

Basic properties

Let W be a subspace of $\mathbb{R}^n$.

Facts:

Orthogonal Complements

Basic properties

Let W be a subspace of R^n.

Facts:

1. $W^{\perp}$ is also a subspace of R^n

Orthogonal Complements

Basic properties

Let W be a subspace of R^n.

Facts:

1. $W^\perp$ is also a subspace of R^n
2. $(W^\perp)^\perp = W$

Orthogonal Complements

Basic properties

Let W be a subspace of R^n.

Facts:

1. $W^\perp$ is also a subspace of R^n
2. $(W^\perp)^\perp = W$
3. $\dim W + \dim W^\perp = n$

Let W be a subspace of R^n.

Facts:

1. $W^\perp$ is also a subspace of R^n
2. $(W^\perp)^\perp = W$
3. $\dim W + \dim W^\perp = n$
4. If $W = \mathrm{Span}\{v_1, v_2, \ldots, v_m\}$, then

$$\begin{aligned} W^\perp &= \text{all vectors orthogonal to each } v_1, v_2, \ldots, v_m \\ &= \{x \text{ in } \mathrm{R}^n \mid x \cdot v_i = 0 \text{ for all } i = 1, 2, \ldots, m\} \\ &= \mathrm{Nul}\begin{pmatrix} — v_1^T — \\ — v_2^T — \\ \vdots \\ — v_m^T — \end{pmatrix}. \end{aligned}$$

Let's check 1.

- Is 0 in $W^\perp$? Yes: $0 \cdot w = 0$ for any w in W.
- Suppose x, y are in $W^\perp$. So $x \cdot w = 0$ and $y \cdot w = 0$ for all w in W. Then $(x + y) \cdot w = x \cdot w + y \cdot w = 0 + 0 = 0$ for all w in W. So $x + y$ is also in $W^\perp$.
- Suppose x is in $W^\perp$. So $x \cdot w = 0$ for all w in W. If c is a scalar, then $(cx) \cdot w = c(x \cdot 0) = c(0) = 0$ for any w in W. So cx is in $W^\perp$.

Problem: if $W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$, compute $W^\perp$.

By property 4, we have to find the null space of the matrix whose rows are $\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$, which we did before:

$$\text{Nul}\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} = \text{Span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}.$$

Orthogonal Complements
Computation

Problem: if $W = \text{Span}\left\{\begin{pmatrix}1\\1\\-1\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$, compute $W^{\perp}$.

By property 4, we have to find the null space of the matrix whose rows are $\begin{pmatrix}1 & 1 & -1\end{pmatrix}$ and $\begin{pmatrix}1 & 1 & 1\end{pmatrix}$, which we did before:

$$\text{Nul}\begin{pmatrix}1 & 1 & -1\\1 & 1 & 1\end{pmatrix} = \text{Span}\left\{\begin{pmatrix}-1\\1\\0\end{pmatrix}\right\}.$$

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^{\perp} = \text{Nul}\begin{pmatrix}— v_1^T —\\— v_2^T —\\\vdots\\— v_m^T —\end{pmatrix}$$

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted Row A.

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted $\operatorname{Row} A$. Equivalently, it is the column span of A^T:

$$\operatorname{Row} A = \operatorname{Col} A^T.$$

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted $\operatorname{Row} A$. Equivalently, it is the column span of A^T:

$$\operatorname{Row} A = \operatorname{Col} A^T.$$

It is a subspace of R–.

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted $\operatorname{Row} A$. Equivalently, it is the column span of A^T:

$$\operatorname{Row} A = \operatorname{Col} A^T.$$

It is a subspace of R^n.

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted Row A. Equivalently, it is the column span of A^T:

$$\text{Row } A = \text{Col } A^T.$$

It is a subspace of $\mathbb{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp =$$

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted Row A. Equivalently, it is the column span of A^T:

$$\text{Row}\, A = \text{Col}\, A^T.$$

It is a subspace of $\mathbb{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}\, A.$$

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted Row A. Equivalently, it is the column span of A^T:

$$\text{Row}\, A = \text{Col}\, A^T.$$

It is a subspace of $\mathbf{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}\, A.$$

Hence we have shown:

Fact: $(\text{Row}\, A)^\perp = \text{Nul}\, A.$

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted $\mathrm{Row}\, A$. Equivalently, it is the column span of A^T:

$$\mathrm{Row}\, A = \mathrm{Col}\, A^T.$$

It is a subspace of $\mathbf{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\mathrm{Span}\{v_1, v_2, \ldots, v_m\}^{\perp} = \mathrm{Nul}\, A.$$

Hence we have shown:

Fact: $(\mathrm{Row}\, A)^{\perp} = \mathrm{Nul}\, A$.

Replacing A by A^T, and remembering $\mathrm{Row}\, A^T = \mathrm{Col}\, A$:

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted Row A. Equivalently, it is the column span of A^T:

$$\text{Row}\, A = \text{Col}\, A^T.$$

It is a subspace of $\mathbf{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}\, A.$$

Hence we have shown:

Fact: $(\text{Row}\, A)^\perp = \text{Nul}\, A$.

Replacing A by A^T, and remembering $\text{Row}\, A^T = \text{Col}\, A$:

Fact: $(\text{Col}\, A)^\perp = \text{Nul}\, A^T$.

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted $\text{Row}\, A$. Equivalently, it is the column span of A^T:

$$\text{Row}\, A = \text{Col}\, A^T.$$

It is a subspace of $\mathbb{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\text{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \text{Nul}\, A.$$

Hence we have shown:

Fact: $(\text{Row}\, A)^\perp = \text{Nul}\, A$.

Replacing A by A^T, and remembering $\text{Row}\, A^T = \text{Col}\, A$:

Fact: $(\text{Col}\, A)^\perp = \text{Nul}\, A^T$.

Using property 2 and taking the orthogonal complements of both sides, we get:

Definition

The **row space** of an $m \times n$ matrix A is the span of the *rows* of A. It is denoted $\operatorname{Row} A$. Equivalently, it is the column span of A^T:

$$\operatorname{Row} A = \operatorname{Col} A^T.$$

It is a subspace of $\mathbb{R}^n$.

We showed before that if A has rows $v_1^T, v_2^T, \ldots, v_m^T$, then

$$\operatorname{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \operatorname{Nul} A.$$

Hence we have shown:

Fact: $(\operatorname{Row} A)^\perp = \operatorname{Nul} A$.

Replacing A by A^T, and remembering $\operatorname{Row} A^T = \operatorname{Col} A$:

Fact: $(\operatorname{Col} A)^\perp = \operatorname{Nul} A^T$.

Using property 2 and taking the orthogonal complements of both sides, we get:

Fact: $(\operatorname{Nul} A)^\perp = \operatorname{Row} A$ and $\operatorname{Col} A = (\operatorname{Nul} A^T)^\perp$.

Orthogonal Complements of Most of the Subspaces We've Seen

For any vectors $v_1, v_2, \ldots, v_m$:

$$\mathrm{Span}\{v_1, v_2, \ldots, v_m\}^\perp = \mathrm{Nul}\begin{pmatrix} - v_1^T - \\ - v_2^T - \\ \vdots \\ - v_m^T - \end{pmatrix}$$

For any matrix A:

$$\mathrm{Row}\,A = \mathrm{Col}\,A^T$$

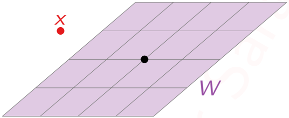
and

$$(\mathrm{Row}\,A)^\perp = \mathrm{Nul}\,A \qquad \mathrm{Row}\,A = (\mathrm{Nul}\,A)^\perp$$

$$(\mathrm{Col}\,A)^\perp = \mathrm{Nul}\,A^T \qquad \mathrm{Col}\,A = (\mathrm{Nul}\,A^T)^\perp$$

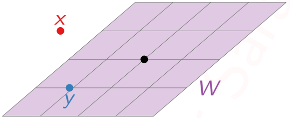
Best Approximation

Suppose you measure a data point x which you know for theoretical reasons must lie on a subspace W.



Best Approximation

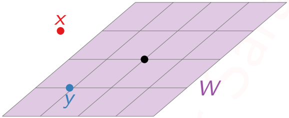
Suppose you measure a data point x which you know for theoretical reasons must lie on a subspace W.



Due to measurement error, though, the measured x is not actually in W. Best approximation: y is the *closest* point to x on W.

Best Approximation

Suppose you measure a data point x which you know for theoretical reasons must lie on a subspace W.

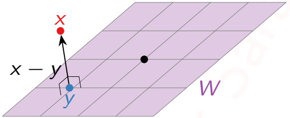


Due to measurement error, though, the measured x is not actually in W. Best approximation: y is the *closest* point to x on W.

How do you know that y is the closest point?

Best Approximation

Suppose you measure a data point x which you know for theoretical reasons must lie on a subspace W.

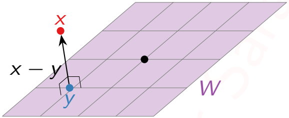


Due to measurement error, though, the measured x is not actually in W. Best approximation: y is the *closest* point to x on W.

How do you know that y is the closest point? The vector from y to x is orthogonal to W:

Best Approximation

Suppose you measure a data point x which you know for theoretical reasons must lie on a subspace W.



Due to measurement error, though, the measured x is not actually in W. Best approximation: y is the *closest* point to x on W.

How do you know that y is the closest point? The vector from y to x is orthogonal to W: it is in the *orthogonal complement* $W^{\perp}$.

Orthogonal Projection onto a Line

Theorem

Let $L = \mathrm{Span}\{u\}$ be a line in R^n, and let x be in R^n. The closest point to x on L is the point

$$\mathrm{proj}_L(x) = \frac{x \cdot u}{u \cdot u}\, u.$$

Orthogonal Projection onto a Line

Theorem

Let $L = \text{Span}\{u\}$ be a line in R^n, and let x be in R^n. The closest point to x on L is the point

$$\text{proj}_L(x) = \frac{x \cdot u}{u \cdot u}\, u.$$

This point is called the **orthogonal projection of** x **onto** L**.**

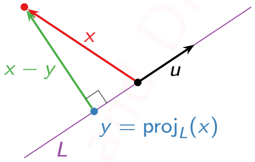
Orthogonal Projection onto a Line

Theorem

Let $L = \text{Span}\{u\}$ be a line in R^n, and let x be in R^n. The closest point to x on L is the point

$$\text{proj}_L(x) = \frac{x \cdot u}{u \cdot u}\, u.$$

This point is called the **orthogonal projection of x onto L.**



Why? Let $y = \text{proj}_L(x)$. We have to verify that $x - y$ is in $L^{\perp}$. This means proving that $u \cdot (x - y) = 0$.

$$u \cdot (x - y) = u \cdot \left(x - \frac{x \cdot u}{u \cdot u} u\right) = u \cdot x - \frac{x \cdot u}{u \cdot u}(u \cdot u) = u \cdot x - x \cdot u = 0.$$

Orthogonal Projection onto a Line

Example

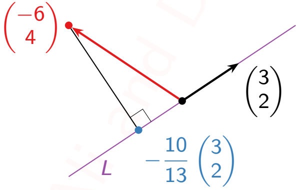
Compute the orthogonal projection of $x = \begin{pmatrix} -6 \\ 4 \end{pmatrix}$ onto the line L spanned by $u = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.

$$y = \mathrm{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u = \frac{-18+8}{9+4} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = -\frac{10}{13} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Orthogonal Projection onto a Line
Example

Compute the orthogonal projection of $x = \begin{pmatrix} -6 \\ 4 \end{pmatrix}$ onto the line L spanned by $u = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$.

$$y = \text{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u = \frac{-18+8}{9+4} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = -\frac{10}{13} \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$



Orthogonal Sets

Definition

A set of *nonzero* vectors is **orthogonal** if each pair of vectors is orthogonal.

Orthogonal Sets

Definition

A set of *nonzero* vectors is **orthogonal** if each pair of vectors is orthogonal. It is **orthonormal** if, in addition, each vector is a unit vector.

Orthogonal Sets

Definition

A set of *nonzero* vectors is **orthogonal** if each pair of vectors is orthogonal. It is **orthonormal** if, in addition, each vector is a unit vector.

Example: $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$ is an orthogonal set. Check:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = 0 \qquad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0 \qquad \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0.$$

Orthogonal Sets

Definition

A set of *nonzero* vectors is **orthogonal** if each pair of vectors is orthogonal. It is **orthonormal** if, in addition, each vector is a unit vector.

Example: $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$ is an orthogonal set. Check:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = 0 \qquad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0 \qquad \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0.$$

Lemma

An orthogonal set of vectors is linearly independent.

