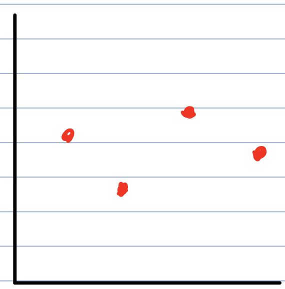
Section 6.5

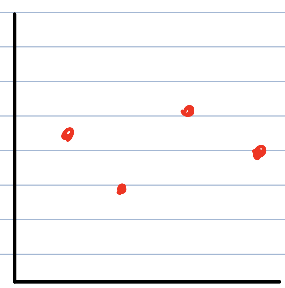
Today we will see least-square problems and how they are essentially just projections

Recall least-squares lines / linear regression



Have data / measurements. Can think of them as outputs.

Would like a function that explains these outputs and predicts other outputs for whatever inputs we supply.



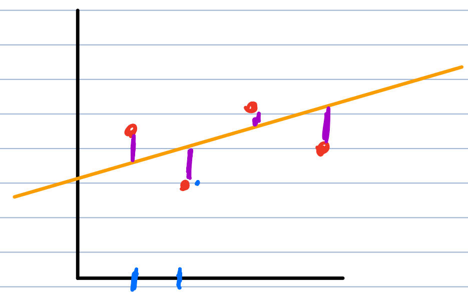
Maybe there is some crazy function that perfectly describes data.

But without more info we will never find it exactly.

What should we do? We should make a model (a guess of the underlying function).

Many ways to do this. Least-squares is one way.

Least squares approach:



We will make very simple guesses about underlying function.

Simplest functions? Lines.

Which line should we choose?

Most likely scenario: no line will fit perfectly. So, should pick a line that minimizes "error".

How do we measure "error"?

"error" = $\sum$ (distance between point and line)^2

This is where term "least squares" comes from.

That is set up. More in 6.6. For now, return to world of matrices.

Will seem very different at first.

Assume you have a fixed $m \times n$ matrix A. Then you are given some vector $\vec{b} \in \mathbb{R}^m$ an output. You want to find inputs $\vec{x} \in \mathbb{R}^n$ such that

$$A\vec{x} = \vec{b}$$

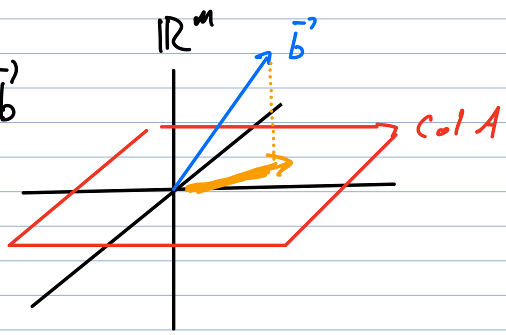
Try to solve system but it is inconsistent.

Do we give up and go home?

If can't find exact solution, would like to find approximate solution.

All the outputs that A can form form the column space, subspace of $\mathbb{R}^m$

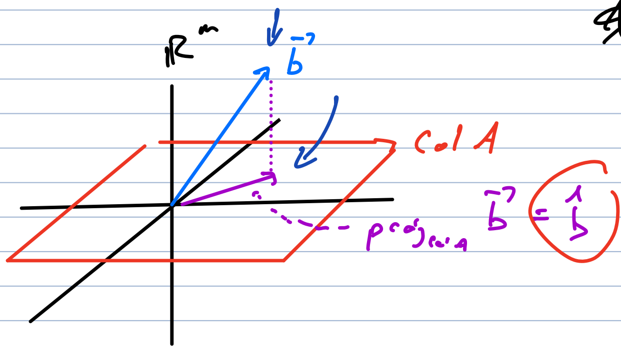
$A\vec{x} = \vec{b}$



End goal: find $\vec{x}$ that gives some $\hat{b}$ very close to $\vec{b}$.

☆ 1st question: what $\hat{b}$'s are close to $\vec{b}$?

Recall "best approximation theorem". The orthogonal projection of vector onto subspace is closest we can get to that vector while in the subspace.



(*) The projection of $\vec{b}$ onto Col A will give me output closest to $\vec{b}$.

Call this closest vector $\hat{b}$.

So $\hat{b} = \text{proj}_{\text{Col} A}\, \vec{b}$

(*) <u>2nd question</u>: $\hat{b}$ is the closest we can get to $\vec{b}$. But what inputs give us output $\hat{b}$.

