Mathematics for Computer Science
Linear Algebra (Part 2)
Least Squares

Karl Southern

Durham University

February 24th, 2025

Thanks to Andrei Krokhin and William Moses for use of slides.

Outline

1. Plan for Today
2. QR Decomposition of Matrices
3. Least Squares Solutions of Inconsistent Linear Systems
   - Setting Things Up
   - Actually Finding a Least Squares Solution
4. Wrapping Things Up

Roadmap for Lectures 5-8

- **End Goal:** Application - linear regression.
- **Using:** Least Squares.
- **Using:** QR decomposition.
- Requires knowledge of some basics: Inner product spaces and Gram-Schmidt Orthogonalisation.

Roadmap for Lectures 5-8

- **End Goal:** Application - linear regression.
- **Using:** Least Squares.
- **Using:** QR decomposition.
- Requires knowledge of some basics: Inner product spaces and Gram-Schmidt Orthogonalisation.

Now we recap last lectures & look at what we'll cover today.

Last Lecture Reminder

- Orthonormal = consisting of unit vectors, which are pairwise orthogonal
- Every finite-dimensional inner product space V has an orthonormal basis, which can be constructed from any basis of V via the Gram-Schmidt process.

Last Lecture Reminder

- Orthonormal = consisting of unit vectors, which are pairwise orthogonal
- Every finite-dimensional inner product space V has an orthonormal basis, which can be constructed from any basis of V via the Gram-Schmidt process.
- For a subspace W of an inner product space V, its orthogonal complement is

$$W^{\perp} = \{\mathbf{x} \in V \mid \langle \mathbf{u}, \mathbf{x} \rangle = 0 \text{ for all } \mathbf{u} \in W\}$$

- If V (or even only W) is finite-dimensional then every vector $\mathbf{u} \in V$ can be uniquely expressed as $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$ where $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W^{\perp}$.

  Then $\mathbf{w}_1 = \mathrm{proj}_W \mathbf{u}$ is the orthogonal projection of $\mathbf{u}$ onto W.

Today's Lecture Contents

- QR decomposition
- Least squares solutions of inconsistent linear systems

Outline

1 Plan for Today

2 QR Decomposition of Matrices

3 Least Squares Solutions of Inconsistent Linear Systems
- Setting Things Up
- Actually Finding a Least Squares Solution

4 Wrapping Things Up

QR Decomposition

Let A be an $m \times n$ matrix with linearly independent columns $\mathbf{u}_1, \dots, \mathbf{u}_n$

Let $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ be the orthonormal set obtained by applying Gram-Schmidt to $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. How does $A = [\mathbf{u}_1 | \dots | \mathbf{u}_n]$ relate to the matrix $Q = [\mathbf{q}_1 | \dots | \mathbf{q}_n]$?

QR Decomposition

Let A be an $m \times n$ matrix with linearly independent columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$

Let $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ be the orthonormal set obtained by applying Gram-Schmidt to $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$.
How does $A = [\mathbf{u}_1| \ldots |\mathbf{u}_n]$ relate to the matrix $Q = [\mathbf{q}_1| \ldots |\mathbf{q}_n]$?

Since $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ is an orthonormal basis for $span(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, we have

$$
\begin{aligned}
\mathbf{u}_1 &= \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_1, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_1, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 &= \langle \mathbf{u}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_2, \mathbf{q}_n \rangle \mathbf{q}_n \\
&\vdots \\
\mathbf{u}_n &= \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{aligned}
$$

QR Decomposition

$$
\begin{aligned}
\mathbf{u}_1 &= \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_1, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_1, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 &= \langle \mathbf{u}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_2, \mathbf{q}_n \rangle \mathbf{q}_n \\
&\vdots \\
\mathbf{u}_n &= \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{aligned}
$$

QR Decomposition

$$\begin{aligned}
\mathbf{u}_1 &= \langle \mathbf{u}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_1, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_1, \mathbf{q}_n \rangle \mathbf{q}_n \\
\mathbf{u}_2 &= \langle \mathbf{u}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_2, \mathbf{q}_n \rangle \mathbf{q}_n \\
&\vdots \\
\mathbf{u}_n &= \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \mathbf{q}_2 + \ldots + \langle \mathbf{u}_n, \mathbf{q}_n \rangle \mathbf{q}_n
\end{aligned}$$

or, in the matrix form,

$$A = [\mathbf{u}_1 | \ldots | \mathbf{u}_n] = [\mathbf{q}_1 | \ldots | \mathbf{q}_n] \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix} = QR.$$

QR Decomposition

What can we say about the matrix R?

