Lecture # 21 - Math 2121

1. Gram-Schmidt process, orthogonal proj, etc
2. Least-squares solutions to linear systems

orthogonal: $u \bullet v = u^T v = 0$

orthonormal: orthogonal + unit vectors $\begin{cases} u \cdot u = 1 \\ v \cdot v = 1 \end{cases}$

1 Last time: orthonormality, projections, bases

Orthogonal vectors $u_1, u_2, \ldots, u_p$ are *orthonormal* if each vector is a unit vector.

In other words, if $u_i \bullet u_j = 0$ when $i \neq j$ and $u_i \bullet u_i = 1$ for all $i = 1, 2, \ldots, p$.

If $v_1, \ldots, v_p$ are orthogonal + nonzero then $\boxed{\frac{1}{\|v_1\|}v_1, \frac{1}{\|v_2\|}v_2, \ldots, \frac{1}{\|v_p\|}v_p \text{ are orthonormal.}}$

Theorem. Let U be an $m \times n$ matrix.

* $\boxed{\text{The columns of } U \text{ are orthonormal vectors if and only if } U^\top U = I_n.}$

If this happens then $(Ux) \bullet (Uy) = x \bullet y$ for all $x, y \in \mathbb{R}^n$.

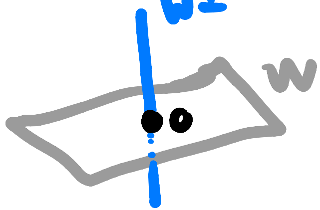
If $m = n$ then $U^\top U = I_n$ means $U^{-1} = U^\top$ and we say U is an **orthogonal matrix**.

Example. The columns of $U = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$ are orthonormal as $U^\top U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

square matrix w/ orthonormal columns

$U^{-1} = U^T \Longleftrightarrow$

picture



here: $n=3$, $\dim W = 2$, $\dim W^\perp = 1$

Let $W \subseteq \mathbb{R}^n$ be any subspace. Recall that $W^\perp = \{w \in \mathbb{R}^n : w \bullet v = 0 \text{ for all } v \in W\}$.

We showed last time that $\boxed{W \cap W^\perp = \{0\}}$ and $\boxed{\dim W + \dim W^\perp = n.}$

Theorem. Let $y \in \mathbb{R}^n$.

Then there is a unique vector $\mathrm{proj}_W(y) \in W$ such that $y - \mathrm{proj}_W(y) \in W^\perp$.

We call $\mathrm{proj}_W(y)$ the *orthogonal projection* of y onto W.

To compute $\mathrm{proj}_W(y)$, find an orthogonal basis $u_1, u_2, \ldots, u_p$ for W; then

$$\mathrm{proj}_W(y) = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \frac{y \bullet u_2}{u_2 \bullet u_2} u_2 + \cdots + \frac{y \bullet u_p}{u_p \bullet u_p} u_p.$$

easy to compute $\mathrm{proj}_W(y)$ (if we have an orthog. basis for W)

