Last time: consider linearly independent vectors

$v_1, \ldots, v_k \in \mathbb{R}^n$

Gram-Schmidt orthogonalization

$$v_1 \rightsquigarrow b_1 = v_1$$

$$v_2 \rightsquigarrow b_2 = v_2 - p_{12} b_1 \quad , \text{ where } p_{ij} = \frac{v_j \cdot b_i}{b_i \cdot b_i}$$

$$\vdots$$

$$v_k \rightsquigarrow b_k = v_k - p_{1k} b_1 - \cdots - p_{k-1,k} b_{k-}$$

$\{b_1, \ldots, b_k\}$ span the same vector space as $\{v_1, \ldots, v_k\}$, but they're orthogonal

Gram-Schmidt orthonormalization: on top of the above

$$b_1 \rightsquigarrow g_1 = \frac{b_1}{\|b_1\|} \qquad\qquad b_1 = d_1 g_1$$

$$\vdots \qquad \Rightarrow \qquad \vdots \qquad , \text{ where } d_i = \|b_i\|$$

$$b_k \rightsquigarrow g_k = \frac{b_k}{\|b_k\|} \qquad\qquad b_k = d_k g_k$$

QR factorization: if $A = (v_1 \ldots v_k)$, then linearly independent

$$A = \underset{\text{orthonormal columns}}{Q}\ \underset{\text{upper triangular square matrix with positive diagonal entries}}{R}, \quad \text{where} \quad Q = (q_1 \ \cdots \ q_k)$$

$$R = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_k \end{pmatrix} \begin{pmatrix} 1 & p_{12} & \cdots & p_{1k} \\ & 1 & \cdots & p_{2k} \\ & & \ddots & \vdots \\ 0 & & 1 & p_{k-1\,k} \\ & & & 1 \end{pmatrix}$$

Today: equivalent way to get coefficients of R once you do G–S and $\{q_1, \dots, q_k\}$ is known

THM 24.1: suppose $A = (v_1 \ \cdots \ v_k)$ and $A = QR$

$Q = (q_1, \dots, q_k)$ orthonormal

Then $\boxed{r_{ij} = q_i \cdot v_j}$ are the coefficients of R.

Proof: $A = QR \quad \big|\ \cdot$ left by Q^T

$$Q^T A = \underbrace{Q^T Q}_{I_k} R$$

$$\begin{pmatrix} q_1^T \\ \hline \vdots \\ \hline q_k^T \end{pmatrix} \cdot (v_1 | \dots | v_k) = R$$

q_k

$$\begin{pmatrix} q_1^T v_1 & q_1^T v_2 & \cdots & q_1^T v_k \\ \vdots & \vdots & & \vdots \\ q_k^T v_1 & q_k^T v_2 & \cdots & q_k^T v_k \end{pmatrix} = R$$

$$\begin{pmatrix} q_1 \cdot v_1 & q_1 \cdot v_2 & \cdots & q_1 \cdot v_k \\ \vdots & \vdots & & \vdots \\ q_k \cdot v_1 & q_k \cdot v_2 & \cdots & q_k \cdot v_k \end{pmatrix} = R \implies r_{ij} = q_i \cdot v_j$$

R is upper triangular because $q_i \cdot v_j = 0 \ \forall i > j$

$q_i \perp v_j$, because G-S makes it so q_i is orthogonal to $v_1, \dots, v_j, \dots v_{i-1}$

$\square$

Ex: find QR factorization of $A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

$\begin{array}{ccc} v_1 & v_2 & v_3 \end{array}$

G-S: $v_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} \rightsquigarrow b_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} \rightsquigarrow q_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$