From Gram-Schmidt, for each $j \geq 2$, $\mathbf{q}_j$ is orthogonal to $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$. Hence R is

$$\begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix} = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix}.$$

QR Decomposition

What can we say about the matrix R?

From Gram-Schmidt, for each $j \geq 2$, $\mathbf{q}_j$ is orthogonal to $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$. Hence R is

$$\begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix} = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix}.$$

From Gram-Schmidt, $\langle \mathbf{u}_i, \mathbf{q}_i \rangle = \langle \mathbf{u}_i, \frac{1}{||\mathbf{v_i}||}\mathbf{v}_i \rangle = \frac{1}{||\mathbf{v_i}||}\langle \mathbf{u}_i, \mathbf{v}_i \rangle = \frac{1}{||\mathbf{v_i}||}\langle \mathbf{v}_i, \mathbf{v}_i \rangle > 0$.

QR Decomposition

What can we say about the matrix R?

From Gram-Schmidt, for each $j \geq 2$, $\mathbf{q}_j$ is orthogonal to $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$. Hence R is

$$\begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{q}_2 \rangle & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \mathbf{u}_1, \mathbf{q}_n \rangle & \langle \mathbf{u}_2, \mathbf{q}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix} = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{q}_2 \rangle \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{u}_n, \mathbf{q}_n \rangle \end{pmatrix}.$$

From Gram-Schmidt, $\langle \mathbf{u}_i, \mathbf{q}_i \rangle = \langle \mathbf{u}_i, \frac{1}{||\mathbf{v_i}||} \mathbf{v}_i \rangle = \frac{1}{||\mathbf{v_i}||} \langle \mathbf{u}_i, \mathbf{v}_i \rangle = \frac{1}{||\mathbf{v_i}||} \langle \mathbf{v}_i, \mathbf{v}_i \rangle > 0.$

Theorem (QR Decomposition)

If A is an $m \times n$ matrix with linearly independent column vectors, then it can be factored as

$$A = QR$$

where Q has orthonormal columns and R is an invertible upper triangular matrix.

QR Decomposition

Theorem (QR Decomposition)

If A is an $m \times n$ matrix with linearly independent column vectors, then it can be factored as

$$A = QR$$

where Q has orthonormal columns and R is an invertible upper triangular matrix.

QR Decomposition

Theorem (QR Decomposition)

If A is an $m \times n$ matrix with linearly independent column vectors, then it can be factored as

$$A = QR$$

where Q has orthonormal columns and R is an invertible upper triangular matrix.

For $m = n$, this theorem says that every invertible matrix has a QR-decomposition.

Example 17.1 QR decomposition

Let $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$, find a QR decomposition of A.

Step 1. From Ex 16.2 we have: $\mathbf{q}_1 = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, $\mathbf{q}_2 = (-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}})$, $\mathbf{q}_3 = (0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$

Step 2. Compute $\begin{pmatrix} \langle \mathbf{u}_1, \mathbf{q}_1 \rangle & \langle \mathbf{u}_2, \mathbf{q}_1 \rangle & \langle \mathbf{u}_3, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{u}_2, \mathbf{q}_2 \rangle & \langle \mathbf{u}_3, \mathbf{q}_2 \rangle \\ 0 & 0 & \langle \mathbf{u}_3, \mathbf{q}_3 \rangle \end{pmatrix} = \begin{pmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$

Step 3. $A = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ 0 & \sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$

Outline

1. Plan for Today
2. QR Decomposition of Matrices
3. Least Squares Solutions of Inconsistent Linear Systems
   - Setting Things Up
   - Actually Finding a Least Squares Solution
4. Wrapping Things Up

Outline

1. Plan for Today
2. QR Decomposition of Matrices
3. Least Squares Solutions of Inconsistent Linear Systems
   - Setting Things Up
   - Actually Finding a Least Squares Solution
4. Wrapping Things Up

Reminder: Column Space of a Matrix

Let $A = [\mathbf{u}_1|\mathbf{u}_2|\dots|\mathbf{u}_n]$ be an $m \times n$ matrix with column vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$.

Reminder: Column Space of a Matrix

Let $A = [\mathbf{u}_1 | \mathbf{u}_2 | \dots | \mathbf{u}_n]$ be an $m \times n$ matrix with column vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$.