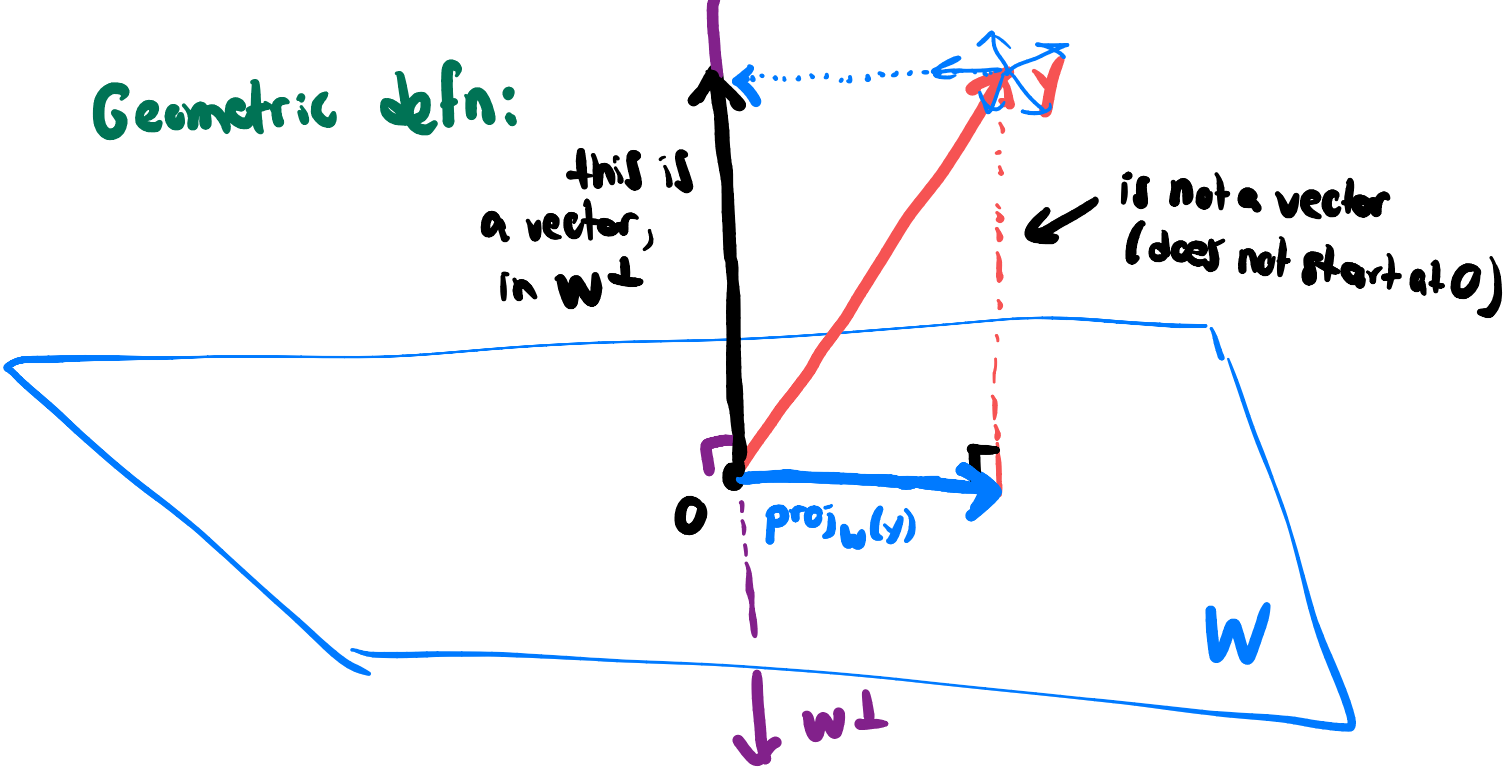
($u_i \cdot u_j = 0$, $i \neq j$)

① $\mathrm{proj}_W(y)$ is in W

② the difference $y - \mathrm{proj}_W(y) \in W^\perp$

Note: if $y \in W$ then $\mathrm{proj}_W(y) = y$, if $y \in W^\perp$ then $\mathrm{proj}_W(y) = 0$

Geometric defn:



$$\boxed{\text{proj}_W(y) = \text{“shadow of } y \text{ cast on } W\text{”}}$$

black vector $= y - \text{proj}_W(y) \in W^\perp$

Properties of orthogonal projections

* We have $\mathrm{proj}_W(y) = y$ if and only if $y \in W$, and $\mathrm{proj}_W(y) = 0$ if and only if $y \in W^\perp$.