Suppose $\{u_1, u_2, \ldots, u_m\}$ is orthogonal. We need to show that the equation

$$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m = 0$$

has only the trivial solution $c_1 = c_2 = \cdots = c_m = 0$.

$$0 = u_1 \cdot (c_1 u_1 + c_2 u_2 + \cdots + c_m u_m) = c_1(u_1 \cdot u_1) + 0 + 0 + \cdots + 0.$$

Hence $c_1 = 0$. Similarly for the other c_i.

Orthogonal Bases

An orthogonal set $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ forms a basis for $W = \operatorname{Span} \mathcal{B}$.

Orthogonal Bases

An orthogonal set $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ forms a basis for $W = \operatorname{Span} \mathcal{B}$.

An advantage of orthogonal bases is it's *very easy* to compute the $\mathcal{B}$-coordinates of a vector in W.

Orthogonal Bases

An orthogonal set $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ forms a basis for $W = \operatorname{Span}\mathcal{B}$.

An advantage of orthogonal bases is it's *very easy* to compute the $\mathcal{B}$-coordinates of a vector in W.

Theorem

Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \operatorname{Span}\mathcal{B}$. Then

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$

Orthogonal Bases

An orthogonal set $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ forms a basis for $W = \operatorname{Span} \mathcal{B}$.

An advantage of orthogonal bases is it's *very easy* to compute the $\mathcal{B}$-coordinates of a vector in W.

Theorem

Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \operatorname{Span} \mathcal{B}$. Then

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m} u_m.$$

In other words, the $\mathcal{B}$-coordinates of x are $\left(\frac{x \cdot u_1}{u_1 \cdot u_1}, \frac{x \cdot u_2}{u_2 \cdot u_2}, \ldots, \frac{x \cdot u_m}{u_m \cdot u_m} \right)$.

Why? If $x = c_1 u_1 + c_2 u_2 + \cdots + c_m u_m$, then

$$x \cdot u_1 = c_1(u_1 \cdot u_1) + 0 + \cdots + 0 \implies c_1 = \frac{x \cdot u_1}{u_1 \cdot u_1}.$$

Similarly for the other c_i.

Orthogonal Bases

Geometric reason

Theorem

Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \operatorname{Span} \mathcal{B}$. Then

proj$_{L_2}(u_2)$

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \boxed{\frac{x \cdot u_2}{u_2 \cdot u_2} u_2} + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m} u_m.$$

If L_i is the line spanned by u_i, then this says

$$x = \operatorname{proj}_{L_1}(x) + \operatorname{proj}_{L_2}(x) + \cdots + \operatorname{proj}_{L_m}(x).$$

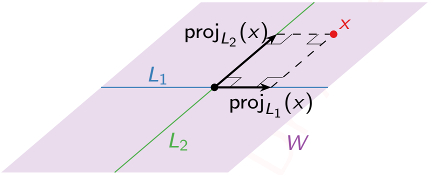
Theorem

Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \operatorname{Span}\mathcal{B}$. Then

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \overset{\mathrm{proj}_{L_2}(u_2)}{\boxed{\frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2}} + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$

If L_i is the line spanned by u_i, then this says

$$x = \mathrm{proj}_{L_1}(x) + \mathrm{proj}_{L_2}(x) + \cdots + \mathrm{proj}_{L_m}(x).$$



Orthogonal Bases

Geometric reason

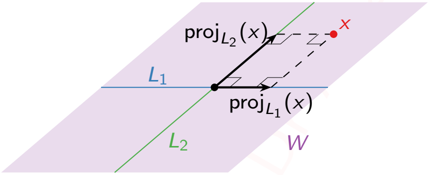
Theorem

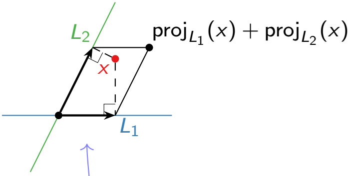
Let $\mathcal{B} = \{u_1, u_2, \ldots, u_m\}$ be an orthogonal set, and let x be a vector in $W = \operatorname{Span}\mathcal{B}$. Then

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \frac{x \cdot u_1}{u_1 \cdot u_1}\, u_1 + \overset{\mathrm{proj}_{L_2}(u_2)}{\boxed{\frac{x \cdot u_2}{u_2 \cdot u_2}\, u_2}} + \cdots + \frac{x \cdot u_m}{u_m \cdot u_m}\, u_m.$$

If L_i is the line spanned by u_i, then this says

$$x = \mathrm{proj}_{L_1}(x) + \mathrm{proj}_{L_2}(x) + \cdots + \mathrm{proj}_{L_m}(x).$$





Warning: This only works for an *orthogonal* basis.

Orthogonal Bases

Example

Problem: Find the $\mathcal{B}$-coordinates of $x = \binom{0}{3}$, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \end{pmatrix} \right\}.$$

Orthogonal Bases

Example

Problem: Find the $\mathcal{B}$-coordinates of $x = \binom{0}{3}$, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \end{pmatrix} \right\}.$$

Old way:

$$\left(\begin{array}{cc|c} 1 & -4 & 0 \\ 2 & 2 & 3 \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{cc|c} 1 & 0 & 6/5 \\ 0 & 1 & 6/20 \end{array}\right) \implies [x]_{\mathcal{B}} = \begin{pmatrix} 6/5 \\ 6/20 \end{pmatrix}.$$

Orthogonal Bases
Example

Problem: Find the $\mathcal{B}$-coordinates of $x = \binom{0}{3}$, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \end{pmatrix} \right\}.$$

Old way:

$$\left(\begin{array}{cc|c} 1 & -4 & 0 \\ 2 & 2 & 3 \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{cc|c} 1 & 0 & 6/5 \\ 0 & 1 & 6/20 \end{array}\right) \implies [x]_{\mathcal{B}} = \begin{pmatrix} 6/5 \\ 6/20 \end{pmatrix}.$$

New way: note $\mathcal{B}$ is an *orthogonal* basis.

$$x = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{3 \cdot 2}{1^2 + 2^2} u_1 + \frac{3 \cdot 2}{(-4)^2 + 2^2} u_2 = \frac{6}{5} u_1 + \frac{6}{20} u_2.$$

So the $\mathcal{B}$-coordinates are $\frac{6}{5}, \frac{6}{20}$.

Orthogonal Bases
Example

Problem: Find the $\mathcal{B}$-coordinates of $x = \binom{0}{3}$, where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \end{pmatrix} \right\}.$$

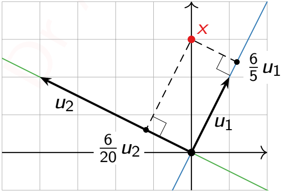
Old way:

$$\left(\begin{array}{cc|c} 1 & -4 & 0 \\ 2 & 2 & 3 \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{cc|c} 1 & 0 & 6/5 \\ 0 & 1 & 6/20 \end{array}\right) \implies [x]_{\mathcal{B}} = \begin{pmatrix} 6/5 \\ 6/20 \end{pmatrix}.$$

New way: note $\mathcal{B}$ is an *orthogonal* basis.

$$x = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{3 \cdot 2}{1^2 + 2^2} u_1 + \frac{3 \cdot 2}{(-4)^2 + 2^2} u_2 = \frac{6}{5} u_1 + \frac{6}{20} u_2.$$

So the $\mathcal{B}$-coordinates are $\frac{6}{5}, \frac{6}{20}$.



Orthogonal Bases

Example

Problem: Find the $\mathcal{B}$-coordinates of $x = (6, 1, -8)$ where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}.$$

Orthogonal Bases
Example

Problem: Find the $\mathcal{B}$-coordinates of $x = (6, 1, -8)$ where

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}.$$

Answer:

$$\begin{aligned}
[x]_{\mathcal{B}} &= \left(\frac{x \cdot u_1}{u_1 \cdot u_1}, \frac{x \cdot u_2}{u_2 \cdot u_2}, \frac{x \cdot u_3}{u_3 \cdot u_3} \right) \\
&= \left(\frac{6 \cdot 1 + 1 \cdot 1 - 8 \cdot 1}{1^2 + 1^2 + 1^2}, \frac{6 \cdot 1 + 1 \cdot (-2) - 8 \cdot 1}{1^2 + (-2)^2 + 1^2}, \frac{6 \cdot 1 + 1 \cdot 0 + (-8) \cdot (-1)}{1^2 + 0^2 + (-1)^2} \right) \\
&= \left(-\frac{1}{3}, -\frac{2}{3}, 7 \right).
\end{aligned}$$

Check:

$$\begin{pmatrix} 6 \\ 1 \\ -8 \end{pmatrix} = -\frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + 7 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

✔

Orthonormal Basis

Definition

A set of vectors in R^n is called an **orthonormal set** if it is an orthogonal set of unit vectors.

Theorem

Let $\{x_1, x_2, ..., x_k\}$ be an orthonormal basis for a subspace W of R^n and let w be any vector in W. Then the unique scalars $c_1, c_2, ..., c_k$ such that

$$w = c_1 x_1 + c_2 x_2 + ... + c_k x_k$$

are given by

$$c_i = w \cdot x_i, \quad i = 1, 2, ..., k.$$

Theorem

The matrix Q (square or rectangular) has orthonormal columns if and only if $Q^TQ = I$

Proof.

If Q has orthonormal columns then,

$$(Q^TQ)_{ij} = q_i \cdot q_j = I.$$

Conversely,
If $Q^TQ = I$, then

$$q_i \cdot q_j = \begin{cases} 1, & \text{if } i \neq j \\ 0, & \text{if} i = j. \end{cases}$$

Orthogonal Matrices

Theorem

Any square matrix Q whose columns form an orthonormal set is called **Orthogonal Matrix**.

Theorem

Let Q be an $n \times n$ matrix. Then the following statements are equivalent:

1. Q is orthogonal.
2. $Q^T = Q^{-1}$.
3. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.
4. $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

   where $\mathbf{x}$ and $\mathbf{y}$ are from R^n.

Orthogonal Matrices

Theorem

Let Q be an orthogonal matrix.

1. Q^{-1} is orthogonal.
2. $det(Q) = \pm 1$
3. If λ is an eigenvalue of Q, then $|\lambda| = 1$.
4. Product of orthogonal matrices of same size is another orthogonal matrix.
5. Rows of Q forms an orthonormal set.

Orthogonal Matrices

Examples

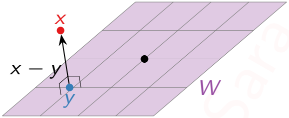
- $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & cos\theta \end{bmatrix}$.
- $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if

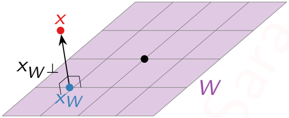
Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^{\perp}$:



Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^\perp$:



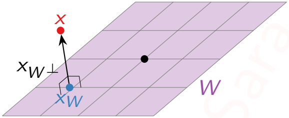
Reformulation: Every vector x can be decompsed uniquely as

$$x = x_W + x_{W^\perp}$$

where $x_W = y$ is the closest vector to x in W, and $x_{W^\perp} = x - y$ is in $W^\perp$.

Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^{\perp}$:



Reformulation: Every vector x can be decompsed uniquely as

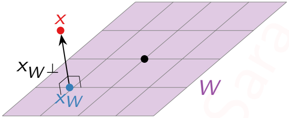
$$x = x_W + x_{W^{\perp}}$$

where $x_W = y$ is the closest vector to x in W, and $x_{W^{\perp}} = x - y$ is in $W^{\perp}$.

Example: Let $u = \binom{3}{2}$ and let $L = \mathrm{Span}\{u\}$. Let $x = \binom{-6}{4}$.

Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^\perp$:



Reformulation: Every vector x can be decompsed uniquely as

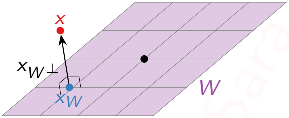
$$x = x_W + x_{W^\perp}$$

where $x_W = y$ is the closest vector to x in W, and $x_{W^\perp} = x - y$ is in $W^\perp$.

Example: Let $u = \binom{3}{2}$ and let $L = \text{Span}\{u\}$. Let $x = \binom{-6}{4}$. Then the closest point to x in L is

Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^\perp$:



Reformulation: Every vector x can be decompsed uniquely as

$$x = x_W + x_{W^\perp}$$

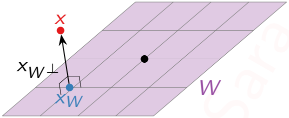
where $x_W = y$ is the closest vector to x in W, and $x_{W^\perp} = x - y$ is in $W^\perp$.

