Linear Systems

Class 28: **Friday April 11**

TITLE Least Squares Approximation
CURRENT READING Poole 7.3

Summary
In practice one often can not solve $A\vec{x} = \vec{b}$. Instead one solves the next best thing, $A^T A\vec{x} = A^T\vec{b}$. This is known as the "least squares approximation."

Homework Assignment
HW #27 Poole, Section 7.3 : **1**,*2*,**4**,**9**,*10*,**25**,**36**. *EXTRA CREDIT 56.*

DEFINITION

1. Line of Best Fit
Suppose you have m data points in the xy-plane and you want to find the line which "fits" the data the best. In other words, we want to minimize the sum of the distance of the points from the line.

Suppose we have the points (x_i, y_i) for $i = 1$ to m. If we assume the line has the form $y = dx + c$ then we're looking for m equations in 2 variables, if we say that the line has to go through all m points.

$$\begin{aligned} y_1 &= dx_1 + c \\ y_2 &= dx_2 + c \\ y_3 &= dx_3 + c \\ \vdots &= \vdots \\ y_m &= dx_m + c \end{aligned}$$

We can write this as a matrix as

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

How many solutions does this linear system $A\vec{x} = \vec{b}$ have? (Think: what is the possible rank of A?)

EXAMPLE
Let's let $m = 3$ and consider the points $(0, 6)$, $(1, 0)$ and $(2, 0)$. What's the equation of the line which fits this data the *best*? (What matrix equation does this correspond to?)

There's multiple ways of thinking about the "best fit" problem:

Using Linear Algebra

We can think of the vector $\vec{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ being split into two orthogonal parts, one part $\vec{p}$ in the column space of $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and the other $\vec{e}$ in the left nullspace of A (nullspace of A^T).

$\vec{p}$ is the projection of the vector $\vec{b}$ on to the vector space spanned by the columns of A, i.e. col(A). According to the orthogonal decomposition theorem, $\vec{b} = \vec{p} + \vec{e} \Leftrightarrow \vec{e} = \vec{b} - \vec{p} = \vec{b} - \text{proj}_{\text{col}(A)}(\vec{b}) = \text{perp}_{\text{col}(A)}(\vec{b})$

Again, recall that $\vec{e}$ is the size of the error (or difference) between the projection $\vec{b}$ on the column space of A and the vector $\vec{b}$ Draw the graphical relationship between these vectors $\vec{e}$, $\vec{b}$ and $\vec{p}$:

Another way to describe the column space of A is the set of vectors $A\vec{x}$ where $\vec{x}$ is every vector in $\mathbb{R}^n$. It's clear from the picture that Error$= ||A\vec{x} - \vec{b}||^2 = ||A\vec{x} - \vec{p}||^2 + ||e||^2$. If we say that there exists a vector $\hat{x}$ such that $\vec{p} = A\hat{x}$ then when $\vec{x} = \hat{x}$, $||A\hat{x} - \vec{b}||^2 = ||e||^2$, (notice $A\hat{x} - \vec{p} = \vec{0}$) is the "least square error." Thus, this vector $\hat{x}$ is the least squares solution to $A\vec{x} = \vec{b}$, i.e. it is the vector which minimizes the distance between vectors in the column space (i.e. $A\vec{x}$) and the vector $\vec{b}$.

Recall that you can find $\hat{x}$ using the formula for projection $\vec{p} = A\hat{x}$ of $\vec{b}$ onto col(A):
$\vec{p} = A(A^TA)^{-1}A^T\vec{b} \Rightarrow \hat{x} = (A^TA)^{-1}A^T\vec{b}$

DEFINITION

If A is a matrix with linearly independent columns, then the **pseudoinverse of** A is the matrix A^+ defined by $A^+ = (A^TA)^{-1}A^T$ Note that when A is an $m \times n$ matrix, A^+ is an $n \times m$ matrix.

Then the least squares solution to $A\vec{x} = \vec{b}$ is $\vec{x} = \hat{x} = A^+\vec{b}$.

Exercise

Find the least square error and the equation of the line of best fit for the points $(0, 6)$, $(1, 0)$ and $(2, 0)$.