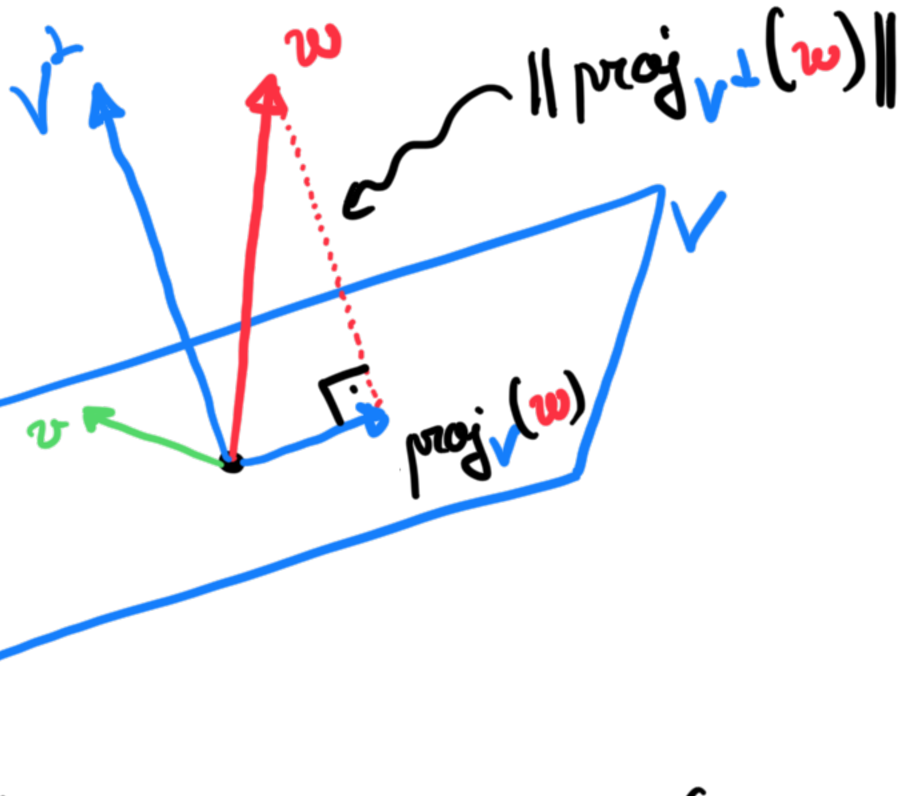
$v_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \rightsquigarrow b_2 = \begin{pmatrix} 1/3 \\ 2/3 \\ 1/3 \\ 0 \end{pmatrix} \rightsquigarrow q_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}$

$$v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix} \rightsquigarrow b_3 = \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \\ 2 \end{pmatrix} \rightsquigarrow g_3 = \frac{1}{\sqrt{18}} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 4 \end{pmatrix}$$

$$Q = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{18}} \\ 0 & 0 & \frac{4}{\sqrt{18}} \end{pmatrix}$$

$$R = \begin{pmatrix} g_1 \cdot v_1 & g_1 \cdot v_2 & g_1 \cdot v_3 \\ 0 & g_2 \cdot v_2 & g_2 \cdot v_3 \\ 0 & 0 & g_3 \cdot v_3 \end{pmatrix} = \begin{pmatrix} \sqrt{3} & \frac{2}{\sqrt{3}} & 0 \\ 0 & \sqrt{\frac{2}{3}} & \frac{3}{\sqrt{6}} \\ 0 & 0 & \frac{3}{\sqrt{2}} \end{pmatrix}$$

New topic: orthogonal projections and distances



$$\mathrm{dist}(w, V) = \min_{v \in V} \mathrm{dist}(w, v)$$

$$= \text{dist}(w, \text{proj}_V(w))$$

$$= \| w - \text{proj}_V(w) \| = \| \text{proj}_{V^\perp}(w) \|$$

you can calculate this using the projection formula.

Ex: $w = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $V = \text{span}\left\{ \underbrace{\begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}}_{v_1}, \underbrace{\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}}_{v_2} \right\}$

$\text{dist}(w, V) = ?$

$v_1 \not\perp v_2$

$$b_1 = v_1 = \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}$$

$$b_2 = v_2 - \text{proj}_{b_1}(v_2) = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} - \frac{b_1 \cdot v_2}{b_1 \cdot b_1} b_1$$

$$= \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} - \frac{-3}{27} \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}$$

$$= \frac{1}{\ } \begin{pmatrix} 19 \\ \ \end{pmatrix}$$

$$= \frac{1}{9}\begin{pmatrix} \\ -4 \\ 1 \end{pmatrix}$$

$$\mathrm{proj}_V(w) = \frac{w \cdot b_1}{b_1 \cdot b_1} b_1 + \frac{w \cdot b_2}{b_2 \cdot b_2} b_2$$