Example: Let $u = \binom{3}{2}$ and let $L = \text{Span}\{u\}$. Let $x = \binom{-6}{4}$. Then the closest point to x in L is $\text{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u$, so

$$x_L =$$

Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^\perp$:



Reformulation: Every vector x can be decompsed uniquely as

$$x = x_W + x_{W^\perp}$$

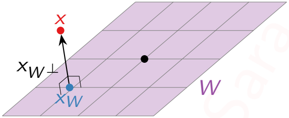
where $x_W = y$ is the closest vector to x in W, and $x_{W^\perp} = x - y$ is in $W^\perp$.

Example: Let $u = \binom{3}{2}$ and let $L = \mathrm{Span}\{u\}$. Let $x = \binom{-6}{4}$. Then the closest point to x in L is $\mathrm{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u$, so

$$x_L = \mathrm{proj}_L(x) = -\frac{10}{13}\begin{pmatrix} 3 \\ 2 \end{pmatrix} \qquad x_{L^\perp} =$$

Idea Behind Orthogonal Projections

If x is not in a subspace W, then y in W is the closest to x if $x - y$ is in $W^\perp$:



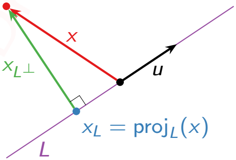
Reformulation: Every vector x can be decompsed uniquely as

$$x = x_W + x_{W^\perp}$$

where $x_W = y$ is the closest vector to x in W, and $x_{W^\perp} = x - y$ is in $W^\perp$.

Example: Let $u = \binom{3}{2}$ and let $L = \text{Span}\{u\}$. Let $x = \binom{-6}{4}$. Then the closest point to x in L is $\text{proj}_L(x) = \frac{x \cdot u}{u \cdot u} u$, so

$$x_L = \text{proj}_L(x) = -\frac{10}{13}\begin{pmatrix} 3 \\ 2 \end{pmatrix} \qquad x_{L^\perp} = x - \text{proj}_L(x) = \begin{pmatrix} -6 \\ 4 \end{pmatrix} + \frac{10}{13}\begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$



Orthogonal Projections

Definition

Let W be a subspace of R^n, and let $\{u_1, u_2, \ldots, u_m\}$ be an *orthogonal* basis for W. The **orthogonal projection** of a vector x onto W is

$$\mathrm{proj}_W(x) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i.$$

Orthogonal Projections

Definition

Let W be a subspace of R^n, and let $\{u_1, u_2, \ldots, u_m\}$ be an *orthogonal* basis for W. The **orthogonal projection** of a vector x onto W is

$$\mathrm{proj}_W(x) \stackrel{\text{def}}{=} \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i.$$

Question: What is the difference between this and the formula for $[x]_{\mathcal{B}}$ from before?

Orthogonal Projections

Definition

Let W be a subspace of R^n, and let $\{u_1, u_2, \ldots, u_m\}$ be an *orthogonal* basis for W. The **orthogonal projection** of a vector x onto W is

$$\mathrm{proj}_W(x) \stackrel{\text{def}}{=} \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i.$$

Question: What is the difference between this and the formula for $[x]_{\mathcal{B}}$ from before?

Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $\mathrm{proj}_W(x)$ is the closest point to x in W.

Orthogonal Projections

Definition

Let W be a subspace of R^n, and let $\{u_1, u_2, \ldots, u_m\}$ be an *orthogonal* basis for W. The **orthogonal projection** of a vector x onto W is

$$\mathrm{proj}_W(x) \stackrel{\mathrm{def}}{=} \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i.$$

Question: What is the difference between this and the formula for $[x]_{\mathcal{B}}$ from before?

Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $\mathrm{proj}_W(x)$ is the closest point to x in W. Therefore

$$\boxed{x_W = \mathrm{proj}_W(x) \qquad x_{W^\perp} = x - \mathrm{proj}_W(x).}$$

Why? Let $y = \mathrm{proj}_W(x)$. We need to show that $x - y$ is in $W^\perp$. In other words, $u_i \cdot (x - y) = 0$ for each i. Let's do u_1:

$$u_1 \cdot (x - y) = u_1 \cdot \left(x - \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i \right) = u_1 \cdot x - \frac{x \cdot u_1}{u_1 \cdot u_1}(u_1 \cdot u_1) - 0 - \cdots = 0.$$

Orthogonal Projections

Easy example

What is the projection of $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ onto the xy-plane?

Answer: The xy-plane is $W = \mathrm{Span}\{e_1, e_2\}$, and $\{e_1, e_2\}$ is an orthogonal basis.

$$x_W = \mathrm{proj}_W \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{x \cdot e_1}{e_1 \cdot e_1} e_1 + \frac{x \cdot e_2}{e_2 \cdot e_2} e_2 = \frac{1 \cdot 1}{1^2} e_1 + \frac{1 \cdot 2}{1^2} e_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

Orthogonal Projections

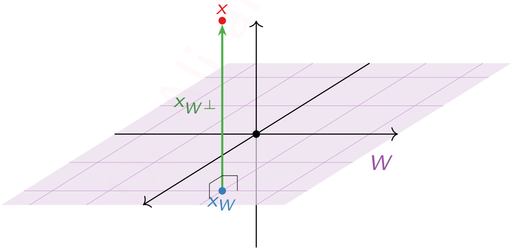
Easy example

What is the projection of $x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ onto the xy-plane?

Answer: The xy-plane is $W = \operatorname{Span}\{e_1, e_2\}$, and $\{e_1, e_2\}$ is an orthogonal basis.

$$x_W = \operatorname{proj}_W \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{x \cdot e_1}{e_1 \cdot e_1} e_1 + \frac{x \cdot e_2}{e_2 \cdot e_2} e_2 = \frac{1 \cdot 1}{1^2} e_1 + \frac{1 \cdot 2}{1^2} e_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}.$$

So this is the same projection as before.



Orthogonal Projections

More complicated example

What is the projection of $x = \begin{pmatrix} -1.1 \\ 1.4 \\ 1.45 \end{pmatrix}$ onto $W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1.1 \\ -.2 \end{pmatrix} \right\}$?

Answer: The basis is orthogonal, so

$$x_W = \text{proj}_W \begin{pmatrix} -1.1 \\ 1.4 \\ 1.45 \end{pmatrix} = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2$$

$$= \frac{(-1.1)(1)}{1^2} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{(1.4)(1.1) + (1.45)(-.2)}{1.1^2 + (-.2)^2} \begin{pmatrix} 0 \\ 1.1 \\ -.2 \end{pmatrix}$$

This turns out to be equal to $u_2 - 1.1u_1$.

Orthogonal Projections

More complicated example

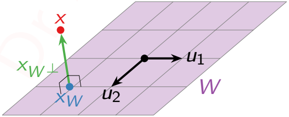
What is the projection of $x = \begin{pmatrix} -1.1 \\ 1.4 \\ 1.45 \end{pmatrix}$ onto $W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1.1 \\ -.2 \end{pmatrix} \right\}$?

Answer: The basis is orthogonal, so

$$x_W = \text{proj}_W \begin{pmatrix} -1.1 \\ 1.4 \\ 1.45 \end{pmatrix} = \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{x \cdot u_2}{u_2 \cdot u_2} u_2$$

$$= \frac{(-1.1)(1)}{1^2} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{(1.4)(1.1) + (1.45)(-.2)}{1.1^2 + (-.2)^2} \begin{pmatrix} 0 \\ 1.1 \\ -.2 \end{pmatrix}$$

This turns out to be equal to $u_2 - 1.1u_1$.



Orthogonal Projections

Picture

Let W be a subspace of R^n, and let $\{u_1, u_2, \ldots, u_m\}$ be an orthogonal basis for W. Let $L_i = \mathrm{Span}\{u_i\}$. Then

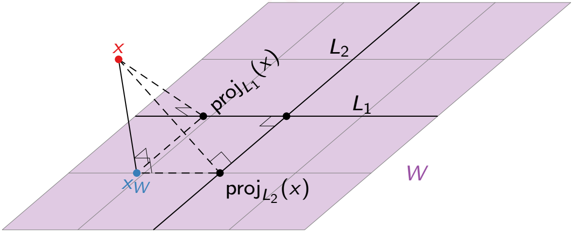
$$\mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \sum_{i=1}^{m} \mathrm{proj}_{L_i}(x).$$

Orthogonal Projections
Picture

Let W be a subspace of R^n, and let $\{u_1, u_2, \ldots, u_m\}$ be an orthogonal basis for W. Let $L_i = \mathrm{Span}\{u_i\}$. Then

$$\mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i = \sum_{i=1}^{m} \mathrm{proj}_{L_i}(x).$$

So the orthogonal projection is formed by adding orthogonal projections onto perpendicular lines.



First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W : \mathrm{R}^n \longrightarrow \mathrm{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathrm{R}^n \longrightarrow \mathrm{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of R^n.

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathrm{R}^n \longrightarrow \mathrm{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of R^n.

1. proj_W is a *linear* transformation.

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathrm{R}^n \longrightarrow \mathrm{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of R^n.

1. proj_W is a *linear* transformation.
2. For every x in W, we have $\mathrm{proj}_W(x) =$

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathrm{R}^n \longrightarrow \mathrm{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of R^n.

1. proj_W is a *linear* transformation.
2. For every x in W, we have $\mathrm{proj}_W(x) = x$.

Orthogonal Projections
Properties

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, and let x be a vector in $\mathbb{R}^n$. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathbb{R}^n \longrightarrow \mathbb{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of $\mathbb{R}^n$.

1. proj_W is a *linear* transformation.
2. For every x in W, we have $\mathrm{proj}_W(x) = x$.
3. For every x in $W^\perp$, we have $\mathrm{proj}_W(x) =$

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of $\mathbf{R}^n$, and let x be a vector in $\mathbf{R}^n$. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathbf{R}^n \longrightarrow \mathbf{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of $\mathbf{R}^n$.

1. proj_W is a *linear* transformation.
2. For every x in W, we have $\mathrm{proj}_W(x) = x$.
3. For every x in $W^\perp$, we have $\mathrm{proj}_W(x) = 0$.

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of R^n, and let x be a vector in R^n. Then $y = \mathrm{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\mathrm{dist}(x, y') \geq \mathrm{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\mathrm{proj}_W \colon \mathrm{R}^n \longrightarrow \mathrm{R}^n \qquad x \mapsto \mathrm{proj}_W(x).$$

Theorem

Let W be a subspace of R^n.

1. proj_W is a *linear* transformation.
2. For every x in W, we have $\mathrm{proj}_W(x) = x$.
3. For every x in $W^\perp$, we have $\mathrm{proj}_W(x) = 0$.
4. The range of proj_W is

First we restate the property we've been using all along.

Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, and let x be a vector in $\mathbb{R}^n$. Then $y = \text{proj}_W(x)$ is the closest point in W to x, in the sense that

$$\text{dist}(x, y') \geq \text{dist}(x, y) \quad \text{for all} \quad y' \text{ in } W.$$

We can think of orthogonal projection as a *transformation*:

$$\text{proj}_W \colon \mathbb{R}^n \longrightarrow \mathbb{R}^n \qquad x \mapsto \text{proj}_W(x).$$

Theorem

Let W be a subspace of $\mathbb{R}^n$.

1. proj_W is a *linear* transformation.
2. For every x in W, we have $\text{proj}_W(x) = x$.
3. For every x in $W^\perp$, we have $\text{proj}_W(x) = 0$.
4. The range of proj_W is W.

Poll

Let W be a subspace of $\mathbb{R}^n$.

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

Poll

Let W be a subspace of R^n.

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

The 1-eigenspace is

Poll

Let W be a subspace of R^n.

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

The 1-eigenspace is W.

Poll

Let W be a subspace of R^n.

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

The 1-eigenspace is W.

The 0-eigenspace is

Poll

Let W be a subspace of R^n.

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

The 1-eigenspace is W.

The 0-eigenspace is $W^\perp$.

Poll

Let W be a subspace of R^n.

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

The 1-eigenspace is W.

The 0-eigenspace is $W^\perp$.

We have $\dim W + \dim W^\perp = n$, so that gives n linearly independent eigenvectors already.

Poll

Let W be a subspace of R^n.

Poll

Let A be the matrix for proj_W. What is/are the eigenvalue(s) of A?

A. 0 B. 1 C. -1 D. $0, 1$ E. $1, -1$ F. $0, -1$ G. $-1, 0, 1$

The 1-eigenspace is W.

The 0-eigenspace is $W^\perp$.

We have $\dim W + \dim W^\perp = n$, so that gives n linearly independent eigenvectors already.

So the answer is D.