$A\vec{x} = \hat{b}$ } Now try to solve this system. It <u>will</u> be consistent because $\hat{b}$ <u>is</u> in column space of A.

The $\vec{x}$ that solves $A\vec{x} = \hat{b}$ is the input that gets us as close as possible to $\vec{b}$.

Call $\vec{x}$ the least squares solution to system $A\vec{x} = \vec{b}$

<u>3rd Question</u>: How do we calculate least squares solution?

Long way: ① Row reduce A to figure out pivot columns
② Make pivot columns of A orthogonal
③ Project $\vec{b}$ onto those orthogonal vectors to get $\hat{b}$

(a) Solve $A\vec{x} = \hat{b}$

Will rarely do it like this. Too much work.

<u>Shortcut</u>:

Remember: vector minus its projection is orthogonal to subspace.

- So $\vec{b} - \hat{b}$ should be orthogonal to Col A.

- Since orthogonal to Col A, should be orthogonal to all columns of A.
  i.e $\vec{a}_i \cdot (\vec{b} - \hat{b}) = 0$

- Dot product of two vectors, $\vec{c} \cdot \vec{d}$, same as matrix multiplication $\vec{c}^{\,T} \vec{d}$ $\quad c, d \in \mathbb{R}^n$

  $$\begin{bmatrix} \quad \vec{c}^{\,T} \quad \end{bmatrix} \begin{bmatrix} \vec{d} \end{bmatrix}_{n \times 1}$$

  So $$\underbrace{\begin{bmatrix} \quad \vec{a}_i^{\,T} \quad \end{bmatrix}}_{1 \times m \text{ matrix}} \underbrace{\begin{bmatrix} \vec{b} - \hat{b} \end{bmatrix}}_{m \times 1 \text{ matrix}} = 0 \quad \text{for all } i.$$

- Make matrix out of rows $\vec{a}_i^{\,T}$. This is exactly matrix A^T. Thus

  $$A^T(\vec{b} - \hat{b}) = \vec{0}$$

Know that $\hat{b} = A\vec{x}$ for some $\vec{x}$'s. So rewrite $\hat{b}$ as $A\vec{x}$.

$$\star \quad A^T(\vec{b} - A\vec{x}) = \vec{0}$$

Distribute matrix A^T

$$A^T\vec{b} - A^TA\vec{x} = \vec{0} \qquad \underbrace{A^T}_{n \times n}\ \underbrace{A}_{m \times n} = n \times n \text{ matrix}$$

Rearrange

"Normal Equations"

$$\underbrace{A^TA}_{n \times n \text{ matrix}}\ \underbrace{\vec{x}}_{\text{vector unknowns}} = \underbrace{A^T\vec{b}}_{\text{vector in } \mathbb{R}^n}$$

Our $\vec{x}$ that gets us as close as possible to $\vec{b}$ will be solution to this new matrix equation $A^TA\vec{x} = A^T\vec{b}$.

Short story: $A\vec{x} = \vec{b}$ inconsistent. To find least squares solution multiply both sides by A^T then solve.

Again, solution $\vec{x}$ to $A^TA\vec{x} = A^Tb$ is the least squares solution to system $A\vec{x} = \vec{b}$

Error: $\hat{x}$ still not an exact solution.
$A\vec{x} = \hat{b}$, not $\vec{b}$. So there is error.

Least Squares Error: the distance between $\vec{b}$ and $\hat{b}$, i.e. length of vector $(\vec{b} - \hat{b})$

$$\|\vec{b} - \hat{b}\| = \sqrt{(\vec{b} - \hat{b}) \cdot (\vec{b} - \hat{b})}$$

Theorem: A an $m \times n$ matrix. TFAE:

1. $A\vec{x} = \vec{b}$ has unique least squares solution for each $\vec{b}$ in $\mathbb{R}^m$
2. Columns of A are linearly independent
3. $A^T A$ is invertible

When any one of these (and thus all of them) are true, least squares solution is given by

$$\star \quad \underline{\vec{x} = (A^T A)^{-1} A^T \vec{b}} \quad \star$$

$$\star \quad (A^T A)\vec{x} = A^T \vec{b}$$

$$\vec{x} = (A^T A)^{-1} A^T \vec{b}$$