$$= \frac{1}{27}\begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix} + \frac{\frac{19}{\cancel{9}}}{\frac{378}{\cancel{81}}} \frac{1}{\cancel{9}}\begin{pmatrix} 19 \\ -4 \\ 1 \end{pmatrix}$$

$$= \frac{1}{126}\begin{pmatrix} 125 \\ -2 \\ 11 \end{pmatrix}$$

$$\mathrm{dist}(w, V) = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{126}\begin{pmatrix} 125 \\ -2 \\ 11 \end{pmatrix} \right\| = \frac{1}{126} \left\| \begin{pmatrix} 1 \\ -2 \\ 11 \end{pmatrix} \right\| = \frac{1}{\sqrt{126}}$$

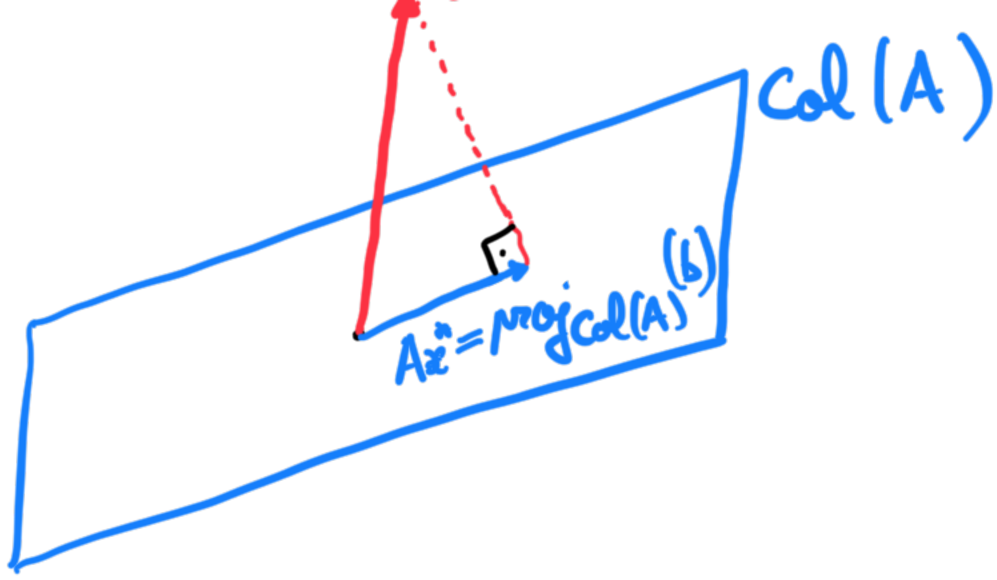
New topic: least squares approximation

$Ax = b$ $\rightsquigarrow$ if $b \in \mathrm{Col}(A)$, $\exists$ solution!

But what if $b \notin \mathrm{Col}(A)$, then $\nexists$ x s.t. $Ax = b$

Let's instead search for an approximation $Ax \sim b$

b



error

Best approximation: that $x = x^*$ such that $\|Ax - b\|$ is minimal

$$\|Ax - b\| = \text{dist}(b, Col(A))$$

minimal when $Ax^* = \text{proj}_{col(A)}(b)$

Goal of least squares: find x^* such that $Ax^* = \text{proj}_{col(A)}(b)$

DEF 24.2: a least squares solution to $Ax = b$ is x^* such that $Ax^* = \text{proj}_{Col(A)}(b)$

minimizes error

$$\left(\|Ax^* - b\| \leq \|Ax - b\| \text{ for all } x \right)$$

A least squares solution always exists because

A least squares solution always exists:

$\mathrm{proj}_{\mathrm{Col}(A)}(b) \in \mathrm{Col}(A)$, so $Ax^* = \mathrm{proj}_{\mathrm{Col}(A)}(b)$ has sols.