Recall $\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$. (usual def of length)

$\boxed{\text{Then } \|y - \mathrm{proj}_W(y)\| < \|y - v\| \text{ for all } v \in W \text{ with } v \neq \mathrm{proj}_W(y).}$

in words: $\mathrm{proj}_W(y)$ is the vector in W with minimum distance to y

(the vector in W that is closest to y)

$$\underset{(v \in W)}{\mathrm{dist}(y,v)} \overset{\text{def}}{=} \|y-v\| > \|y - \mathrm{proj}_W(y)\| = \mathrm{dist}(y, \mathrm{proj}_W(y))$$

if $v \neq \mathrm{proj}_W(y)$

a sequence of formulas that converts a basis for W to an orthogonal basis

Gram-Schmidt process. Let $W \subseteq \mathbb{R}^n$ with basis $x_1, x_2, \ldots, x_p$.

Define $v_1, v_2, \ldots, v_p \in W$ inductively by the following formulas:

1

$$v_1 = x_1.$$

2

$$v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1.$$

3

$$v_3 = x_3 - \frac{x_3 \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2.$$

$$\vdots$$

p

$$v_p = x_p - \frac{x_p \bullet v_1}{v_1 \bullet v_1} v_1 - \frac{x_p \bullet v_2}{v_2 \bullet v_2} v_2 - \frac{x_p \bullet v_3}{v_3 \bullet v_3} v_3 - \cdots - \frac{x_p \bullet v_{p-1}}{v_{p-1} \bullet v_{p-1}} v_{p-1}.$$

For each i, the vectors $v_1, v_2, \ldots, v_i$ are an orthogonal basis for the subspace

$$\mathbb{R}\text{-span}\{x_1, x_2, \ldots, x_i\} = \mathbb{R}\text{-span}\{v_1, v_2, \ldots, v_i\}. = W_i$$

The full list of vectors $v_1, v_2, \ldots, v_p$ is an orthogonal basis for W.

why this works: $v_{i+1} \overset{\text{def}}{=} \underbrace{x_{i+1} - \text{proj}_{W_i}(x_{i+1})}_{\text{will be in } W_i^{\perp}\text{, hence will be orthogonal to } v_1, \ldots, v_i}$ where

$W \subseteq \mathbb{R}^4$

Example. Let $\boxed{W = \text{Nul}\left(\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}\right)} = \{w \in \mathbb{R}^4 : w_1 + w_2 + w_3 + w_4 = 0\}$.

A basis for W is given by $x_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$, $x_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$, $x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$.

(not orthogonal)

$x_1 \cdot x_2 = -1$

$x_1 \cdot x_3 = 0$

$x_2 \cdot x_3 = -1$

GS :

1. To find an orthogonal basis, we let $\boxed{v_1 = x_1.}$

2. $$\boxed{v_2 = x_2 - \frac{x_2 \bullet v_1}{v_1 \bullet v_1} v_1} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}}_{x_2} + \frac{1}{2} \underbrace{\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}}_{v_1} = \begin{bmatrix} 1/2 \\ 1/2 \\ -1 \\ 0 \end{bmatrix}.$$

$$-\frac{x_2 \cdot v_1}{v_1 \cdot v_1} = -\frac{\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}} = \frac{+1}{1+1} = +\frac{1}{2}$$

Check:

$$v_1 \cdot v_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1/2 \\ 1/2 \\ -1 \\ 0 \end{bmatrix} = 0 \checkmark$$

$x_3 \cdot v_1 = x_3 \cdot x_1 = 0$

$-\dfrac{x_3 \bullet v_2}{v_2 \bullet v_2}$

x_3

3

$$v_3 = x_3 - \cancel{\underbrace{\frac{x_3 \bullet v_1}{v_1 \bullet v_1}}_{=0} v_1} - \frac{x_3 \bullet v_2}{v_2 \bullet v_2} v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} + \frac{2}{3} \begin{bmatrix} 1/2 \\ 1/2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \\ -1 \end{bmatrix}.$$

v_2

Thus $\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1/2 \\ 1/2 \\ -1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \\ -1 \end{bmatrix}$ are an orthogonal basis for W.