What is the matrix for $\mathrm{proj}_W \colon \mathrm{R}^3 \to \mathrm{R}^3$, where

$$W = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}?$$

Answer: Recall how to compute the matrix for a linear transformation:

$$A = \begin{pmatrix} | & | & | \\ \mathrm{proj}_W(e_1) & \mathrm{proj}_W(e_2) & \mathrm{proj}_W(e_3) \\ | & | & | \end{pmatrix}.$$

We compute:

$$\mathrm{proj}_W(e_1) = \frac{e_1 \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{e_1 \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5/6 \\ 1/3 \\ -1/6 \end{pmatrix}$$

$$\mathrm{proj}_W(e_2) = \frac{e_2 \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{e_2 \cdot u_2}{u_2 \cdot u_2} u_2 = 0 + \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}$$

$$\mathrm{proj}_W(e_3) = \frac{e_3 \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{e_3 \cdot u_2}{u_2 \cdot u_2} u_2 = -\frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1/6 \\ 1/3 \\ 5/6 \end{pmatrix}$$

Therefore $A = \begin{pmatrix} 5/6 & 1/3 & -1/6 \\ 1/3 & 1/3 & 1/3 \\ -1/6 & 1/3 & 5/6 \end{pmatrix}$.

Orthogonal Projections

Matrix facts

Let W be an m-dimensional subspace of $\mathbb{R}^n$, let $\text{proj}_W \colon \mathbb{R}^n \to W$ be the projection, and let A be the matrix for proj_L.

Orthogonal Projections

Matrix facts

Let W be an m-dimensional subspace of $\mathbf{R}^n$, let $\text{proj}_W \colon \mathbf{R}^n \to W$ be the projection, and let A be the matrix for proj_L.

Fact 1: A is diagonalizable with eigenvalues 0 and 1; it is similar to the diagonal matrix with m ones and $n - m$ zeros on the diagonal.

Why? Let $v_1, v_2, \ldots, v_m$ be a basis for W, and let $v_{m+1}, v_{m+2}, \ldots, v_n$ be a basis for $W^\perp$. These are (linearly independent) eigenvectors with eigenvalues 1 and 0, respectively, and they form a basis for $\mathbf{R}^n$ because there are n of them.

Example: If W is a plane in $\mathbf{R}^3$, then A is similar to projection onto the xy-plane:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Orthogonal Projections

Matrix facts

Let W be an m-dimensional subspace of $\mathbf{R}^n$, let $\text{proj}_W \colon \mathbf{R}^n \to W$ be the projection, and let A be the matrix for proj_L.

Fact 1: A is diagonalizable with eigenvalues 0 and 1; it is similar to the diagonal matrix with m ones and $n-m$ zeros on the diagonal.

Why? Let $v_1, v_2, \ldots, v_m$ be a basis for W, and let $v_{m+1}, v_{m+2}, \ldots, v_n$ be a basis for $W^\perp$. These are (linearly independent) eigenvectors with eigenvalues 1 and 0, respectively, and they form a basis for $\mathbf{R}^n$ because there are n of them.

Example: If W is a plane in $\mathbf{R}^3$, then A is similar to projection onto the xy-plane:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Fact 2: $A^2 = A$.

Why? Projecting twice is the same as projecting once:

$$\text{proj}_W \circ \text{proj}_W = \text{proj}_W \implies A \cdot A = A.$$

What is the distance from e_1 to $W = \text{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$?

Answer: The closest point on W to e_1 is $\text{proj}_W(e_1) = \begin{pmatrix} 5/6 \\ 1/3 \\ -1/6 \end{pmatrix}$.

The distance from e_1 to this point is

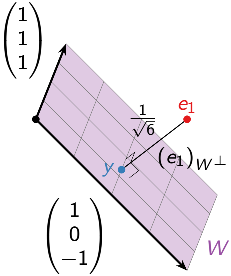
$$\begin{aligned}
\text{dist}\big(e_1, \text{proj}_W(e_1)\big) &= \|(e_1)_{W^\perp}\| \\
&= \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5/6 \\ 1/3 \\ -1/6 \end{pmatrix} \right\| \\
&= \left\| \begin{pmatrix} 1/6 \\ -1/3 \\ -1/6 \end{pmatrix} \right\| \\
&= \sqrt{(1/6)^2 + (-1/3)^2 + (-1/6)^2} \\
&= \frac{1}{\sqrt{6}}.
\end{aligned}$$

What is the distance from e_1 to $W = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$?

Answer: The closest point on W to e_1 is $\mathrm{proj}_W(e_1) = \begin{pmatrix} 5/6 \\ 1/3 \\ -1/6 \end{pmatrix}$.

The distance from e_1 to this point is

$$\begin{aligned}
\mathrm{dist}\big(e_1, \mathrm{proj}_W(e_1)\big) &= \|(e_1)_{W^\perp}\| \\
&= \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5/6 \\ 1/3 \\ -1/6 \end{pmatrix} \right\| \\
&= \left\| \begin{pmatrix} 1/6 \\ -1/3 \\ -1/6 \end{pmatrix} \right\| \\
&= \sqrt{(1/6)^2 + (-1/3)^2 + (-1/6)^2} \\
&= \frac{1}{\sqrt{6}}.
\end{aligned}$$



Motivation

All of the procedures we learned require an *orthogonal* basis $\{u_1, u_2, \ldots, u_m\}$.

Motivation

All of the procedures we learned require an *orthogonal* basis $\{u_1, u_2, \ldots, u_m\}$.

- Finding the $\mathcal{B}$-coordinates of a vector x using dot products:

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i$$

Motivation

All of the procedures we learned require an *orthogonal* basis $\{u_1, u_2, \ldots, u_m\}$.

- Finding the $\mathcal{B}$-coordinates of a vector x using dot products:

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i$$

- Finding the orthogonal projection of a vector x onto the span W of $u_1, u_2, \ldots, u_m$:

$$\text{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i.$$

Motivation

All of the procedures we learned require an *orthogonal* basis $\{u_1, u_2, \ldots, u_m\}$.

- Finding the $\mathcal{B}$-coordinates of a vector x using dot products:

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i$$

- Finding the orthogonal projection of a vector x onto the span W of $u_1, u_2, \ldots, u_m$:

$$\mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i}\, u_i.$$

Problem: What if your basis isn't orthogonal?

Motivation

All of the procedures we learned require an *orthogonal* basis $\{u_1, u_2, \ldots, u_m\}$.

- Finding the $\mathcal{B}$-coordinates of a vector x using dot products:

$$x = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i$$

- Finding the orthogonal projection of a vector x onto the span W of $u_1, u_2, \ldots, u_m$:

$$\mathrm{proj}_W(x) = \sum_{i=1}^{m} \frac{x \cdot u_i}{u_i \cdot u_i} u_i.$$

Problem: What if your basis isn't orthogonal?

Solution: The Gram–Schmidt process: take any basis and make it orthogonal.

The Gram–Schmidt Process

Let $\{v_1, v_2, \ldots, v_m\}$ be a basis for a subspace W of R^n. Define:

1. $u_1 = v_1$

The Gram–Schmidt Process

Let $\{v_1, v_2, \ldots, v_m\}$ be a basis for a subspace W of R^n. Define:

1. $u_1 = v_1$
2. $u_2 = v_2 - \mathrm{proj}_{\mathrm{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1$

The Gram–Schmidt Process

Let $\{v_1, v_2, \ldots, v_m\}$ be a basis for a subspace W of R^n. Define:

1. $u_1 = v_1$
2. $u_2 = v_2 - \operatorname{proj}_{\mathrm{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1$
3. $u_3 = v_3 - \operatorname{proj}_{\mathrm{Span}\{u_1, u_2\}}(v_3) \qquad = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

The Gram–Schmidt Process

Let $\{v_1, v_2, \ldots, v_m\}$ be a basis for a subspace W of R^n. Define:

1. $u_1 = v_1$
2. $u_2 = v_2 - \mathrm{proj}_{\mathrm{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1$
3. $u_3 = v_3 - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2\}}(v_3) \qquad = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$\vdots$

m. $u_m = v_m - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2, \ldots, u_{m-1}\}}(v_m) = v_m - \displaystyle\sum_{i=1}^{m-1} \frac{v_m \cdot u_i}{u_i \cdot u_i}\, u_i$

The Gram–Schmidt Process

Let $\{v_1, v_2, \ldots, v_m\}$ be a basis for a subspace W of R^n. Define:

1. $u_1 = v_1$
2. $u_2 = v_2 - \mathrm{proj}_{\mathrm{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1$
3. $u_3 = v_3 - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2\}}(v_3) \qquad = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$\vdots$

m. $u_m = v_m - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2, \ldots, u_{m-1}\}}(v_m) \ = v_m - \displaystyle\sum_{i=1}^{m-1} \frac{v_m \cdot u_i}{u_i \cdot u_i}\, u_i$

Then $\{u_1, u_2, \ldots, u_m\}$ is an *orthogonal* basis for the same subspace W.

The Gram–Schmidt Process

Let $\{v_1, v_2, \dots, v_m\}$ be a basis for a subspace W of R^n. Define:

1. $u_1 = v_1$
2. $u_2 = v_2 - \mathrm{proj}_{\mathrm{Span}\{u_1\}}(v_2) \qquad = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1$
3. $u_3 = v_3 - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2\}}(v_3) \qquad = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$\vdots$

m. $u_m = v_m - \mathrm{proj}_{\mathrm{Span}\{u_1, u_2, \dots, u_{m-1}\}}(v_m) \; = v_m - \displaystyle\sum_{i=1}^{m-1} \frac{v_m \cdot u_i}{u_i \cdot u_i}\, u_i$

Then $\{u_1, u_2, \dots, u_m\}$ is an *orthogonal* basis for the same subspace W.

Remark

In fact, for every i between 1 and n, the set $\{u_1, u_2, \dots, u_i\}$ is an orthogonal basis for $\mathrm{Span}\{v_1, v_2, \dots, v_i\}$.

The Gram–Schmidt Process

Two vectors

Find an orthogonal basis $\{u_1, u_2\}$ for $W = \mathrm{Span}\{v_1, v_2\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Run Gram–Schmidt:

$$\text{1. } u_1 = v_1 \qquad \text{2. } u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Why does this work?

- First we take $u_1 = v_1$.
- Now we're sad because $u_1 \cdot v_2 \neq 0$, so we can't take $u_2 = v_2$.
- Fix: let $L_1 = \mathrm{Span}\{u_1\}$, and let $u_2 = (v_2)_{L_1^\perp} = v_2 - \mathrm{proj}_{L_1}(v_2)$.
- By construction, $u_1 \cdot u_2 = 0$, because $L_1 \perp u_2$.

The Gram–Schmidt Process

Two vectors

Find an orthogonal basis $\{u_1, u_2\}$ for $W = \text{Span}\{v_1, v_2\}$, where

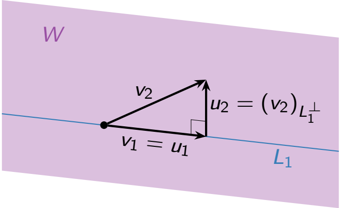
$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Run Gram–Schmidt:

$$\textcolor{red}{1.}\ u_1 = v_1 \qquad \textcolor{red}{2.}\ u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Why does this work?

- First we take $u_1 = v_1$.
- Now we're sad because $u_1 \cdot v_2 \neq 0$, so we can't take $u_2 = v_2$.
- Fix: let $L_1 = \text{Span}\{u_1\}$, and let $u_2 = (v_2)_{L_1^\perp} = v_2 - \text{proj}_{L_1}(v_2)$.
- By construction, $u_1 \cdot u_2 = 0$, because $L_1 \perp u_2$.



The Gram–Schmidt Process

Two vectors

Find an orthogonal basis $\{u_1, u_2\}$ for $W = \mathrm{Span}\{v_1, v_2\}$, where

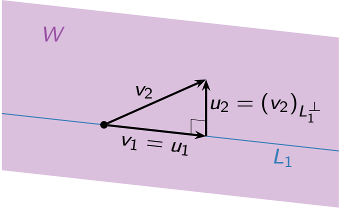
$$v_1 = \begin{pmatrix}1\\1\\0\end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix}1\\1\\1\end{pmatrix}.$$

Run Gram–Schmidt:

1. $u_1 = v_1$ 2. $u_2 = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1 = \begin{pmatrix}1\\1\\1\end{pmatrix} - \dfrac{2}{2}\begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\end{pmatrix}.$

Why does this work?

- First we take $u_1 = v_1$.
- Now we're sad because $u_1 \cdot v_2 \neq 0$, so we can't take $u_2 = v_2$.
- Fix: let $L_1 = \mathrm{Span}\{u_1\}$, and let $u_2 = (v_2)_{L_1^\perp} = v_2 - \mathrm{proj}_{L_1}(v_2)$.
- By construction, $u_1 \cdot u_2 = 0$, because $L_1 \perp u_2$.



Important: $\mathrm{Span}\{u_1, u_2\} = \mathrm{Span}\{v_1, v_2\} = W$: this is an *orthogonal* basis for the *same* subspace.