THM 24.3: x^* is a least squares solution to $Ax = b$ $\iff$ $A^TAx^* = A^Tb$

rectangular inconsistent — square consistent

Proof: x^* is a least squares solution to $Ax = b$

$$\Updownarrow$$

$$Ax^* = \mathrm{proj}_{\mathrm{Col}(A)}(b)$$

$$\Updownarrow$$

$$Ax^* - b \in \mathrm{Col}(A)^{\perp} = \mathrm{Ker}(A^T)$$

$$\Updownarrow$$

$$A^T(Ax^* - b) = 0$$

$\square$

Ex: find a least squares solution to

$$\underbrace{\begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}}_{A} x = \underbrace{\begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}}_{b}$$

By Thm 24.3, $A^T A\, x^* = A^T b = \begin{pmatrix} 6 \\ -2 \end{pmatrix}$

$$A^T A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix}$$

$$\rightsquigarrow \begin{pmatrix} 6 & 0 \\ 0 & 5 \end{pmatrix} x^* = \begin{pmatrix} 6 \\ -2 \end{pmatrix}$$

$$\rightsquigarrow x^* = \begin{pmatrix} 1 \\ -2/5 \end{pmatrix}$$

Explicit formula for x^*: assumes columns of A are linearly independent (Ax = linear combi of columns of A)

$\Downarrow$

$A = QR$ (Q orthonormal columns)

THM 24.4: for $A = QR$ as above, a least squares solution to $Ax = b$ is

$$\boxed{x^* = R^{-1} Q^T b}$$

Proof: $A x^* = A R^{-1} Q^T b = Q \underbrace{R R^{-1}}_{} Q^T b$

I_k

$$= QQ^T b$$

Thm 23.1

$$= \mathrm{proj}_{Col(Q)}(b)$$

By G-S, $Col(A) = Col(Q)$

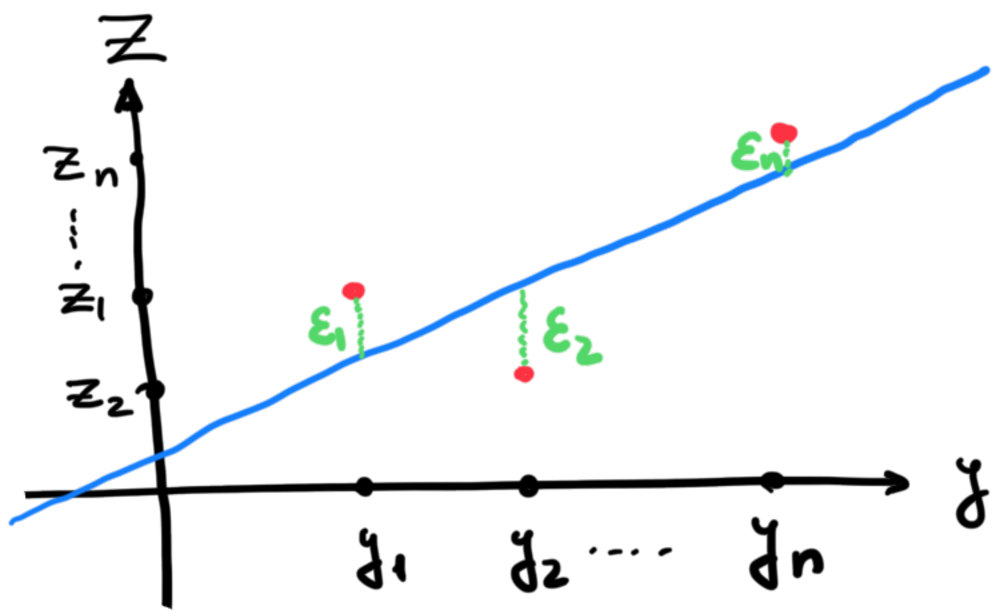
$$= \mathrm{proj}_{Col(A)}(b) \quad \square$$

THM 24.5: a least squares solution x^* always exists, but it is unique $\iff$ A^TA is invertible $\iff$ A has ind. columns

Why is this called least squares?