The column space of A is a subspace of $\mathbb{R}^m$, denoted by $\mathcal{C}(A)$ and defined as

$$\mathcal{C}(A) = span(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = \{x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \dots + x_n\mathbf{u}_n \mid x_1, \dots, x_n \in \mathbb{R}\}$$

Reminder: Column Space of a Matrix

Let $A = [\mathbf{u}_1|\mathbf{u}_2|\ldots|\mathbf{u}_n]$ be an $m \times n$ matrix with column vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

The column space of A is a subspace of $\mathbb{R}^m$, denoted by $\mathcal{C}(A)$ and defined as

$$\mathcal{C}(A) = span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \ldots + x_n\mathbf{u}_n \mid x_1, \ldots, x_n \in \mathbb{R}\}$$

which we can re-write in matrix notation as

$$\mathcal{C}(A) = span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{[\mathbf{u}_1|\mathbf{u}_2|\ldots|\mathbf{u}_n]\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mid \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n\}$$

Reminder: Column Space of a Matrix

Let $A = [\mathbf{u}_1|\mathbf{u}_2|\ldots|\mathbf{u}_n]$ be an $m \times n$ matrix with column vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

The column space of A is a subspace of $\mathbb{R}^m$, denoted by $\mathcal{C}(A)$ and defined as

$$\mathcal{C}(A) = span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \ldots + x_n\mathbf{u}_n \mid x_1, \ldots, x_n \in \mathbb{R}\}$$

which we can re-write in matrix notation as

$$\mathcal{C}(A) = span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{[\mathbf{u}_1|\mathbf{u}_2|\ldots|\mathbf{u}_n]\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mid \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n\}$$

and hence also as

$$\mathcal{C}(A) = span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}.$$

Solving Inconsistent Linear Systems

- Assume that we have an inconsistent linear system $A\mathbf{x} = \mathbf{b}$.
- Can we find a vector that comes as close as possible to a being a solution?

Solving Inconsistent Linear Systems

- Assume that we have an inconsistent linear system $A\mathbf{x} = \mathbf{b}$.
- Can we find a vector that comes as close as possible to a being a solution?

Definition (Least Squares Problem)

Given a linear system $A\mathbf{x} = \mathbf{b}$ with m equations and n variables, find a vector $\mathbf{x}$ that minimises $||\mathbf{b} - A\mathbf{x}||$ (w.r.t. the Euclidean inner product on $\mathbb{R}^m$).

We call such a vector $\mathbf{x}$ a least squares solution to the system, the vector $\mathbf{b} - A\mathbf{x}$ is the least squares error vector, and the number $||\mathbf{b} - A\mathbf{x}||$ is the least squares error.

Solving Inconsistent Linear Systems

- Assume that we have an inconsistent linear system $A\mathbf{x} = \mathbf{b}$.
- Can we find a vector that comes as close as possible to a being a solution?

Definition (Least Squares Problem)

Given a linear system $A\mathbf{x} = \mathbf{b}$ with m equations and n variables, find a vector $\mathbf{x}$ that minimises $||\mathbf{b} - A\mathbf{x}||$ (w.r.t. the Euclidean inner product on $\mathbb{R}^m$).

We call such a vector $\mathbf{x}$ a least squares solution to the system, the vector $\mathbf{b} - A\mathbf{x}$ is the least squares error vector, and the number $||\mathbf{b} - A\mathbf{x}||$ is the least squares error.

"Least squares" - because the norm is the (square root of the) sum of squares:

if $A\mathbf{x} = \mathbf{a}$ then $||\mathbf{b} - A\mathbf{x}|| = ||\mathbf{b} - \mathbf{a}|| = \sqrt{(b_1 - a_1)^2 + \ldots + (b_m - a_m)^2}$.

Solving Inconsistent Linear Systems

- Assume that we have an inconsistent linear system $A\mathbf{x} = \mathbf{b}$.
- Can we find a vector that comes as close as possible to a being a solution?

Definition (Least Squares Problem)

Given a linear system $A\mathbf{x} = \mathbf{b}$ with m equations and n variables, find a vector $\mathbf{x}$ that minimises $||\mathbf{b} - A\mathbf{x}||$ (w.r.t. the Euclidean inner product on $\mathbb{R}^m$).

We call such a vector $\mathbf{x}$ a least squares solution to the system, the vector $\mathbf{b} - A\mathbf{x}$ is the least squares error vector, and the number $||\mathbf{b} - A\mathbf{x}||$ is the least squares error.