The Gram–Schmidt Process

Three vectors

Find an orthogonal basis $\{u_1, u_2, u_3\}$ for $W = \mathrm{Span}\{v_1, v_2, v_3\} = \mathrm{R}^3$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

Run Gram–Schmidt:

1. $u_1 = v_1$
2. $u_2 = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \dfrac{2}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$
3. $u_3 = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$$= \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} - \frac{4}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{1}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

The Gram–Schmidt Process

Three vectors

Find an orthogonal basis $\{u_1, u_2, u_3\}$ for $W = \text{Span}\{v_1, v_2, v_3\} = \mathbb{R}^3$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

Run Gram–Schmidt:

1. $u_1 = v_1$
2. $u_2 = v_2 - \dfrac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \dfrac{2}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$
3. $u_3 = v_3 - \dfrac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \dfrac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2$

$$= \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} - \frac{4}{2}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - \frac{1}{1}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

Important: $\text{Span}\{u_1, u_2, u_3\} = \text{Span}\{v_1, v_2, v_3\} = W$: this is an *orthogonal* basis for the *same* subspace.

The Gram–Schmidt Process

Three vectors, continued

$$v_1 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ v_2 = \begin{pmatrix}1\\1\\1\end{pmatrix},\ v_3 = \begin{pmatrix}3\\1\\1\end{pmatrix} \overset{\text{G–S}}{\rightsquigarrow} u_1 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ u_2 = \begin{pmatrix}0\\0\\1\end{pmatrix},\ u_3 = \begin{pmatrix}1\\-1\\0\end{pmatrix}$$

Why does this work?

- Once we have u_1 and u_2, then we're sad because v_3 is not orthogonal to u_1 and u_2.
- Fix: let $W_2 = \mathrm{Span}\{u_1, u_2\}$, and let $u_3 = (v_3)_{W_2^\perp} = v_3 - \mathrm{proj}_{W_3}(u_3)$.
- By construction, $u_1 \cdot u_3 = 0 = u_2 \cdot u_3$ because $W_2 \perp u_3$.

Check:

$$u_1 \cdot u_2 = 0 \ ✓$$
$$u_1 \cdot u_3 = 0 \ ✓$$
$$u_2 \cdot u_3 = 0 \ ✓$$

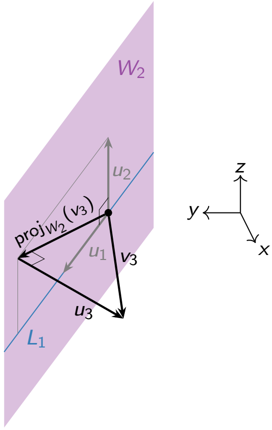
$$v_1 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ v_2 = \begin{pmatrix}1\\1\\1\end{pmatrix},\ v_3 = \begin{pmatrix}3\\1\\1\end{pmatrix} \overset{\text{G–S}}{\rightsquigarrow} u_1 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ u_2 = \begin{pmatrix}0\\0\\1\end{pmatrix},\ u_3 = \begin{pmatrix}1\\-1\\0\end{pmatrix}$$

Why does this work?

- Once we have u_1 and u_2, then we're sad because v_3 is not orthogonal to u_1 and u_2.
- Fix: let $W_2 = \mathrm{Span}\{u_1, u_2\}$, and let $u_3 = (v_3)_{W_2^\perp} = v_3 - \mathrm{proj}_{W_3}(u_3)$.
- By construction, $u_1 \cdot u_3 = 0 = u_2 \cdot u_3$ because $W_2 \perp u_3$.

Check:

$$u_1 \cdot u_2 = 0$$
$$u_1 \cdot u_3 = 0$$
$$u_2 \cdot u_3 = 0$$



The Gram–Schmidt Process

Three vectors in $\mathbb{R}^4$

Find an orthogonal basis $\{u_1, u_2, u_3\}$ for $W = \text{Span}\{v_1, v_2, v_3\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_2 = \begin{pmatrix} -1 \\ 4 \\ 4 \\ -1 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 4 \\ -2 \\ -2 \\ 0 \end{pmatrix}.$$

Run Gram–Schmidt:

1. $u_1 = v_1$
2. $u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = \begin{pmatrix} -1 \\ 4 \\ 4 \\ -1 \end{pmatrix} - \frac{6}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{pmatrix}$
3. $u_3 = v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2$

$$= \begin{pmatrix} 4 \\ -2 \\ -2 \\ 0 \end{pmatrix} - \frac{0}{24} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{-20}{25} \begin{pmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \\ -2 \end{pmatrix}$$

Poll

Poll

What happens if you try to run Gram–Schmidt on a linearly *de*pendent set of vectors $\{v_1, v_2, \ldots, v_m\}$?

A. You get an inconsistent equation.

B. For some i you get $u_i = u_{i-1}$.

C. For some i you get $u_i = 0$.

D. You create a rift in the space-time continuum.

Poll

Poll

What happens if you try to run Gram–Schmidt on a linearly *dependent* set of vectors $\{v_1, v_2, \dots, v_m\}$?

A. You get an inconsistent equation.

B. For some i you get $u_i = u_{i-1}$.

C. For some i you get $u_i = 0$.

D. You create a rift in the space-time continuum.

If $\{v_1, v_2, \dots, v_m\}$ is linearly dependent, then some v_i is in $\mathrm{Span}\{v_1, v_2, \dots, v_{i-1}\} = \mathrm{Span}\{u_1, u_2, \dots, u_{i-1}\}$.

Poll

Poll

What happens if you try to run Gram–Schmidt on a linearly *dependent* set of vectors $\{v_1, v_2, \ldots, v_m\}$?

A. You get an inconsistent equation.

B. For some i you get $u_i = u_{i-1}$.

C. For some i you get $u_i = 0$.

D. You create a rift in the space-time continuum.

If $\{v_1, v_2, \ldots, v_m\}$ is linearly dependent, then some v_i is in $\operatorname{Span}\{v_1, v_2, \ldots, v_{i-1}\} = \operatorname{Span}\{u_1, u_2, \ldots, u_{i-1}\}$.

This means

$$\begin{aligned} v_i &= \operatorname{proj}_{\operatorname{Span}\{u_1, u_2, \ldots, u_{i-1}\}}(v_i) \\ \implies u_i &= v_i - \operatorname{proj}_{\operatorname{Span}\{u_1, u_2, \ldots, u_{i-1}\}}(v_i) = 0. \end{aligned}$$

Poll

Poll

What happens if you try to run Gram–Schmidt on a linearly *dependent* set of vectors $\{v_1, v_2, \ldots, v_m\}$?

A. You get an inconsistent equation.

B. For some i you get $u_i = u_{i-1}$.

C. For some i you get $u_i = 0$.

D. You create a rift in the space-time continuum.

If $\{v_1, v_2, \ldots, v_m\}$ is linearly dependent, then some v_i is in $\mathrm{Span}\{v_1, v_2, \ldots, v_{i-1}\} = \mathrm{Span}\{u_1, u_2, \ldots, u_{i-1}\}$.

This means

$$v_i = \mathrm{proj}_{\mathrm{Span}\{u_1,u_2,\ldots,u_{i-1}\}}(v_i)$$

$$\implies u_i = v_i - \mathrm{proj}_{\mathrm{Span}\{u_1,u_2,\ldots,u_{i-1}\}}(v_i) = 0.$$

In this case, you can simply discard u_i and v_i and continue: so Gram–Schmidt produces an orthogonal basis from any spanning set!

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **orthonormal** if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **orthonormal** if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

Check: A matrix Q has orthonormal columns if and only if $Q^T Q = I$.

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **orthonormal** if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

Check: A matrix Q has orthonormal columns if and only if $Q^T Q = I$.

The columns of A are a basis for $W = \operatorname{Col} A$.

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **orthonormal** if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

Check: A matrix Q has orthonormal columns if and only if $Q^T Q = I$.

The columns of A are a basis for $W = \operatorname{Col} A$. The columns of Q come from Gram–Schmidt as applied to the columns of A, after normalizing to unit vectors.

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **orthonormal** if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

Check: A matrix Q has orthonormal columns if and only if $Q^T Q = I$.

The columns of A are a basis for $W = \operatorname{Col} A$. The columns of Q come from Gram–Schmidt as applied to the columns of A, after normalizing to unit vectors. The columns of R come from the steps in Gram–Schmidt.

QR Factorization

QR Factorization Theorem

Let A be a matrix with linearly independent columns. Then

$$A = QR$$

where Q has orthonormal columns and R is upper-triangular with positive diagonal entries.

Recall: A set of vectors $\{v_1, v_2, \ldots, v_m\}$ is **orthonormal** if they are orthogonal unit vectors: $v_i \cdot v_j = 0$ when $i \neq j$, and $v_i \cdot v_i = 1$.

Check: A matrix Q has orthonormal columns if and only if $Q^T Q = I$.

The columns of A are a basis for $W = \operatorname{Col} A$. The columns of Q come from Gram–Schmidt as applied to the columns of A, after normalizing to unit vectors. The columns of R come from the steps in Gram–Schmidt.

Here is the procedure for producing a QR factorization.

QR Factorization

Example

Find the QR factorization of $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

QR Factorization
Example

Find the QR factorization of $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

(The columns of A are the vectors v_1, v_2, v_3 from a previous example.)

QR Factorization

Example

Find the QR factorization of $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

(The columns of A are the vectors v_1, v_2, v_3 from a previous example.)

Step 1: Run Gram–Schmidt

$$u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = v_2 - 1\, u_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\begin{aligned} u_3 &= v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2 \\ &= v_3 - 2\, u_1 - 1\, u_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \end{aligned}$$

QR Factorization

Example

Find the QR factorization of $A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.

(The columns of A are the vectors v_1, v_2, v_3 from a previous example.)

Step 1: Run Gram–Schmidt and solve for v_1, v_2, v_3 in terms of u_1, u_2, u_3.

$$u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \qquad v_1 = u_1$$

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1}\, u_1 = v_2 - 1\, u_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad v_2 = u_1 + u_2$$

$$\begin{aligned} u_3 &= v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1}\, u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2}\, u_2 \\ &= v_3 - 2\, u_1 - 1\, u_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \end{aligned} \qquad v_3 = 2u_1 + u_2 + u_3$$

QR Factorization

Example, continued

$$v_1 = 1\,u_1 \qquad v_2 = 1\,u_1 + 1\,u_2 \qquad v_3 = 2\,u_1 + 1\,u_2 + 1\,u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

$$v_1 = 1\,u_1 \qquad v_2 = 1\,u_1 + 1\,u_2 \qquad v_3 = 2\,u_1 + 1\,u_2 + 1\,u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

Do this by putting the above equations in matrix form:

$$A \longrightarrow \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = \underbrace{\begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix}}_{\widehat{Q}} \underbrace{\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\widehat{R}}$$

$$v_1 = 1\, u_1 \qquad v_2 = 1\, u_1 + 1\, u_2 \qquad v_3 = 2\, u_1 + 1\, u_2 + 1\, u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

Do this by putting the above equations in matrix form:

$$A \longrightarrow \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = \underbrace{\begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix}}_{\widehat{Q}} \underbrace{\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\widehat{R}}$$

$$\text{first column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1 u_1 = v_1$$

QR Factorization

Example, continued

$$v_1 = 1\,u_1 \qquad v_2 = 1\,u_1 + 1\,u_2 \qquad v_3 = 2\,u_1 + 1\,u_2 + 1\,u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

Do this by putting the above equations in matrix form:

$$\underbrace{\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix}}_{\widehat{Q}} \underbrace{\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\widehat{R}}$$

$$\text{first column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1u_1 = v_1$$

$$\text{second column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 1u_1 + 1u_2 = v_2$$

QR Factorization
Example, continued

$$v_1 = 1u_1 \qquad v_2 = 1u_1 + 1u_2 \qquad v_3 = 2u_1 + 1u_2 + 1u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

Do this by putting the above equations in matrix form:

$$\underbrace{\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix}}_{\widehat{Q}} \underbrace{\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}}_{\widehat{R}}$$

$$\text{first column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1u_1 = v_1$$

$$\text{second column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 1u_1 + 1u_2 = v_2$$

$$\text{third column of } A = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = 2u_1 + 1u_2 + 1u_3 = v_3$$

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \quad 0 & \quad 1 & \quad \\ 1 & \quad 0 & \quad -1 & \quad \\ 0 & \quad 1 & \quad 0 & \quad \end{pmatrix} \begin{pmatrix} 1 & \quad 1 & \quad 2 & \quad \\ 0 & \quad 1 & \quad 1 & \quad \\ 0 & \quad 0 & \quad 1 & \quad \end{pmatrix}.$$

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0 & 1 \\ 1/\sqrt{2} & 0 & -1 \\ 0/\sqrt{2} & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \cdot \sqrt{2} & 1 \cdot \sqrt{2} & 2 \cdot \sqrt{2} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0/1 & 1 \\ 1/\sqrt{2} & 0/1 & -1 \\ 0/\sqrt{2} & 1/1 & 0 \end{pmatrix} \begin{pmatrix} 1 \cdot \sqrt{2} & 1 \cdot \sqrt{2} & 2 \cdot \sqrt{2} \\ 0 \cdot 1 & 1 \cdot 1 & 1 \cdot 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0/1 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0/1 & -1/\sqrt{2} \\ 0/\sqrt{2} & 1/1 & 0/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1\cdot\sqrt{2} & 1\cdot\sqrt{2} & 2\cdot\sqrt{2} \\ 0\cdot 1 & 1\cdot 1 & 1\cdot 1 \\ 0\cdot\sqrt{2} & 0\cdot\sqrt{2} & 1\cdot\sqrt{2} \end{pmatrix}.$$

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0/1 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0/1 & -1/\sqrt{2} \\ 0/\sqrt{2} & 1/1 & 0/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 \cdot \sqrt{2} & 1 \cdot \sqrt{2} & 2 \cdot \sqrt{2} \\ 0 \cdot 1 & 1 \cdot 1 & 1 \cdot 1 \\ 0 \cdot \sqrt{2} & 0 \cdot \sqrt{2} & 1 \cdot \sqrt{2} \end{pmatrix}.$$

Note that the entries in the ith column of Q multiply by the entries in the ith row of R, so this doesn't change the product.

