Announcements

Wednesday, November 29

- 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 $\operatorname{Col} A$.

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

Solution

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

Least Squares Solutions

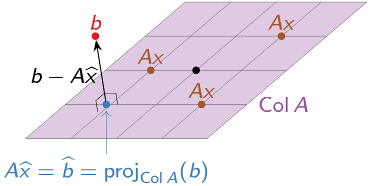
Definition

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

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

> A least squares solution $\widehat{x}$ solves $Ax = b$ *as closely as possible*.

Note that $b - A\widehat{x}$ is in $(\mathrm{Col}\,A)^{\perp}$.



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

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

Least Squares Solutions: Orthogonal case

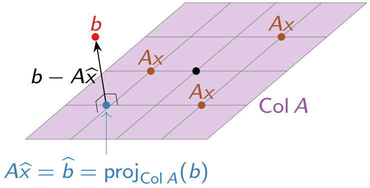
Theorem

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:

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



Least Squares Solutions: General Solution

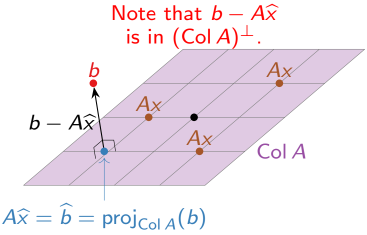
Theorem

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

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

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

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



Why is this true?

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

Now, $b - A\widehat{x}$ is in $(\mathrm{Col}\, A)^\perp$ if and only if

$$A^T(b - A\widehat{x}) = 0.$$

In other words, $A^T A\widehat{x} = A^T b$.

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

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

and

$$A^T b = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

Second: Solve the new system; 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 *unique* least squares *solution is* $\widehat{x} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$.

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

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

and

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

Second: Solve the new system; 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 *unique* 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 $\widehat{x}$?

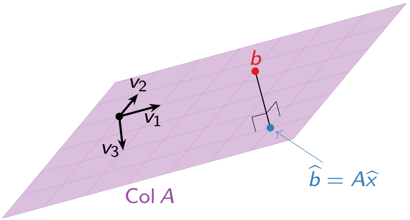
- $A^T A$ is always a square matrix, but it need not be invertible.

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 $\mathbf{R}^n$, which is

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



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

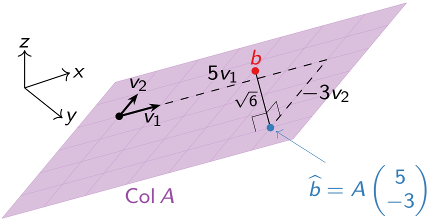
Extra: More details

From Example 1

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

1. The solution $\widehat{x}$ makes the *distance from b to its approximation:*

$$\|b - A\widehat{x}\| = \left\| \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\| = \sqrt{6}.$$



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

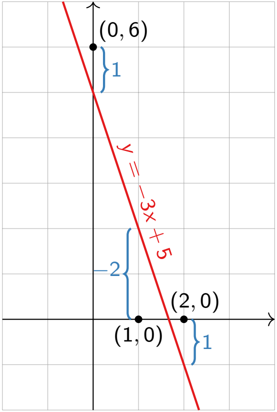
$$\begin{aligned} c + d \cdot 0 &= 6 \\ c + d \cdot 1 &= 0 \\ c + d \cdot 2 &= 0. \end{aligned}$$

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

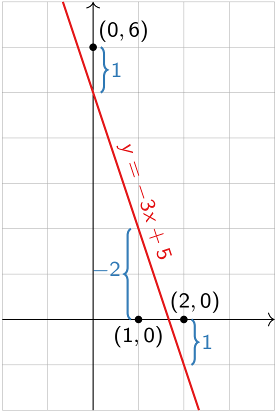
$$y = -3x + 5.$$



$$A \begin{pmatrix} 5 \\ -3 \end{pmatrix} - \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

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:

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

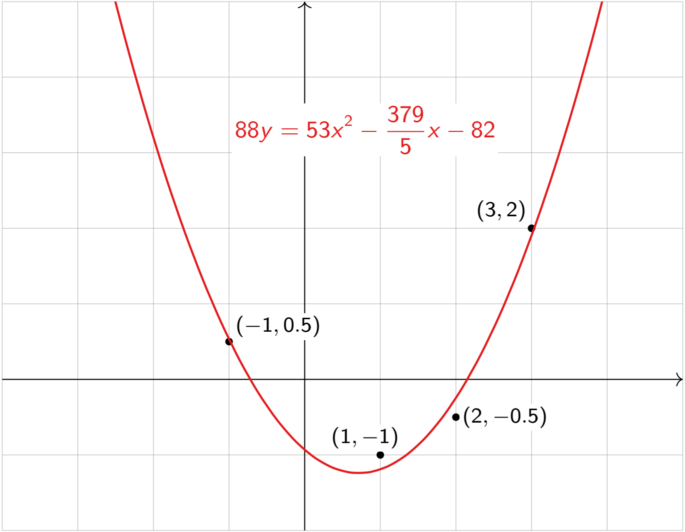
In matrix form:

$$\begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0.5 \\ -1 \\ -0.5 \\ 2 \end{pmatrix}.$$

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

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:

$$\begin{aligned}
(0)^2 + A(2)^2 + B(0)(2) + C(0) + D(2) + E &= 0 \\
(2)^2 + A(1)^2 + B(2)(1) + C(2) + D(1) + E &= 0 \\
(1)^2 + A(-1)^2 + B(1)(-1) + C(1) + D(-1) + E &= 0 \\
(-1)^2 + A(-2)^2 + B(-1)(-2) + C(-1) + D(-2) + E &= 0 \\
(-3)^2 + A(1)^2 + B(-3)(1) + C(-3) + D(1) + E &= 0
\end{aligned}$$

In matrix form:

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

Data modeling: best fit ellipse

Complete procedure

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

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

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:

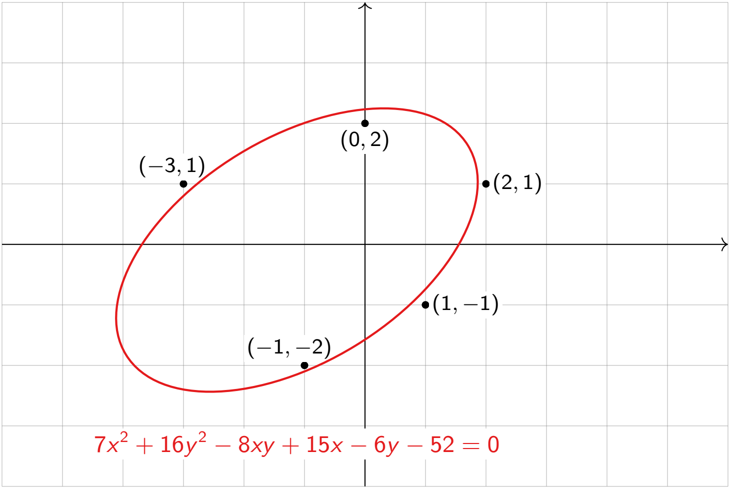
$$x^2 + \frac{16}{7}y^2 - \frac{8}{7}xy + \frac{15}{7}x - \frac{6}{7}y - \frac{52}{7} = 0$$

or

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

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	$f(x,y)$
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

$$\begin{aligned} a(1) + \quad b(0) + c &= 0 \\ a(0) + \quad b(1) + c &= 1 \\ a(-1) + \quad b(0) + c &= 3 \\ a(0) + b(-1) + c &= 4 \end{aligned}$$

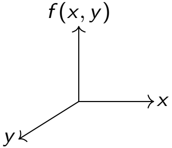
In matrix form:

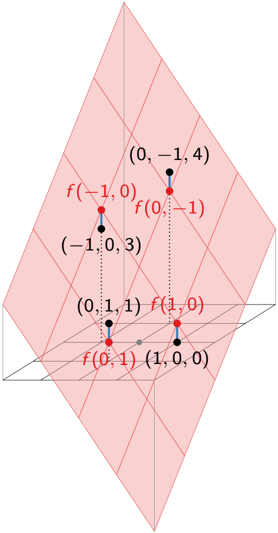
$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}.$$

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

Extra: Best fit linear function

Picture





Graph of

$$f(x,y) = -\frac{3}{2}x - \frac{3}{2}y + 2$$