Check!

* The rescaled vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix}$ are also an orthogonal basis for W.

more equations than variables

2 Least-squares problems

Many linear systems $Ax = b$ that arise in applications are $\boxed{\textit{overdetermined}}$ (meaning they have more equations than variables, so the matrix A has more rows than columns) and often inconsistent (meaning they have no exact solution $x \in \mathbb{R}^n$).

There may be no input vector $x \in \mathbb{R}^n$ such that $Ax = b$. When no exact solution is available, the next best thing to provide is an input vector $x \in \mathbb{R}^n$ such that Ax is as "close" to the vector $b \in \mathbb{R}^m$ as possible.

goal: find x such that Ax is close to b

least-squares solution = (a certain kind of approximate solution to a linear system)

What is a good definition of an approximate solution to $Ax = b$?

(imprecise)

→ a vector x that makes Ax as close to b as possible

→ more precisely, a vector x that makes the distance $\|Ax - b\|$ as small as possible

(from Ax to b)

(precise)

There is more than one way to quantify how close two vectors are to each other. One common method is to use the distance function we have already seen: define the *distance* between vectors $u, v \in \mathbb{R}^n$ to be

$$\|u - v\| = \sqrt{(u-v) \bullet (u-v)} = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}.$$

Two vectors are close if their distance in this sense is small. The distance function $\|\cdot\|$ is called the *Euclidean distance* or *L^2-distance*. In two and three dimensions, this distance is the usual way that we measure distance between points in space.

*

Definition. If A is an $m \times n$ matrix and $b \in \mathbb{R}^m$, then a *least-squares solution* to the linear system $\boxed{Ax = b}$ is a vector $s \in \mathbb{R}^n$ with $\|b - As\| \leq \|b - Ax\|$ for all $x \in \mathbb{R}^n$.

So a least-squares solution to $Ax = b$ is a vector $s \in \mathbb{R}^n$ that minimizes $\|b - As\|$.

A vector that minimizes $\|b - As\|$ will also minimize $\|b - As\|^2$, which is the sum of the squares of the entries in the vector $b - As$. Hence "least-squares."

$$\|v\|^2 = v_1^2 + v_2^2 + \dots + v_n^2$$

Least-squares problems (that is, problems requiring us to find a least-squares solution to some linear system) arise frequently in engineering and statistics.

Being able to solve such problems is an important application of the material covered in this course. Goal today: describe the general solution to least-squares problems.

Here are the key points:

- If $Ax = b$ is a consistent linear system then **every least-squares solution is also an exact solution.**
- There **may be more than one least-squares solution** to a given linear system $Ax = b$. (just like exact solutions)
- However, in contrast to exact solutions, there is **always at least one least-squares solution** even if $Ax = b$ is inconsistent.

The last fact is not obvious from the definition of a least-squares solution.

and this is easy to compute

Setup: $Ax = b$

Solving least-squares problems in general.

Fix an $m \times n$ matrix A and a vector $b \in \mathbb{R}^m$

A least-squares solution $s \in \mathbb{R}^n$ to $Ax = b$ is **defined to be** a vector such that $\|As - b\|$ is as small as possible.

If $s \in \mathbb{R}^n$ then we necessarily have $\boxed{As \in \operatorname{Col} A.}$

As mentioned earlier, if $v \in \operatorname{Col} A$ minimizes $\|v - b\|$ then $\boxed{v = \operatorname{proj}_{\operatorname{Col} A}(b)}$. Therefore:

Lemma. The least-squares solutions to $Ax = b$ are precisely those $s \in \mathbb{R}^n$ such that