QR Factorization

Example, continued

$$A = \widehat{Q}\widehat{R} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Scale the columns of $\widehat{Q}$ to get unit vectors, and scale the rows of $\widehat{R}$ by the opposite factor, to get Q and R.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0/1 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0/1 & -1/\sqrt{2} \\ 0/\sqrt{2} & 1/1 & 0/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 \cdot \sqrt{2} & 1 \cdot \sqrt{2} & 2 \cdot \sqrt{2} \\ 0 \cdot 1 & 1 \cdot 1 & 1 \cdot 1 \\ 0 \cdot \sqrt{2} & 0 \cdot \sqrt{2} & 1 \cdot \sqrt{2} \end{pmatrix}.$$

Note that the entries in the ith column of Q multiply by the entries in the ith row of R, so this doesn't change the product.

The final QR decomposition is:

$$A = QR \qquad Q = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix} \qquad R = \begin{pmatrix} \sqrt{2} & \sqrt{2} & 2\sqrt{2} \\ 0 & 1 & 1 \\ 0 & 0 & \sqrt{2} \end{pmatrix}$$

QR Factorization
Another example

Find the QR factorization of $A = \begin{pmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & -2 \\ 1 & -1 & 0 \end{pmatrix}$.

QR Factorization

Another example

Find the QR factorization of $A = \begin{pmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & -2 \\ 1 & -1 & 0 \end{pmatrix}$.

(The columns are vectors from a previous example.)

Step 1: Run Gram–Schmidt and solve for v_1, v_2, v_3 in terms of u_1, u_2, u_3:

$$u_1 = v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad v_1 = u_1$$

$$u_2 = v_2 - \frac{v_2 \cdot u_1}{u_1 \cdot u_1} u_1 = v_2 - \frac{3}{2} u_1 = \begin{pmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{pmatrix} \qquad v_2 = \frac{3}{2} u_1 + u_2$$

$$u_3 = v_3 - \frac{v_3 \cdot u_1}{u_1 \cdot u_1} u_1 - \frac{v_3 \cdot u_2}{u_2 \cdot u_2} u_2 = v_3 + \frac{4}{5} u_2 = \begin{pmatrix} 2 \\ 0 \\ 0 \\ -2 \end{pmatrix} \qquad v_3 = -\frac{4}{5} u_2 + u_3$$

QR Factorization

Another example, continued

$$v_1 = 1\, u_1 \qquad v_2 = \frac{3}{2}\, u_1 + 1\, u_2 \qquad v_3 = 0\, u_1 - \frac{4}{5}\, u_2 + 1\, u_3$$

Step 2: Write $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ has *orthogonal* columns u_1, u_2, u_3 and $\widehat{R}$ is upper-triangular with 1s on the diagonal.

$$\widehat{Q} = \begin{pmatrix} | & | & | \\ u_1 & u_2 & u_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix}$$

$$\widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

QR Factorization

Another example, continued

$$A = \widehat{Q}\widehat{R} \qquad \widehat{Q} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix} \qquad \widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Normalize the columns of $\widehat{Q}$ and the rows of $\widehat{R}$ to get Q and R:

$$Q = \begin{pmatrix} | & | & | \\ u_1/\|u_1\| & u_2/\|u_2\| & u_3/\|u_3\| \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} \end{pmatrix}$$

$$R = \begin{pmatrix} 1 \cdot \|u_1\| & 3/2 \cdot \|u_1\| & 0 \cdot \|u_1\| \\ 0 & 1 \cdot \|u_2\| & -4/5 \cdot \|u_2\| \\ 0 & 0 & 1 \cdot \|u_3\| \end{pmatrix} = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 5 & -4 \\ 0 & 0 & 2\sqrt{2} \end{pmatrix}$$

QR Factorization

Another example, continued

$$A = \widehat{Q}\widehat{R} \qquad \widehat{Q} = \begin{pmatrix} 1 & -5/2 & 2 \\ 1 & 5/2 & 0 \\ 1 & 5/2 & 0 \\ 1 & -5/2 & -2 \end{pmatrix} \qquad \widehat{R} = \begin{pmatrix} 1 & 3/2 & 0 \\ 0 & 1 & -4/5 \\ 0 & 0 & 1 \end{pmatrix}$$

Step 3: Normalize the columns of $\widehat{Q}$ and the rows of $\widehat{R}$ to get Q and R:

$$Q = \begin{pmatrix} | & | & | \\ u_1/\|u_1\| & u_2/\|u_2\| & u_3/\|u_3\| \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} \end{pmatrix}$$

$$R = \begin{pmatrix} 1 \cdot \|u_1\| & 3/2 \cdot \|u_1\| & 0 \cdot \|u_1\| \\ 0 & 1 \cdot \|u_2\| & -4/5 \cdot \|u_2\| \\ 0 & 0 & 1 \cdot \|u_3\| \end{pmatrix} = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 5 & -4 \\ 0 & 0 & 2\sqrt{2} \end{pmatrix}$$

The final QR decomposition is

$$A = QR \qquad Q = \begin{pmatrix} 1/2 & -1/2 & 1/\sqrt{2} \\ 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1/2 & -1/2 & -1/\sqrt{2} \end{pmatrix} \qquad R = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 5 & -4 \\ 0 & 0 & 2\sqrt{2} \end{pmatrix}.$$

QR Factorization

Application: computing determinants

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

QR Factorization

Application: computing determinants

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

Recall: Since Q has orthonormal columns, $Q^T Q = I_n$, so $Q^T = Q^{-1}$.

QR Factorization

Application: computing determinants

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

Recall: Since Q has orthonormal columns, $Q^T Q = I_n$, so $Q^T = Q^{-1}$.

But $\det(Q^T) = \det(Q)$, so

$$1 = \det(I_n) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2.$$

QR Factorization

Application: computing determinants

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

Recall: Since Q has orthonormal columns, $Q^T Q = I_n$, so $Q^T = Q^{-1}$.
But $\det(Q^T) = \det(Q)$, so

$$1 = \det(I_n) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2.$$

It follows that $\det(Q) = \pm 1$.

QR Factorization

Application: computing determinants

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

Recall: Since Q has orthonormal columns, $Q^T Q = I_n$, so $Q^T = Q^{-1}$.

But $\det(Q^T) = \det(Q)$, so

$$1 = \det(I_n) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2.$$

It follows that $\det(Q) = \pm 1$.

(Since $\det(R) > 0$, in fact $\det(Q)$ has the same sign as $\det(A)$.)

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

Recall: Since Q has orthonormal columns, $Q^T Q = I_n$, so $Q^T = Q^{-1}$.
But $\det(Q^T) = \det(Q)$, so

$$1 = \det(I_n) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2.$$

It follows that $\det(Q) = \pm 1$.

(Since $\det(R) > 0$, in fact $\det(Q)$ has the same sign as $\det(A)$.)

Therefore,

$$\det(A) = \det(Q)\det(R) = \pm\det(R).$$

QR Factorization

Application: computing determinants

Let A be an *invertible* $n \times n$ matrix. Consider its QR factorization

$$A = QR.$$

Recall: Since Q has orthonormal columns, $Q^T Q = I_n$, so $Q^T = Q^{-1}$.

But $\det(Q^T) = \det(Q)$, so

$$1 = \det(I_n) = \det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)^2.$$

It follows that $\det(Q) = \pm 1$.

(Since $\det(R) > 0$, in fact $\det(Q)$ has the same sign as $\det(A)$.)

Therefore,

$$\det(A) = \det(Q)\det(R) = \pm\det(R).$$

But R is upper-triangular, so it's easy to compute its determinant!

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

Problem

Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

Problem

Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

To say $Ax = b$ does not have a solution means that b is not in $\operatorname{Col} A$.

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

Problem

Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

To say $Ax = b$ does not have a solution means that b is not in $\operatorname{Col} A$.

The closest possible $\widehat{b}$ for which $Ax = \widehat{b}$ does have a solution is

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

Problem

Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

To say $Ax = b$ does not have a solution means that b is not in $\operatorname{Col} A$.

The closest possible $\widehat{b}$ for which $Ax = \widehat{b}$ does have a solution is $\widehat{b} = \operatorname{proj}_{\operatorname{Col} A}(b)$.

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

Problem

Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

To say $Ax = b$ does not have a solution means that b is not in $\operatorname{Col} A$.

The closest possible $\widehat{b}$ for which $Ax = \widehat{b}$ does have a solution is $\widehat{b} = \operatorname{proj}_{\operatorname{Col} A}(b)$.

Then $A\widehat{x} = \widehat{b}$ is a consistent equation.

Motivation

We now are in a position to solve the motivating problem of this third part of the course:

> **Problem**
>
> Suppose that $Ax = b$ does not have a solution. What is the best possible approximate solution?

To say $Ax = b$ does not have a solution means that b is not in $\operatorname{Col} A$.

The closest possible $\widehat{b}$ for which $Ax = \widehat{b}$ does have a solution is $\widehat{b} = \operatorname{proj}_{\operatorname{Col} A}(b)$.

Then $A\widehat{x} = \widehat{b}$ is a consistent equation.

A solution $\widehat{x}$ to $A\widehat{x} = \widehat{b}$ is a **least squares solution.**

Least Squares Solutions

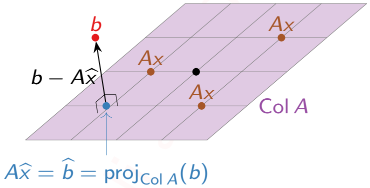
Let A be an $m \times n$ matrix.

Definition

A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbb{R}^n$ such that

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

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



Least Squares Solutions

Let A be an $m \times n$ matrix.

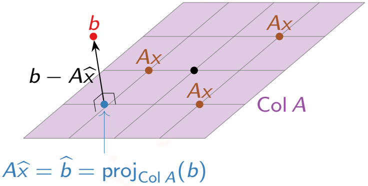
Definition

A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

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

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

Note that $b - A\widehat{x}$ is in $(\mathrm{Col}\, A)^{\perp}$.



Least Squares Solutions

Let A be an $m \times n$ matrix.

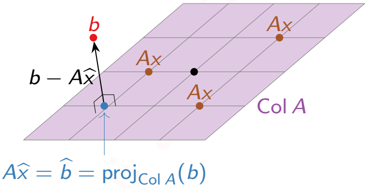
Definition

A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$\|b - A\widehat{x}\| \le \|b - Ax\|$$

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

Note that $b - A\widehat{x}$ is in $(\mathrm{Col}\,A)^{\perp}$.



In other words, a least squares solution $\widehat{x}$ solves $Ax = b$ *as closely as possible*.

Least Squares Solutions

Let A be an $m \times n$ matrix.

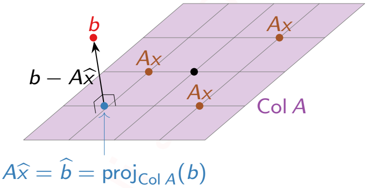
Definition

A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in R^n such that

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

for all x in R^n.

Note that $b - A\widehat{x}$ is in $(\mathrm{Col}\,A)^{\perp}$.



In other words, a least squares solution $\widehat{x}$ solves $Ax = b$ *as closely as possible*.