Using Calculus

Another way to think about this problem is as an optmization problem. That is, find c and d in the equation of the line $y = dx + c$ which minimizes the square error between the line and the given data.

$$\begin{aligned} E &= (dx_1 + c - y_1)^2 + (dx_2 + c - y_2)^2 + (dx_3 + c - y_3)^2 \\ &= (d \cdot 0 + c - 6)^2 + (d \cdot 1 + c - 0)^2 + (d \cdot 2 + c - 0)^2 \end{aligned}$$

So, it's clear that the square error function E is a function of c and d, i.e. $E(d, c)$. To find the minimum value of a function of multiple variables one needs to discover where all of its partial derivatives are simultaneously equal to zero. In other words, solve $\dfrac{\partial E}{\partial c} = 0$ and $\dfrac{\partial E}{\partial d} = 0$

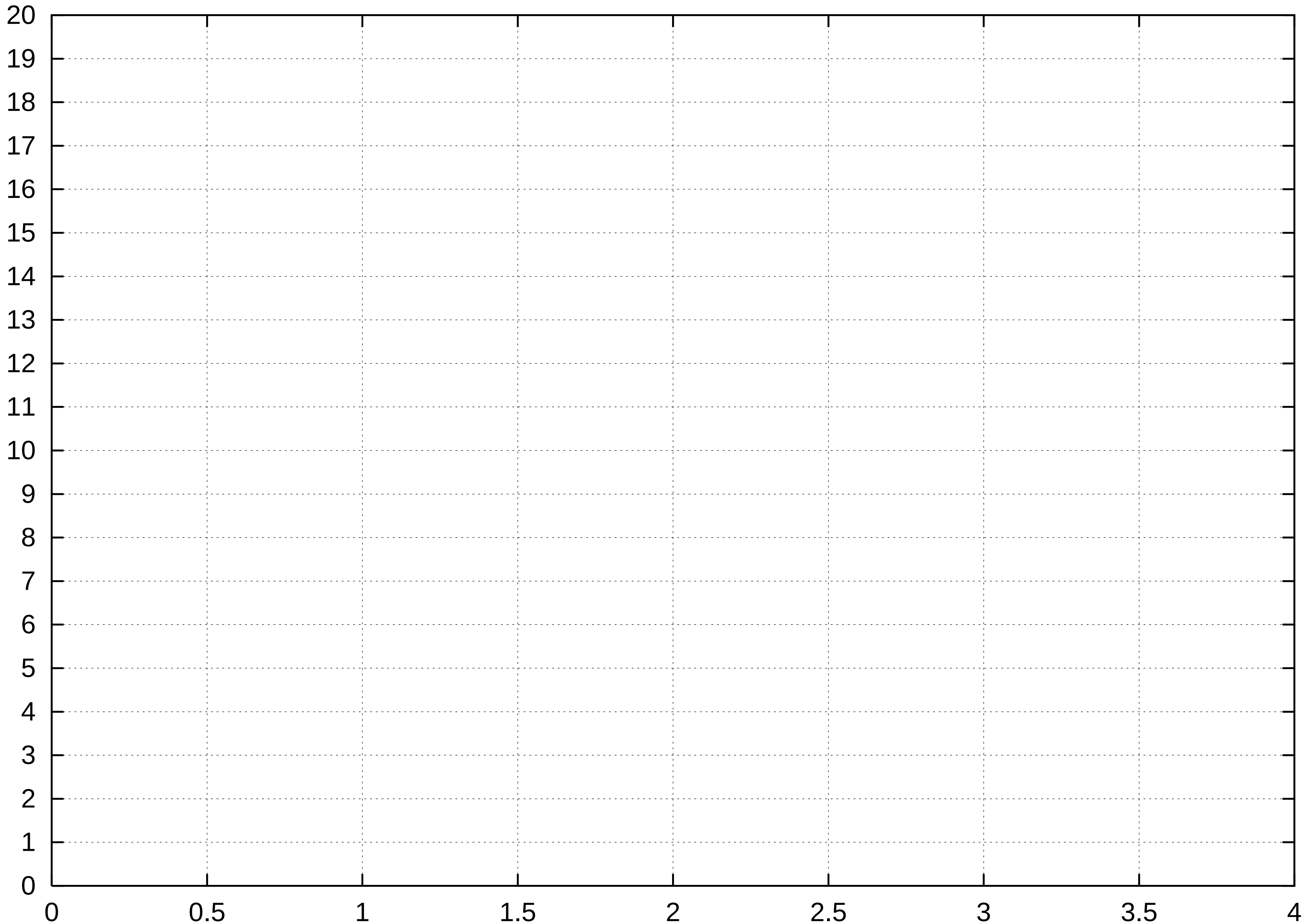
Recognize the answer? Solving these simultaneous equations is equivalent to solving the linear system: $A^T A\hat{x} = A^T\vec{b}$. These equation are known as the **normal equations**.

$\boxed{\text{EXERCISE}}$

Strang, page 216, No. 9. Find the equation of the parabola of best fit through the points $(0, 8, 8, 20)$ when $t = 0, 1, 3, 4$.

Also find the equation of the line of best fit through the same points. Which curve has the lowest least square error?

Graph the points and the curves of best fit on the axes on the next page.


20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
0.5
1
1.5
2
2.5
3
3.5
4