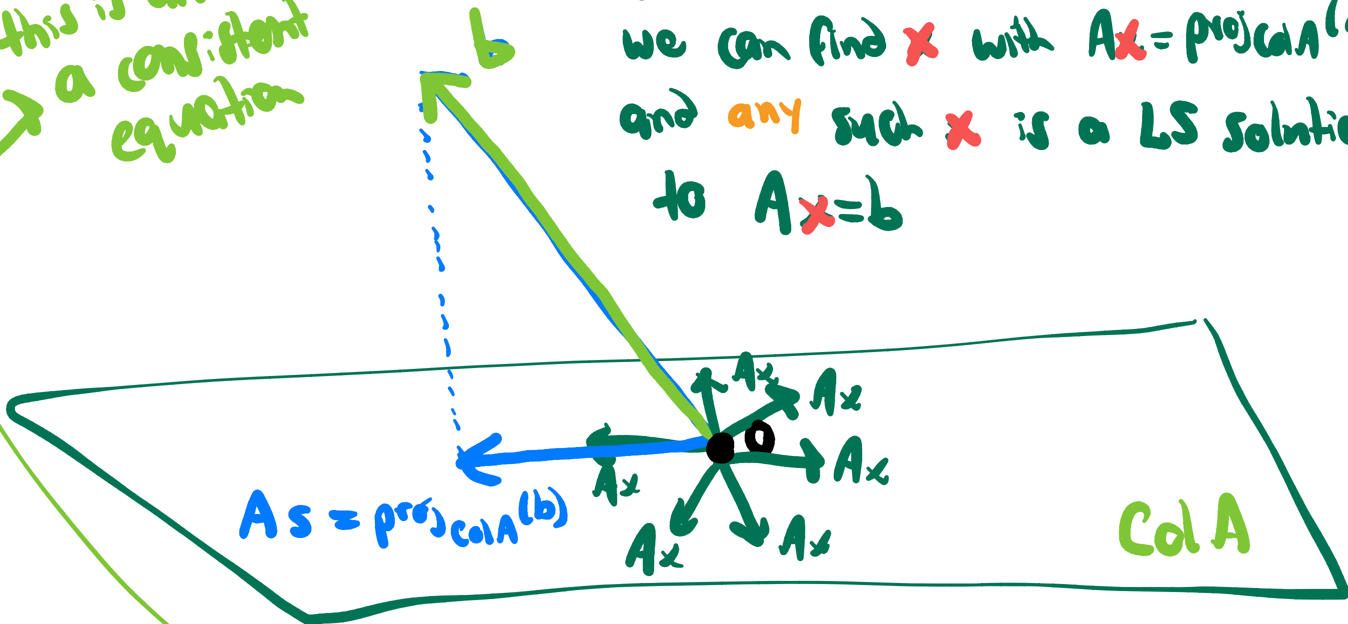
$$As = \operatorname{proj}_{\operatorname{Col} A}(b).$$

Picture of a least-squares solution to $Ax=b$ (inconsistent)

when system is

this is always a consistent equation

because $\text{proj}_{\text{Col}A}(b) \in \text{Col}A$, we can find x with $Ax = \text{proj}_{\text{Col}A}(b)$ and any such x is a LS solution to $Ax=b$



no vector Ax is equal to b so best we can do is find x such that $\boxed{Ax = \text{proj}_{\text{Col}A}(b)}$ as then Ax is as close to b as possible

Using this lemma, we can prove something even more explicit:

** important!

Theorem. The least-squares solutions to $Ax = b$ are the exact solutions to

$$A^\top A x = A^\top b.$$

← know how to solve

This new linear system is always consistent so its set of solutions is nonempty.

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

* Since $b - \widehat{b} \in (\mathrm{Col}\,A)^\perp = \mathrm{Nul}\,A^\top$, we have $A^\top(b - \widehat{b}) = 0$ and $A^\top \widehat{b} = A^\top b$.