"Least squares" - because the norm is the (square root of the) sum of squares:

if $A\mathbf{x} = \mathbf{a}$ then $||\mathbf{b} - A\mathbf{x}|| = ||\mathbf{b} - \mathbf{a}|| = \sqrt{(b_1 - a_1)^2 + \ldots + (b_m - a_m)^2}$.

If we trust different measurements/equations differently, we can use the weighted Euclidean inner product to compute the norm and get the weighted least squares.

Best Approximation Theorem

Theorem

If W is a finite-dimensional subspace in an inner product space V and $\mathbf{b} \in V$ then $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W in the sense that

$$||\mathbf{b} - \mathrm{proj}_W \mathbf{b}|| \leq ||\mathbf{b} - \mathbf{w}||$$

for each vector $\mathbf{w} \in W$, and the inequality is strict for all $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$.

Best Approximation Theorem

Theorem

If W is a finite-dimensional subspace in an inner product space V and $\mathbf{b} \in V$ then $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W in the sense that

$$||\mathbf{b} - \mathrm{proj}_W \mathbf{b}|| \leq ||\mathbf{b} - \mathbf{w}||$$

for each vector $\mathbf{w} \in W$, and the inequality is strict for all $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$.

Proof.

For any vector $\mathbf{w} \in W$, write $\mathbf{b} - \mathbf{w} = \underbrace{(\mathbf{b} - \mathrm{proj}_W \mathbf{b})}_{\text{in } W^{\perp}} + \underbrace{(\mathrm{proj}_W \mathbf{b} - \mathbf{w})}_{\text{in } W}$.

Best Approximation Theorem

Theorem

If W is a finite-dimensional subspace in an inner product space V and $\mathbf{b} \in V$ then $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W in the sense that

$$||\mathbf{b} - \mathrm{proj}_W \mathbf{b}|| \leq ||\mathbf{b} - \mathbf{w}||$$

for each vector $\mathbf{w} \in W$, and the inequality is strict for all $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$.

Proof.

For any vector $\mathbf{w} \in W$, write $\mathbf{b} - \mathbf{w} = \underbrace{(\mathbf{b} - \mathrm{proj}_W \mathbf{b})}_{\text{in } W^{\perp}} + \underbrace{(\mathrm{proj}_W \mathbf{b} - \mathbf{w})}_{\text{in } W}$.

By Pythagoras' theorem (if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ then $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$),

Best Approximation Theorem

Theorem

If W is a finite-dimensional subspace in an inner product space V and $\mathbf{b} \in V$ then $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W in the sense that

$$||\mathbf{b} - \mathrm{proj}_W \mathbf{b}|| \leq ||\mathbf{b} - \mathbf{w}||$$

for each vector $\mathbf{w} \in W$, and the inequality is strict for all $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$.

Proof.

For any vector $\mathbf{w} \in W$, write $\mathbf{b} - \mathbf{w} = \underbrace{(\mathbf{b} - \mathrm{proj}_W \mathbf{b})}_{\text{in } W^{\perp}} + \underbrace{(\mathrm{proj}_W \mathbf{b} - \mathbf{w})}_{\text{in } W}$.

By Pythagoras' theorem (if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ then $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$), we have
$||\mathbf{b} - \mathbf{w}||^2 = ||\mathbf{b} - \mathrm{proj}_W \mathbf{b}||^2 + ||\mathrm{proj}_W \mathbf{b} - \mathbf{w}||^2$

Best Approximation Theorem

Theorem

If W is a finite-dimensional subspace in an inner product space V and $\mathbf{b} \in V$ then $\mathrm{proj}_W \mathbf{b}$ is the best approximation *to $\mathbf{b}$ from W in the sense that*

$$||\mathbf{b} - \mathrm{proj}_W \mathbf{b}|| \leq ||\mathbf{b} - \mathbf{w}||$$

for each vector $\mathbf{w} \in W$, and the inequality is strict for all $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$.

Proof.

For any vector $\mathbf{w} \in W$, write $\mathbf{b} - \mathbf{w} = \underbrace{(\mathbf{b} - \mathrm{proj}_W \mathbf{b})}_{\text{in } W^\perp} + \underbrace{(\mathrm{proj}_W \mathbf{b} - \mathbf{w})}_{\text{in } W}$.

By Pythagoras' theorem (if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ then $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$), we have $||\mathbf{b} - \mathbf{w}||^2 = ||\mathbf{b} - \mathrm{proj}_W \mathbf{b}||^2 + ||\mathrm{proj}_W \mathbf{b} - \mathbf{w}||^2 \geq ||\mathbf{b} - \mathrm{proj}_W \mathbf{b}||^2$.