Equivalently, a least squares solution to $Ax = b$ is a vector $\widehat{x}$ in R^n such that

$$A\widehat{x} = \widehat{b} = \mathrm{proj}_{\mathrm{Col}\,A}(b).$$

Least Squares Solutions

Let A be an $m \times n$ matrix.

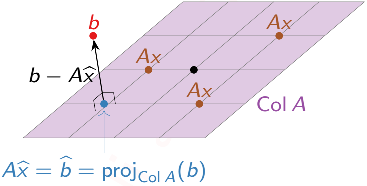
Definition

A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

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

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

Note that $b - A\widehat{x}$ is in $(\mathrm{Col}\, A)^{\perp}$.



In other words, a least squares solution $\widehat{x}$ solves $Ax = b$ *as closely as possible*.

Equivalently, a least squares solution to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$A\widehat{x} = \widehat{b} = \mathrm{proj}_{\mathrm{Col}\, A}(b).$$

This is because $\widehat{b}$ is the closest vector to b such that $A\widehat{x} = \widehat{b}$ is consistent.

Theorem
The least squares solutions to $Ax = b$ are the solutions to

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

Theorem

The least squares solutions to $Ax = b$ are the solutions to

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

This is just another $Ax = b$ problem, but with a *square* matrix $A^T A$!

Theorem

The least squares solutions to $Ax = b$ are the solutions to

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

This is just another $Ax = b$ problem, but with a *square* matrix $A^T A$!
Note we compute $\widehat{x}$ directly, without computing $\widehat{b}$ first.

Theorem

The least squares solutions to $Ax = b$ are the solutions to

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

This is just another $Ax = b$ problem, but with a *square* matrix $A^T A$!
Note we compute $\widehat{x}$ directly, without computing $\widehat{b}$ first.

Why is this true?

- We want to find $\widehat{x}$ such that $A\widehat{x} = \text{proj}_{\text{Col}\,A}(b)$.
- This means $b - A\widehat{x}$ is in $(\text{Col}\,A)^{\perp}$.
- Recall that $(\text{Col}\,A)^{\perp} = \text{Nul}(A^T)$.
- So $b - A\widehat{x}$ is in $(\text{Col}\,A)^{\perp}$ if and only if $A^T(b - A\widehat{x}) = 0$.
- In other words, $A^T A\widehat{x} = A^T b$.

Theorem

The least squares solutions to $Ax = b$ are the solutions to

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

This is just another $Ax = b$ problem, but with a *square* matrix $A^T A$!
Note we compute $\widehat{x}$ directly, without computing $\widehat{b}$ first.

Why is this true?

- We want to find $\widehat{x}$ such that $A\widehat{x} = \mathrm{proj}_{\mathrm{Col}\,A}(b)$.
- This means $b - A\widehat{x}$ is in $(\mathrm{Col}\,A)^\perp$.
- Recall that $(\mathrm{Col}\,A)^\perp = \mathrm{Nul}(A^T)$.
- So $b - A\widehat{x}$ is in $(\mathrm{Col}\,A)^\perp$ if and only if $A^T(b - A\widehat{x}) = 0$.
- In other words, $A^T A\widehat{x} = A^T b$.

Alternative when A has orthogonal columns $v_1, v_2, \ldots, v_n$:

$$\widehat{b} = \mathrm{proj}_{\mathrm{Col}\,A}(b) = \sum_{i=1}^{n} \frac{b \cdot v_i}{v_i \cdot v_i}\, v_i$$

The right hand side equals $A\widehat{x}$, where $\widehat{x} = \left(\frac{b \cdot v_1}{v_1 \cdot v_1}, \frac{b \cdot v_2}{v_2 \cdot v_2}, \cdots, \frac{b \cdot v_n}{v_n \cdot v_n}\right)$.

Least Squares Solutions

Example

Find the least squares solutions to $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We have

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

Row reduce:

$$\left(\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -3 \end{array}\right).$$

Least Squares Solutions

Example

Find the least squares solutions to $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We have

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

Row reduce:

$$\left(\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -3 \end{array}\right).$$

So the only least squares solution is $\widehat{x} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$.

Least Squares Solutions

Example, continued

How close did we get?

$$\widehat{b} = A\widehat{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

The distance from b is

$$\|b - A\widehat{x}\| = \left\| \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\| = \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6}.$$

Least Squares Solutions

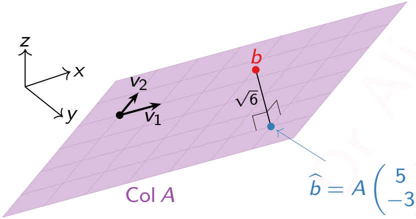
Example, continued

How close did we get?

$$\widehat{b} = A\widehat{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

The distance from b is

$$\|b - A\widehat{x}\| = \left\| \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\| = \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6}.$$



$\widehat{b} = A\begin{pmatrix} 5 \\ -3 \end{pmatrix}$

Let

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

be the columns of A, and let $\mathcal{B} = \{v_1, v_2\}$.

Least Squares Solutions

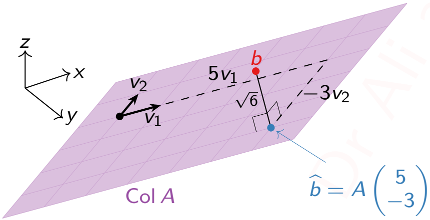
Example, continued

How close did we get?

$$\widehat{b} = A\widehat{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

The distance from b is

$$\|b - A\widehat{x}\| = \left\| \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\| = \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6}.$$



Let

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

be the columns of A, and let $\mathcal{B} = \{v_1, v_2\}$.

Note $\widehat{x} = \binom{5}{-3}$ is just the $\mathcal{B}$-coordinates of $\widehat{b}$, in Col $A = \text{Span}\{v_1, v_2\}$.

Least Squares Solutions

Second example

Find the least squares solutions to $Ax = b$ where:

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

We have

$$A^T A = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}.$$

Row reduce:

$$\left(\begin{array}{cc|c} 5 & -1 & 2 \\ -1 & 5 & -2 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & -1/3 \end{array}\right).$$

Find the least squares solutions to $Ax = b$ where:

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

We have

$$A^T A = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}.$$

Row reduce:

$$\left(\begin{array}{cc|c} 5 & -1 & 2 \\ -1 & 5 & -2 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & -1/3 \end{array}\right).$$

So the only least squares solution is $\widehat{x} = \begin{pmatrix} 1/3 \\ -1/3 \end{pmatrix}$.

Least Squares Solutions

Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Least Squares Solutions

Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

Least Squares Solutions

Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in R^n.

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in R^n.
2. The columns of A are linearly independent.

Least Squares Solutions

Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in R^n.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

Least Squares Solutions

Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in R^n.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

In this case, the least squares solution is $(A^T A)^{-1}(A^T b)$.

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in R^n.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

In this case, the least squares solution is $(A^T A)^{-1}(A^T b)$.

Why?

Least Squares Solutions
Uniqueness

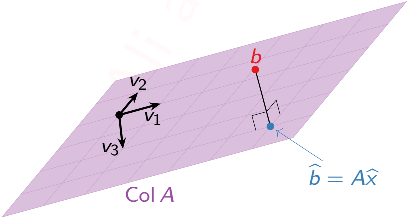
When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

Theorem
Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in $\mathbf{R}^n$.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

In this case, the least squares solution is $(A^T A)^{-1}(A^T b)$.

Why? If the columns of A are linearly *de*pendent, then $A\widehat{x} = \widehat{b}$ has many solutions:



Least Squares Solutions

Uniqueness

When does $Ax = b$ have a *unique* least squares solution $\widehat{x}$?

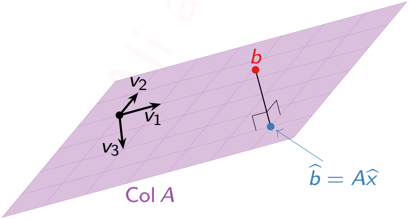
Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in R^n.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

In this case, the least squares solution is $(A^T A)^{-1}(A^T b)$.

Why? If the columns of A are linearly *de*pendent, then $A\widehat{x} = \widehat{b}$ has many solutions:



Note: $A^T A$ is always a square matrix, but it need not be invertible.

Application

Data modeling: best fit line

Find the best fit line through $(0, 6)$, $(1, 0)$, and $(2, 0)$.

The general equation of a line is

$$y = C + Dx.$$

So we want to solve:

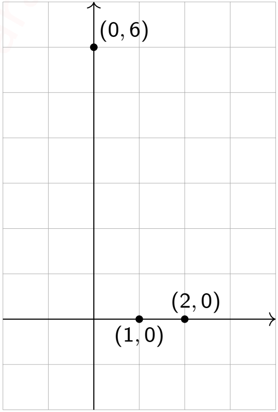
$$\begin{aligned} 6 &= C + D \cdot 0 \\ 0 &= C + D \cdot 1 \\ 0 &= C + D \cdot 2. \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We already saw: the least squares solution is $\binom{5}{-3}$. So the best fit line is

$$y = -3x + 5.$$



Application

Data modeling: best fit line

Find the best fit line through $(0,6)$, $(1,0)$, and $(2,0)$.

The general equation of a line is

$$y = C + Dx.$$

So we want to solve:

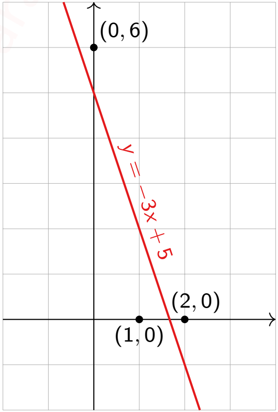
$$\begin{aligned} 6 &= C + D \cdot 0 \\ 0 &= C + D \cdot 1 \\ 0 &= C + D \cdot 2. \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We already saw: the least squares solution is $\binom{5}{-3}$. So the best fit line is

$$y = -3x + 5.$$



Application

Data modeling: best fit line

Find the best fit line through $(0,6)$, $(1,0)$, and $(2,0)$.

The general equation of a line is

$$y = C + Dx.$$

So we want to solve:

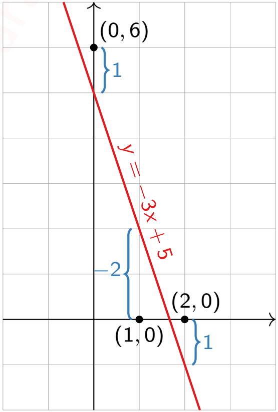
$$\begin{aligned} 6 &= C + D \cdot 0 \\ 0 &= C + D \cdot 1 \\ 0 &= C + D \cdot 2. \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We already saw: the least squares solution is $\binom{5}{-3}$. So the best fit line is

$$y = -3x + 5.$$



$$A\begin{pmatrix} 5 \\ -3 \end{pmatrix} - \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

Poll

Poll

What does the best fit line minimize?

A. The sum of the squares of the distances from the data points to the line.

B. The sum of the squares of the vertical distances from the data points to the line.

C. The sum of the squares of the horizontal distances from the data points to the line.

D. The maximal distance from the data points to the line.

Poll

Poll

What does the best fit line minimize?

A. The sum of the squares of the distances from the data points to the line.

B. The sum of the squares of the vertical distances from the data points to the line.

C. The sum of the squares of the horizontal distances from the data points to the line.

D. The maximal distance from the data points to the line.

Answer: B. See the picture on the previous slide.

Application

Best fit ellipse

Find the best fit ellipse for the points $(0, 2)$, $(2, 1)$, $(1, -1)$, $(-1, -2)$, $(-3, 1)$.

Application

Best fit ellipse

Find the best fit ellipse for the points $(0, 2)$, $(2, 1)$, $(1, -1)$, $(-1, -2)$, $(-3, 1)$.

The general equation for an ellipse is

$$x^2 + Ay^2 + Bxy + Cx + Dy + E = 0$$

Application

Best fit ellipse

Find the best fit ellipse for the points $(0, 2)$, $(2, 1)$, $(1, -1)$, $(-1, -2)$, $(-3, 1)$.

The general equation for an ellipse is

$$x^2 + Ay^2 + Bxy + Cx + Dy + E = 0$$

So we want to solve:

$$\begin{aligned}
(0)^2 + \quad A(2)^2 + \quad B(0)(2) + \quad C(0) + \quad D(2) + E &= 0 \\
(2)^2 + \quad A(1)^2 + \quad B(2)(1) + \quad C(2) + \quad D(1) + E &= 0 \\
(1)^2 + A(-1)^2 + \quad B(1)(-1) + \quad C(1) + D(-1) + E &= 0 \\
(-1)^2 + A(-2)^2 + B(-1)(-2) + C(-1) + D(-2) + E &= 0 \\
(-3)^2 + \quad A(1)^2 + \quad B(-3)(1) + C(-3) + \quad D(1) + E &= 0
\end{aligned}$$

Application

Best fit ellipse

Find the best fit ellipse for the points $(0, 2)$, $(2, 1)$, $(1, -1)$, $(-1, -2)$, $(-3, 1)$.

