## 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 $A x=b$ does not have a solution. What is the best possible approximate solution?

Saying $A x=b$ has no solution means that $b$ is not in $\operatorname{Col} A$.

- Using $\widehat{b}=\operatorname{proj}_{\text {Col } A}(b)$, 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 $A x=b$ is a vector $\widehat{x}$ in $\mathbf{R}^{n}$ such that

$$
A \widehat{x}=\widehat{b}=\operatorname{proj}_{C o l} A(b)
$$

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

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


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

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

## 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 $A x=b$ 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:

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



$$
A \widehat{x}=\widehat{b}=\operatorname{proj}_{C o l} A(b)
$$

## Least Squares Solutions: General Solution

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

$$
\left(A^{T} A\right) \hat{x}=A^{T} b .
$$

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 $A^{\top} A$ !


## Least Squares Solutions

## Example 1

Find the least squares solutions to $A x=b$ where:

$$
A=\left(\begin{array}{ll}
1 & 0 \\
1 & 1 \\
1 & 2
\end{array}\right) \quad b=\left(\begin{array}{l}
6 \\
0 \\
0
\end{array}\right)
$$

First: Compute new matrix and vector

$$
A^{\top} A=\left(\begin{array}{lll}
1 & 1 & 1 \\
0 & 1 & 2
\end{array}\right)\left(\begin{array}{ll}
1 & 0 \\
1 & 1 \\
1 & 2
\end{array}\right)=\left(\begin{array}{ll}
3 & 3 \\
3 & 5
\end{array}\right)
$$

## Least Squares Solutions

## Example 2

Find the least squares solutions to $A x=b$ where:

$$
A=\left(\begin{array}{rr}
2 & 0 \\
-1 & 1 \\
0 & 2
\end{array}\right) \quad b=\left(\begin{array}{r}
1 \\
0 \\
-1
\end{array}\right)
$$

First: Compute new matrix and vector

$$
A^{T} A=\left(\begin{array}{rrr}
2 & -1 & 0 \\
0 & 1 & 2
\end{array}\right)\left(\begin{array}{rr}
2 & 0 \\
-1 & 1 \\
0 & 2
\end{array}\right)=\left(\begin{array}{rr}
5 & -1 \\
-1 & 5
\end{array}\right)
$$

## Least Squares Solutions: Uniqueness

When does $A x=b$ 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. $A x=b$ has a unique least squares solution for all $b$ in $\mathbf{R}^{n}$, which is

$$
\left(A^{T} A\right)^{-1}\left(A^{T} b\right)
$$

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


## Extra: More details

$$
A \widehat{x}=\left(\begin{array}{ll}
1 & 0 \\
1 & 1 \\
1 & 2
\end{array}\right)\binom{5}{-3}=\left(\begin{array}{c}
5 \\
2 \\
-1
\end{array}\right)=\widehat{b}
$$


2. If $A^{T} A$ is invertible: Let $v_{1}, v_{2}$ be the columns of $A$, and $\mathcal{B}=\left\{v_{1}, v_{2}\right\}$, then $\widehat{x}=\binom{5}{-3}$ are the $\mathcal{B}$-coordinates of $\widehat{b}$, in $\operatorname{Col} A=\operatorname{Span}\left\{v_{1}, v_{2}\right\}$.

## 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

$$
c+d x=y
$$

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:

$$
\left(\begin{array}{ll}
1 & 0 \\
1 & 1 \\
1 & 2
\end{array}\right)\binom{c}{d}=\left(\begin{array}{l}
6 \\
0 \\
0
\end{array}\right)
$$

We already saw: the least squares solution is

$\binom{5}{-3}$. So the best fit line has $\widehat{c}=5$ and $\widehat{d}=-3$ :

$$
y=-3 x+5
$$

$$
A\binom{5}{-3}-\left(\begin{array}{l}
6 \\
0 \\
0
\end{array}\right)=\left(\begin{array}{c}
1 \\
-2 \\
1
\end{array}\right)
$$