Best Approximation Theorem

Theorem

If W is a finite-dimensional subspace in an inner product space V and $\mathbf{b} \in V$ then $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W in the sense that

$$||\mathbf{b} - \mathrm{proj}_W \mathbf{b}|| \leq ||\mathbf{b} - \mathbf{w}||$$

for each vector $\mathbf{w} \in W$, and the inequality is strict for all $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$.

Proof.

For any vector $\mathbf{w} \in W$, write $\mathbf{b} - \mathbf{w} = \underbrace{(\mathbf{b} - \mathrm{proj}_W \mathbf{b})}_{\text{in } W^\perp} + \underbrace{(\mathrm{proj}_W \mathbf{b} - \mathbf{w})}_{\text{in } W}.$

By Pythagoras' theorem (if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ then $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$), we have
$||\mathbf{b} - \mathbf{w}||^2 = ||\mathbf{b} - \mathrm{proj}_W \mathbf{b}||^2 + ||\mathrm{proj}_W \mathbf{b} - \mathbf{w}||^2 \geq ||\mathbf{b} - \mathrm{proj}_W \mathbf{b}||^2.$
Moreover, the inequality is strict whenever $\mathbf{w} \neq \mathrm{proj}_W \mathbf{b}$. $\square$

Outline

1. Plan for Today
2. QR Decomposition of Matrices
3. Least Squares Solutions of Inconsistent Linear Systems
   - Setting Things Up
   - Actually Finding a Least Squares Solution
4. Wrapping Things Up

Least Squares Solutions of Linear Systems

- Let $W = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ be the column space of A.
- Since $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W, least squares solutions to $A\mathbf{x} = \mathbf{b}$ (i.e., vectors $\mathbf{x}$ minimising $||\mathbf{b} - A\mathbf{x}||$) are exactly solutions to

$$A\mathbf{x} = \mathrm{proj}_W \mathbf{b}.$$

- We can compute $\mathrm{proj}_W \mathbf{b}$ and solve the system, but there's a more useful way.

Least Squares Solutions of Linear Systems

- Let $W = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ be the column space of A.
- Since $\mathrm{proj}_W\mathbf{b}$ is the best approximation to $\mathbf{b}$ from W, least squares solutions to $A\mathbf{x} = \mathbf{b}$ (i.e., vectors $\mathbf{x}$ minimising $||\mathbf{b} - A\mathbf{x}||$) are exactly solutions to

$$A\mathbf{x} = \mathrm{proj}_W\mathbf{b}.$$

- We can compute $\mathrm{proj}_W\mathbf{b}$ and solve the system, but there's a more useful way.
- The representation $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$ where $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W^\perp$ is unique, so the equation $A\mathbf{x} = \mathrm{proj}_W\mathbf{b}$ is equivalent to the condition $\mathbf{b} - A\mathbf{x} \in W^\perp$.

Least Squares Solutions of Linear Systems

- Let $W = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ be the column space of A.
- Since $\mathrm{proj}_W\mathbf{b}$ is the best approximation to $\mathbf{b}$ from W, least squares solutions to $A\mathbf{x} = \mathbf{b}$ (i.e., vectors $\mathbf{x}$ minimising $||\mathbf{b} - A\mathbf{x}||$) are exactly solutions to
$$A\mathbf{x} = \mathrm{proj}_W\mathbf{b}.$$
- We can compute $\mathrm{proj}_W\mathbf{b}$ and solve the system, but there's a more useful way.
- The representation $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$ where $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W^\perp$ is unique, so the equation $A\mathbf{x} = \mathrm{proj}_W\mathbf{b}$ is equivalent to the condition $\mathbf{b} - A\mathbf{x} \in W^\perp$.
- The columns of A are the rows of A^T, so the condition $\mathbf{b} - A\mathbf{x} \in W^\perp$ is equivalent to $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$,

Least Squares Solutions of Linear Systems

- Let $W = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$ be the column space of A.
- Since $\mathrm{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W, least squares solutions to $A\mathbf{x} = \mathbf{b}$ (i.e., vectors $\mathbf{x}$ minimising $||\mathbf{b} - A\mathbf{x}||$) are exactly solutions to

$$A\mathbf{x} = \mathrm{proj}_W \mathbf{b}.$$

- We can compute $\mathrm{proj}_W \mathbf{b}$ and solve the system, but there's a more useful way.
- The representation $\mathbf{b} = \mathbf{w}_1 + \mathbf{w}_2$ where $\mathbf{w}_1 \in W$ and $\mathbf{w}_2 \in W^\perp$ is unique, so the equation $A\mathbf{x} = \mathrm{proj}_W \mathbf{b}$ is equivalent to the condition $\mathbf{b} - A\mathbf{x} \in W^\perp$.
- The columns of A are the rows of A^T, so the condition $\mathbf{b} - A\mathbf{x} \in W^\perp$ is equivalent to $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, which we can re-write as

$$A^T A\mathbf{x} = A^T \mathbf{b}.$$

- This is the normal equation (or normal system) associated with $A\mathbf{x} = \mathbf{b}$.