Thus, if $s \in \mathbb{R}^n$ satisfies $\boxed{As = \widehat{b} \text{ then } A^\top A s = A^\top \widehat{b} = A^\top b.}$ → in this case, if s is a LS-solution to $Ax=b$ then this shows that s is also an exact solution to $A^\top A x = A^\top b$

more rows than columns

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} x = \begin{bmatrix} \\ b \\ \\ \end{bmatrix} \rightsquigarrow \begin{bmatrix} A^T A \end{bmatrix} x = \begin{bmatrix} A^T b \end{bmatrix}$$

square

$A^T A s - A^T b = 0$

so $A^T(As-b)=0$ so

$As - b \in \operatorname{Nul} A^T$

$= ((\operatorname{Col} A)^\perp$

so $b - As \in ((\operatorname{Col} A)^\perp$

Conversely, if $s \in \mathbb{R}^n$ satisfies $A^\top A s = A^\top b$, then $A^\top(As - b) = 0$ so

$$As - b \in \operatorname{Nul} A^\top = (\operatorname{Col} A)^\perp.$$

In this case, the uniqueness of orthogonal projections implies

then

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

this means s is a LS-solution to $Ax=b$

This shows that the following are all the same:

$$\begin{cases} \text{the exact solutions to } Ax = \widehat{b} \\ \text{the least-squares solutions to } Ax = b \\ \text{exact solutions to } A^\top A x = A^\top b. \end{cases}$$

see picture

We claimed that the linear system $A^\top A x = A^\top b$ is always consistent.

Since $\widehat{b} \in \operatorname{Col} A$, by definition there must exist some $s \in \mathbb{R}^n$ such that $As = \widehat{b}$.

Therefore $A^\top A x = A^\top b$ is consistent as it holds that $A^\top A s = A^\top \widehat{b} = A^\top b$. $\square$

always true that $A^T \hat{b} = A^T b$ (on prev slide)

as $As \in \operatorname{Col} A$ and $b - As \in ((\operatorname{Col} A)^\perp$, must have

$As = \operatorname{proj}_{\operatorname{Col} A}(b)$ by def of proj approximate

exact

try $[A|b] = \left[\begin{array}{cc|c} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 1 & 11 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$

↑ pivot in last col, hence no (exact) solution

Example. Try to solve $\boxed{Ax = b \text{ for } A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \text{ and } b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}}$. We compute

$$A^\top A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \quad \text{and} \quad A^\top b = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

The least-squared solutions are the exact solutions to $A^\top Ax = A^\top b$. We have:

$$\left[\begin{array}{cc|c} 17 & 1 & 19 \\ 1 & 5 & 11 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 5 & 11 \\ 17 & 1 & 19 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 5 & 11 \\ 0 & -84 & -168 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 5 & 11 \\ 0 & 1 & 2 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

$[A^\top A \,|\, b]$

unique LS solution

$$\boxed{x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}}$$

$A\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} \neq b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$

In this case $A^\top Ax = A^\top b$ has a unique solution $s = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ which is also the unique least-squares solution to $Ax = b$. Note that $As \neq b$ as