## 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 $A x=b$ finds the best parabola through the points $(-1,0.5),(1,-1),(2,-0.5),(3,2)$ ?
The general equation for a parabola is

$$
a x^{2}+b x+c=y
$$

So we want to solve:

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

In matrix form:

$$
\left(\begin{array}{rrr}
1 & -1 & 1 \\
1 & 1 & 1 \\
4 & 2 & 1 \\
9 & 3 & 1
\end{array}\right)\left(\begin{array}{l}
a \\
b \\
c
\end{array}\right)=\left(\begin{array}{r}
0.5 \\
-1 \\
-0.5 \\
2
\end{array}\right)
$$

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

## 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

$$
x^{2}+a y^{2}+b x y+c x+d y+e=0
$$

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:

$$
\left(\begin{array}{rrrrr}
4 & 0 & 0 & 2 & 1 \\
1 & 2 & 2 & 1 & 1 \\
1 & -1 & 1 & -1 & 1 \\
4 & 2 & -1 & -2 & 1 \\
1 & -3 & -3 & 1 & 1
\end{array}\right)\left(\begin{array}{l}
a \\
b \\
c \\
d \\
e
\end{array}\right)=\left(\begin{array}{r}
0 \\
-4 \\
-1 \\
-1 \\
-9
\end{array}\right)
$$

## Data modeling: best fit ellipse

## Complete procedure

$$
\begin{aligned}
A & =\left(\begin{array}{rrrrr}
4 & 0 & 0 & 2 & 1 \\
1 & 2 & 2 & 1 & 1 \\
1 & -1 & 1 & -1 & 1 \\
4 & 2 & -1 & -2 & 1 \\
1 & -3 & -3 & 1 & 1
\end{array}\right) & b=\left(\begin{array}{r}
0 \\
-4 \\
-1 \\
-1 \\
-9
\end{array}\right) . \\
A^{T} A & =\left(\begin{array}{rrrrr}
35 & 6 & -4 & 1 & 11 \\
6 & 18 & 10 & -4 & 0 \\
-4 & 10 & 15 & 0 & -1 \\
1 & -4 & 0 & 11 & 1 \\
11 & 0 & -1 & 1 & 5
\end{array}\right) & A^{T} b=\left(\begin{array}{r}
-18 \\
18 \\
19 \\
-10 \\
-15
\end{array}\right)
\end{aligned}
$$

Row reduce:

$$
\left(\begin{array}{rrrrr|r}
35 & 6 & -4 & 1 & 11 & -18 \\
6 & 18 & 10 & -4 & 0 & 18 \\
-4 & 10 & 15 & 0 & -1 & 19 \\
1 & -4 & 0 & 11 & 1 & -10 \\
11 & 0 & -1 & 1 & 5 & -15
\end{array}\right) \leadsto\left(\begin{array}{lllll|r}
1 & 0 & 0 & 0 & 0 & 16 / 7 \\
0 & 1 & 0 & 0 & 0 & -8 / 7 \\
0 & 0 & 1 & 0 & 0 & 15 / 7 \\
0 & 0 & 0 & 1 & 0 & -6 / 7 \\
0 & 0 & 0 & 0 & 1 & -52 / 7
\end{array}\right)
$$

Best fit ellipse:

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

or

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

## 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 $A x=b$ 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 |

So we want to solve

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

In matrix form:

$$
\left(\begin{array}{rrr}
1 & 0 & 1 \\
0 & 1 & 1 \\
-1 & 0 & 1 \\
0 & -1 & 1
\end{array}\right)\left(\begin{array}{l}
a \\
b \\
c
\end{array}\right)=\left(\begin{array}{l}
0 \\
1 \\
3 \\
4
\end{array}\right) .
$$

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

## Extra: Best fit linear function



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