The general equation for an ellipse is

$$x^2 + Ay^2 + Bxy + Cx + Dy + E = 0$$

So we want to solve:

$$\begin{aligned}
(0)^2 + A(2)^2 + B(0)(2) + C(0) + D(2) + E &= 0 \\
(2)^2 + A(1)^2 + B(2)(1) + C(2) + D(1) + E &= 0 \\
(1)^2 + A(-1)^2 + B(1)(-1) + C(1) + D(-1) + E &= 0 \\
(-1)^2 + A(-2)^2 + B(-1)(-2) + C(-1) + D(-2) + E &= 0 \\
(-3)^2 + A(1)^2 + B(-3)(1) + C(-3) + D(1) + E &= 0
\end{aligned}$$

In matrix form:

$$\begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \\ E \end{pmatrix} = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

Application

Best fit ellipse, continued

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

Application

Best fit ellipse, continued

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Application

Best fit ellipse, continued

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Row reduce:

$$\left(\begin{array}{ccccc|c} 35 & 6 & -4 & 1 & 11 & -18 \\ 6 & 18 & 10 & -4 & 0 & 18 \\ -4 & 10 & 15 & 0 & -1 & 19 \\ 1 & -4 & 0 & 11 & 1 & -10 \\ 11 & 0 & -1 & 1 & 5 & -15 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 16/7 \\ 0 & 1 & 0 & 0 & 0 & -8/7 \\ 0 & 0 & 1 & 0 & 0 & 15/7 \\ 0 & 0 & 0 & 1 & 0 & -6/7 \\ 0 & 0 & 0 & 0 & 1 & -52/7 \end{array}\right)$$

Application

Best fit ellipse, continued

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Row reduce:

$$\left(\begin{array}{ccccc|c} 35 & 6 & -4 & 1 & 11 & -18 \\ 6 & 18 & 10 & -4 & 0 & 18 \\ -4 & 10 & 15 & 0 & -1 & 19 \\ 1 & -4 & 0 & 11 & 1 & -10 \\ 11 & 0 & -1 & 1 & 5 & -15 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 16/7 \\ 0 & 1 & 0 & 0 & 0 & -8/7 \\ 0 & 0 & 1 & 0 & 0 & 15/7 \\ 0 & 0 & 0 & 1 & 0 & -6/7 \\ 0 & 0 & 0 & 0 & 1 & -52/7 \end{array}\right)$$

Best fit ellipse:

$$x^2 + \frac{16}{7}y^2 - \frac{8}{7}xy + \frac{15}{7}x - \frac{6}{7}y - \frac{52}{7} = 0$$

Application

Best fit ellipse, continued

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Row reduce:

$$\left(\begin{array}{ccccc|c} 35 & 6 & -4 & 1 & 11 & -18 \\ 6 & 18 & 10 & -4 & 0 & 18 \\ -4 & 10 & 15 & 0 & -1 & 19 \\ 1 & -4 & 0 & 11 & 1 & -10 \\ 11 & 0 & -1 & 1 & 5 & -15 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 16/7 \\ 0 & 1 & 0 & 0 & 0 & -8/7 \\ 0 & 0 & 1 & 0 & 0 & 15/7 \\ 0 & 0 & 0 & 1 & 0 & -6/7 \\ 0 & 0 & 0 & 0 & 1 & -52/7 \end{array}\right)$$

Best fit ellipse:

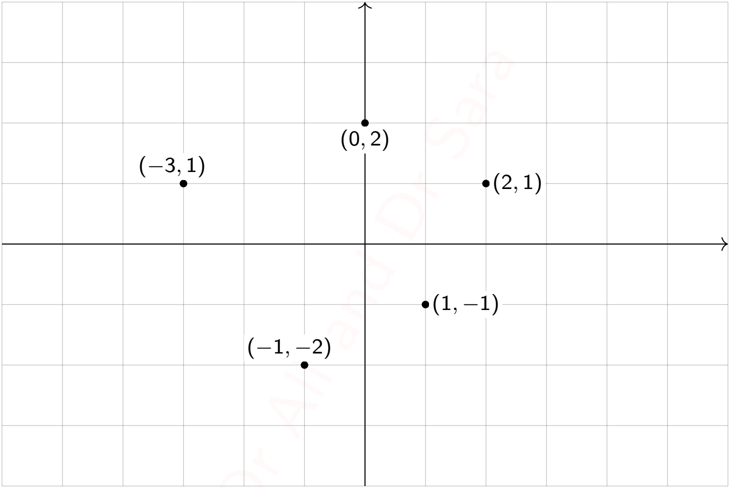
$$x^2 + \frac{16}{7}y^2 - \frac{8}{7}xy + \frac{15}{7}x - \frac{6}{7}y - \frac{52}{7} = 0$$

or

$$7x^2 + 16y^2 - 8xy + 15x - 6y - 52 = 0.$$

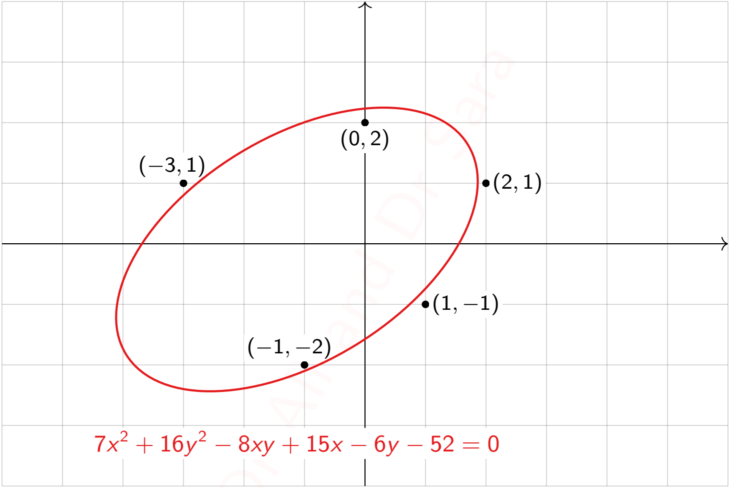
Application

Best fit ellipse, picture



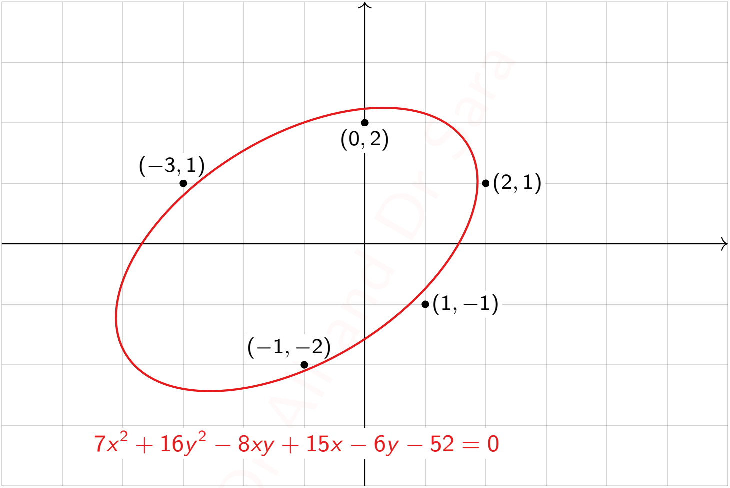
Application

Best fit ellipse, picture



Application

Best fit ellipse, picture



Remark: Gauss invented the method of least squares to do exactly this: he predicted the (elliptical) orbit of the asteroid Ceres as it passed behind the sun in 1801.

What least squares problem $Ax = b$ finds the best parabola through the points $(-1, 0.5)$, $(1, -1)$, $(2, -0.5)$, $(3, 2)$?

The general equation for a parabola is

$$y = Ax^2 + Bx + C.$$

So we want to solve:

$$\begin{aligned} 0.5 &= A(-1)^2 + B(-1) + C \\ -1 &= A(1)^2 + B(1) + C \\ -0.5 &= A(2)^2 + B(2) + C \\ 2 &= A(3)^2 + B(3) + C \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1 \\ -0.5 \\ 2 \end{pmatrix}.$$

Application

Best fit parabola

What least squares problem $Ax = b$ finds the best parabola through the points $(-1, 0.5)$, $(1, -1)$, $(2, -0.5)$, $(3, 2)$?

The general equation for a parabola is

$$y = Ax^2 + Bx + C.$$

So we want to solve:

$$\begin{aligned} 0.5 &= A(-1)^2 + B(-1) + C \\ -1 &= A(1)^2 + B(1) + C \\ -0.5 &= A(2)^2 + B(2) + C \\ 2 &= A(3)^2 + B(3) + C \end{aligned}$$

In matrix form:

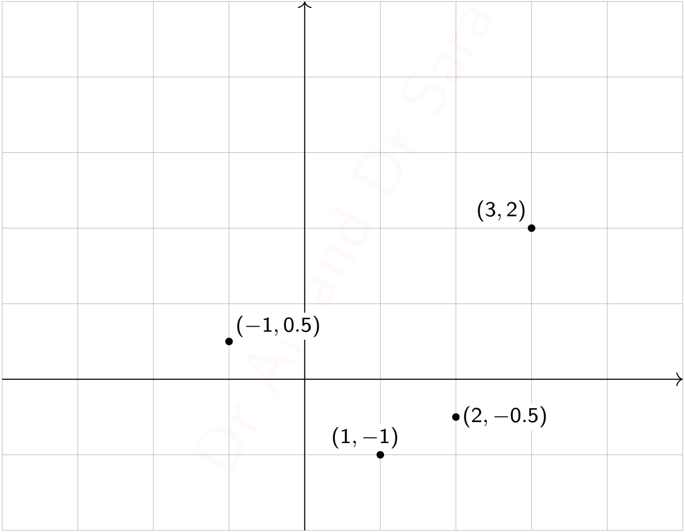
$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1 \\ -0.5 \\ 2 \end{pmatrix}.$$

Answer:

$$88y = 53x^2 - \frac{379}{5}x - 82$$

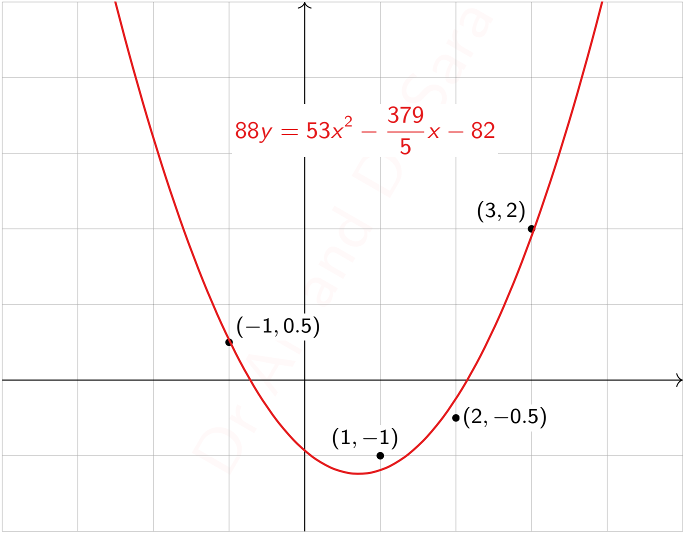
Application

Best fit parabola, picture



Application

Best fit parabola, picture



What least squares problem $Ax = b$ finds the best linear function $f(x, y)$ fitting the following data?

x	y	$f(x,y)$
1	0	0
0	1	1
-1	0	3
0	-1	4

The general equation for a linear function in two variables is

$$f(x, y) = Ax + By + C.$$

So we want to solve

$$\begin{aligned} A(1) + \quad B(0) + C &= 0 \\ A(0) + \quad B(1) + C &= 1 \\ A(-1) + \quad B(0) + C &= 3 \\ A(0) + B(-1) + C &= 4 \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}.$$

What least squares problem $Ax = b$ finds the best linear function $f(x, y)$ fitting the following data?

x	y	$f(x,y)$
1	0	0
0	1	1
-1	0	3
0	-1	4

The general equation for a linear function in two variables is

$$f(x, y) = Ax + By + C.$$

So we want to solve

$$\begin{aligned} A(1) + \quad B(0) + C &= 0 \\ A(0) + \quad B(1) + C &= 1 \\ A(-1) + \quad B(0) + C &= 3 \\ A(0) + B(-1) + C &= 4 \end{aligned}$$

In matrix form:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}.$$

Answer:

$$f(x, y) = -\frac{3}{2}x - \frac{3}{2}y + 2$$