It's because it comes from approximation

Data points •



Find a line s.t. Error $\varepsilon = \sqrt{\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2}$ is minimal

$z = ay + b$, for $a, b \in \mathbb{R}$

$$\varepsilon_1 = z_1 - ay_1 - b$$
$$\varepsilon_2 = z_2 - ay_2 - b$$
$$\vdots$$
$$\varepsilon_n = z_n - ay_n - b$$

$$\varepsilon = \sqrt{(z_1 - ay_1 - b)^2 + \cdots + (z_n - ay_n - b)^2}$$

$$= \left\| \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} - \begin{pmatrix} ay_1 + b \\ \vdots \\ ay_n + b \end{pmatrix} \right\|$$

$$= \Bigg\| \underbrace{\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}}_{b} - \underbrace{\begin{pmatrix} y_1 & 1 \\ \vdots & \vdots \\ y_n & 1 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} a \\ b \end{pmatrix}}_{x^*} \Bigg\|$$

minimize this length

Know minimum is achieved when x^* is a least squares solution, i.e. $A^T A x^* = A^T b$

$$\Updownarrow$$

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

invertible because $y_1, \dots, y_n$ are distinct

Ex: $(y_1, z_1) = (1,0)$

$(y_2, z_2) = (2,3)$

$(y_3, z_3) = (3,3)$

$(y_4, z_4) = (4,2)$

$$\rightsquigarrow A = \begin{pmatrix} y_1 & 1 \\ \vdots & \vdots \\ y_4 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}$$

$$b = \begin{pmatrix} z_1 \\ \vdots \\ z_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 3 \\ 2 \end{pmatrix}$$

$$A^T A = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}, \quad (A^T A)^{-1} = \frac{1}{20}\begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}$$

$$A^T b = \begin{pmatrix} 23 \\ 8 \end{pmatrix}$$

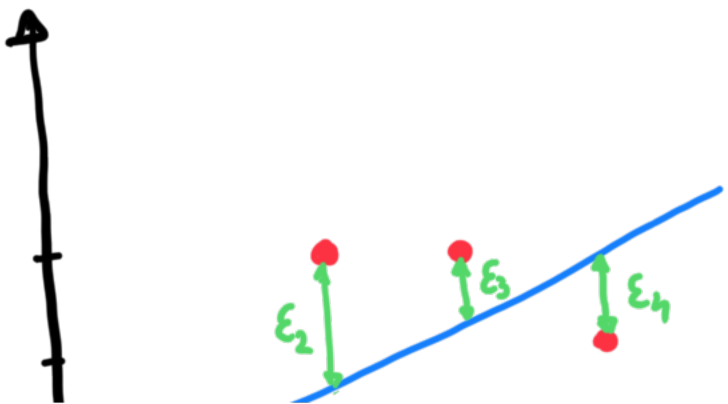
$$\longrightarrow x^* = \frac{1}{20}\begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}\begin{pmatrix} 23 \\ 8 \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 12 \\ 10 \end{pmatrix}$$

$$= \begin{pmatrix} 3/5 \\ 1/2 \end{pmatrix}$$

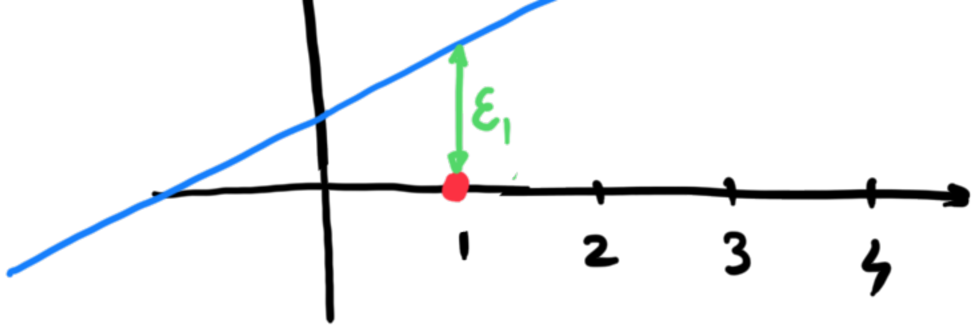
Best linear fit for our 4 data points is the line

↳ smallest Error ε

$$\boxed{z = \frac{3}{5} y + \frac{1}{2}}$$



$\sqrt{\varepsilon_1^2 + \varepsilon_2^2 + \varepsilon_3^2 + \varepsilon_4^2}$ is minimal for this



choice of time