Please fill out the CIOS form online.

- It is important for me to get responses from most of the class: I use these for preparing future iterations of this course.
- If we get an 80% response rate before the final, I'll drop the two lowest quiz grades instead of one.
- Office hours: Wednesday 5–6pm, Friday 10:00–12:00pm
 - As always, TAs' office hours are posted on the website.
 - Math Lab is also a good place to visit.
 - Extra review sessions will be announced later.
- There is no quiz on Friday, but this will be the only opportunity to discuss chapter 6 in recitation.
- ► I will post details about the final exam, a practice final by Monday
- WeBWorK assignments 6.1, 6.2, 6.3 are due today.
- WeBWorK assignments 6.4 and 6.5, if posted, are only for practice—the scores do not count.

Section 6.5

Least Squares Problems

Motivation

The motivating problem of last part of the course:

Problem

Suppose that Ax = b does not have a solution. What is the *best possible approximate* solution?

Saying Ax = b has no solution means that b is not in Col A.

- Using $\hat{b} = \operatorname{proj}_{\operatorname{Col} A}(b)$, then $A\hat{x} = \hat{b}$ is a *consistent equation*.
- **Plus:** \hat{b} is the *closest vector to b* such that $A\hat{x} = \hat{b}$ is consistent.

A solution
$$\hat{x}$$
 to $A\hat{x} = \hat{b}$ is a **least squares solution**.

Definition

Let A be an $m \times n$ matrix. A least squares solution to Ax = b is a vector \hat{x} in \mathbb{R}^n such that

$$A\widehat{x} = \widehat{b} = \operatorname{proj}_{\operatorname{Col} A}(b).$$

A least squares solution \hat{x} solves Ax = b as closely as possible.



In *distance terms*, for all x in \mathbb{R}^n :

$$\|b - A\widehat{x}\| \le \|b - Ax\|$$

Theorem

Let A be a $m \times n$ matrix with orthogonal columns v_1, v_2, \ldots, v_n . The least squares solution to Ax = b is the vector

$$\widehat{x} = \left(\frac{b \cdot v_1}{v_1 \cdot v_1}, \frac{b \cdot v_2}{v_2 \cdot v_2}, \cdots, \frac{b \cdot v_n}{v_n \cdot v_n}\right).$$

This is because we have formulas for the \mathcal{B} -coordinates of orthogonal basis:

$$A\widehat{x} = \sum_{i=1}^{n} \frac{b \cdot v_i}{v_i \cdot v_i} v_i = \operatorname{proj}_{\operatorname{Col} A}(b)$$



Theorem Let A be a $m \times n$ matrix. Least squares solutions to Ax = b are any of the solutions to

 $(A^{\mathsf{T}}A)\widehat{x} = A^{\mathsf{T}}b.$

Now we can solve the problem without computing \hat{b} first.

This is just another sysmetm of equations, but now it *is consistent* and uses *square matrix* $A^T A!$



Why is this true?

Least Squares Solutions Example 1

Find the least squares solutions to Ax = b where:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$$

First: Compute new matrix and vector

$$\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$$

Least Squares Solutions Example 2

Find the least squares solutions to Ax = b where:

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

First: Compute new matrix and vector

$$A^{T}A = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix}$$

When does Ax = b have a *unique* least squares solution \hat{x} ?

► A^TA is always a square matrix, but it need not be invertible.

Theorem

Let A be an $m \times n$ matrix. The following *are equivalent*:

- 1. $A^T A$ is invertible.
- 2. The columns of A are *linearly independent*.
- 3. Ax = b has a **unique least squares solution** for all b in **R**ⁿ, which is

$$(A^{\mathsf{T}}A)^{-1}(A^{\mathsf{T}}b).$$



• If the columns of A are *linearly dependent*, then $A\hat{x} = \hat{b}$ has many solutions.

Extra: More details From Example 1

$$A\widehat{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} = \widehat{b}$$



2. If $A^T A$ is invertible: Let v_1, v_2 be the columns of A, and $\mathcal{B} = \{v_1, v_2\}$, then $\hat{x} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$ are the \mathcal{B} -coordinates of \hat{b} , in Col $A = \text{Span}\{v_1, v_2\}$.

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

c + dx = y.

So we want to solve:

 $c + d \cdot 0 = 6$ $c + d \cdot 1 = 0$ $c + d \cdot 2 = 0.$

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 has $\hat{c} = 5$ and $\hat{d} = -3$:

$$y = -3x + 5.$$



Data Modeling: Best fit line

What does it minimize?

Best fit line minimizes the **sum of the squares** of the *vertical distances from the data points* to the line.



Data modeling: 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

$$ax^2 + bx + c = y.$$

So we want to solve:

$$\begin{aligned} a(-1)^2 + b(-1) + c &= 0.5\\ a(1)^2 + b(1) + c &= -1\\ a(2)^2 + b(2) + c &= -0.5\\ a(3)^2 + b(3) + c &= 2 \end{aligned}$$

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: $\hat{a} = \frac{53}{88}, \hat{b} = \frac{379}{440}, \hat{c} = \frac{82}{88}$ so best fit is: $53x^2 - \frac{379}{5}x - 82 = 88y$

Data modeling: best fit parabola Picture



Data modeling: 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:

In matrix form:

$$\begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}$$

Data modeling: best fit ellipse

Complete procedure

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$
$$A^{T}A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^{T}b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Row reduce:

$$\begin{pmatrix} 35 & 6 & -4 & 1 & 11 & | & -18 \\ 6 & 18 & 10 & -4 & 0 & | & 18 \\ -4 & 10 & 15 & 0 & -1 & | & 19 \\ 1 & -4 & 0 & 11 & 1 & | & -10 \\ 11 & 0 & -1 & 1 & 5 & | & -15 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & | & 16/7 \\ 0 & 1 & 0 & 0 & 0 & | & -8/7 \\ 0 & 0 & 1 & 0 & 0 & | & 15/7 \\ 0 & 0 & 0 & 1 & 0 & | & -6/7 \\ 0 & 0 & 0 & 0 & 1 & | & -52/7 \end{pmatrix}$$

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

$$7x^2 + 16y^2 - 8xy + 15x - 6y - 52 = 0$$

Data modeling: 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.

Extra: Best fit linear function

What least squares problem Ax = b finds the best linear function f(x, y) fitting the following data?

x	y y	f(x,y)
1	0	0
0	1	1
$^{-1}$	0	3
0	$^{-1}$	4

So we want to solve

$$a(1) + b(0) + c = 0$$

$$a(0) + b(1) + c = 1$$

$$a(-1) + b(0) + c = 3$$

$$a(0) + b(-1) + c = 4$$

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: $\hat{a} = -\frac{3}{2}, \hat{b} = -\frac{3}{2}, \hat{c} = 2$ so best fit is: $f(x, y) = -\frac{3}{2}x - \frac{3}{2}y + 2$

Extra: Best fit linear function

Picture