Least Squares Solutions of Linear Systems

On the previous slide, we proved the following.

Theorem

1. *For every linear system* $A\mathbf{x} = \mathbf{b}$*, the associated normal system* $A^T A\mathbf{x} = A^T\mathbf{b}$ *is consistent, and its solutions are exactly least square solutions of* $A\mathbf{x} = \mathbf{b}$*.*
2. *Moreover, if* W *is the column space of* A *and* $\mathbf{x}_0$ *is any least squares solution of* $A\mathbf{x} = \mathbf{b}$ *then* $A\mathbf{x}_0 = \mathrm{proj}_W\mathbf{b}$*.*

Example 17.2: Computing Least Squares Solutions

Find least squares solutions for the linear system, using the euclidean dot product.

$$\begin{aligned} x_1 - x_2 &= 4 \\ 3x_1 + 2x_2 &= 1 \\ -2x_1 + 4x_2 &= 3 \end{aligned} \qquad A\mathbf{x} = \begin{pmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix} = \mathbf{b}.$$

The associated normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ is

$$\begin{pmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 3 & 2 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 3 & -2 \\ -1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}.$$

Computing the matrix products, we get

$$\begin{pmatrix} 14 & -3 \\ -3 & 21 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 10 \end{pmatrix}.$$

Solving this yields a unique least squares solution $x_1 = 17/95$ and $x_2 = 143/285$.

(If needed, can now easily compute the error vector $\mathbf{b} - A\mathbf{x}$ and error $||\mathbf{b} - A\mathbf{x}||$.)

Outline

1. Plan for Today
2. QR Decomposition of Matrices
3. Least Squares Solutions of Inconsistent Linear Systems
   - Setting Things Up
   - Actually Finding a Least Squares Solution
4. Wrapping Things Up

Example exam question

(b) Give the QR decomposition of $A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \\ 2 & 1 \\ 2 & 1 \end{pmatrix}$. **[10 Marks]**

Wrapping Things Up

Today:

- QR decomposition
- Least squares fitting to data

Next time:

- Linear regression

The End

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^TA\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff A^TA is invertible.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^TA\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff A^TA is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- A^TA is square, so it is invertible iff $A^TA\mathbf{x} = \mathbf{0}$ has only the trivial solution.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- $A^T A$ is square, so it is invertible iff $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^T A\mathbf{x} = \mathbf{0}$.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^TA\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff A^TA is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- A^TA is square, so it is invertible iff $A^TA\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^TA\mathbf{x} = \mathbf{0}$.
- Let $\mathbf{x}_0$ be a solution of $A^TA\mathbf{x} = \mathbf{0}$, i.e., $A^TA\mathbf{x}_0 = \mathbf{0}$.

  Then $A\mathbf{x}_0$ is both in the column space of A and in the null space of A^T (which are orthogonal complements of each other). Hence, $A\mathbf{x}_0 = \mathbf{0}$.

Supplemental Slide: When are Least Squares Solutions Unique?

Equivalently: when does a normal system $A^T A\mathbf{x} = A^T\mathbf{b}$ have a unique solution?

Theorem

For any $m \times n$ matrix A, A has linearly independent columns iff $A^T A$ is invertible.

Proof.

- The columns of A are linearly indep iff $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- $A^T A$ is square, so it is invertible iff $A^T A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
- Each solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^T A\mathbf{x} = \mathbf{0}$.
- Let $\mathbf{x}_0$ be a solution of $A^T A\mathbf{x} = \mathbf{0}$, i.e., $A^T A\mathbf{x}_0 = \mathbf{0}$.

  Then $A\mathbf{x}_0$ is both in the column space of A and in the null space of A^T (which are orthogonal complements of each other). Hence, $A\mathbf{x}_0 = \mathbf{0}$.
- Thus $A\mathbf{x} = \mathbf{0}$ has only the trivial solution iff the same is true for $A^T A\mathbf{x} = \mathbf{0}$.

$\square$