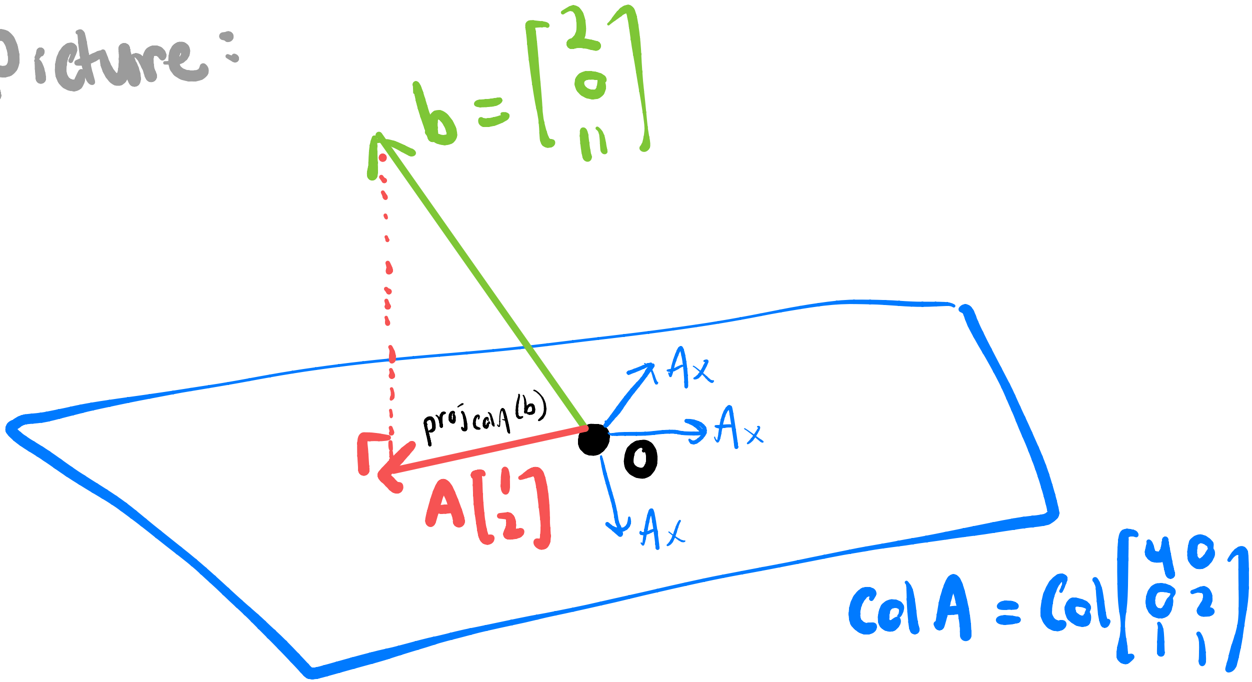
$$\|As - b\| = \left\| \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \right\| = \sqrt{4 + 16 + 64} = \sqrt{84}.$$

Geometrically, we interpret the least-squares solution as meaning that

$$As = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$$

is the point in the plane spanned by the columns of A in $\mathbb{R}^3$ that is closest to b.

Picture:



LS solution is $s = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

unique exact sol to $Ax=b$ for all $b \iff A^{-1}$ exists

A linear system $Ax = b$ has a unique solution for every $b \in \mathbb{R}^m$ if and only if the matrix A is invertible. The following theorem describes, analogously, when $Ax = b$ has a unique least-squares solution.

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

(a) $Ax = b$ has a unique least-squares solution for each $b \in \mathbb{R}^m$.

(b) The columns of A are linearly independent.

(c) $A^\top A$ is invertible.

equivalent

When these properties hold, the unique least-squares solution to $Ax = b$ is $\boxed{s = (A^\top A)^{-1} A^\top b}$, which is the unique exact solution to $A^\top Ax = A^\top b$.

Remark. The product $(A^\top A)^{-1}A^\top b$ is rarely computed for a large linear system. It is more efficient to find s by solving the system $A^\top Ax = A^\top b$ via row reduction.

$$[A^\top A \mid A^\top b]$$

$A^TAx = A^Tb$

Easy to compute a least-squares solution to $Ax = b$ if columns of A are orthogonal:

because A^TA is diagonal

Example. Suppose $A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}$ and $b = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$.

so $[A^TA \mid A^Tb]$ is already in echelon form

The orthogonal projection of b onto $\operatorname{Col} A$ is therefore

$$\operatorname{proj}_{\operatorname{Col} A}(b) = \frac{\begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix} \bullet \begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix}}{\begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix} \bullet \begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix}} \begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -6 \\ -2 \\ 1 \\ 7 \end{bmatrix}.$$

We deduce that $s = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$ is a least-squares solution to $Ax = b$ since $As = \operatorname{proj}_{\operatorname{Col} A}(b)$.