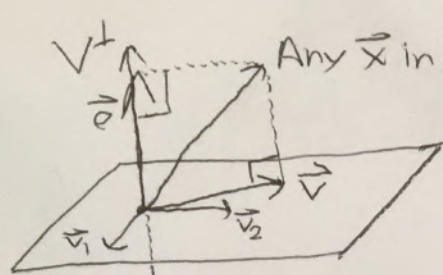


Last time: Orthogonal projection onto a subspace V of \mathbb{R}^n : (109)



Subspace V ,
Basis $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$

Any \vec{x} in $\mathbb{R}^n = \vec{v} + \vec{e}$
 \vec{v} ← "best approximation" vector in V
 \vec{e} ← "Error" vector in V^\perp

Projection matrix: $\vec{v} = P \vec{x}$

$P = P^T$ and $P^2 = P$

How to find P : Write $A = \begin{bmatrix} | & | & \dots & | \\ \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_m \\ | & | & \dots & | \end{bmatrix}$ ← $n \times m$ matrix, not usually invertible.

Then $P = A(A^T A)^{-1} A^T$

Example: $V =$ plane in \mathbb{R}^3 with equation $x + 2y + 3z = 0$

① Find basis: $V =$ all $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ such that $x = -2y - 3z$

$$= \text{all } \begin{bmatrix} -2y - 3z \\ y \\ z \end{bmatrix} = \text{all } y \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}$$

linearly independent spanning set, so basis for V

② $A = \begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$ Notice A is not invertible, but $A^T A$ is!

③ $A^T A = \begin{bmatrix} -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ 6 & 10 \end{bmatrix}$, $(A^T A)^{-1} = \frac{1}{50-36} \begin{bmatrix} 10 & -6 \\ -6 & 5 \end{bmatrix}$
 $\frac{1}{14}$

$$P = A(A^T A)^{-1} A^T = \begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{14} \begin{bmatrix} 10 & -6 \\ -6 & 5 \end{bmatrix} \begin{bmatrix} -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{14} \begin{bmatrix} -2 & -3 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -2 & 10 & -6 \\ -3 & -6 & 5 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 13 & -2 & -3 \\ -2 & 10 & -6 \\ -3 & -6 & 5 \end{bmatrix}$$

Should look familiar?

④ Let's project $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ onto V : $\vec{v} = P \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{14} \begin{bmatrix} 13 \\ -2 \\ -3 \end{bmatrix}$

Here, V is the orthogonal complement of the line from ~~last time~~ ^{example} last time

So our new P is actually $I - (\text{old } P)$

for the line spanned by $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

Question: What is the rank of our new P ?

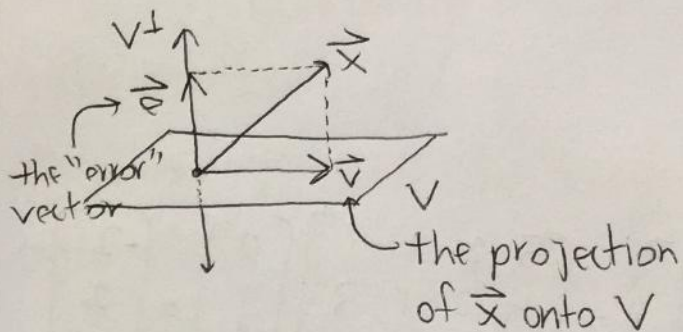
$$\text{rank}(P) = \dim \text{ of } C(P) \leftarrow \text{all vectors like } P\vec{x}$$

= all vectors in V

$$= \dim V = 2$$

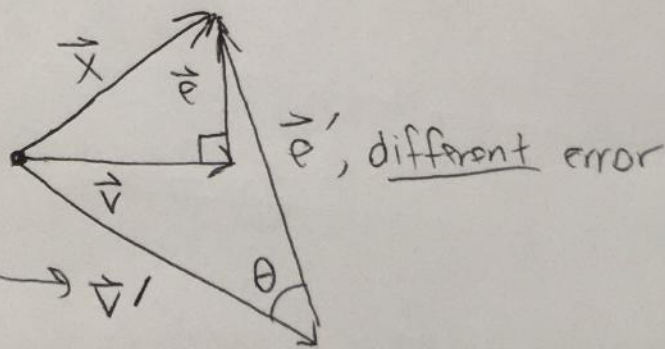
In general: rank of a projection matrix = dimension of the subspace you are projecting onto.

Section 4.3 Least Squares Approximation



Now we want to emphasize: \vec{v} is the vector in V that is closest to \vec{x} . It gives the smallest error, $\|\vec{e}\|$.

Why does \vec{v} minimize the error? Consider any different approximation.

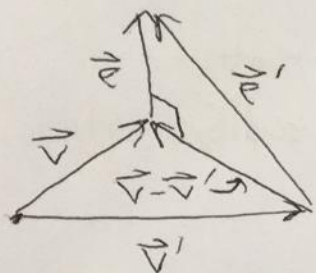


In the picture, it looks like $\|\vec{e}'\| > \|\vec{e}\|$ (worse approximation) (11)
but let's prove it:

First: $\vec{x} = \vec{v}' + \vec{e}'$, so $\|\vec{e}'\| = \|\vec{x} - \vec{v}'\| = \|\vec{e} + (\vec{v} - \vec{v}')\|$
 $\vec{x} = \vec{v} + \vec{e}$ also

Next: $\vec{v} - \vec{v}'$ is still a vector in V , and $\vec{e} \perp V$, so $\vec{e} \perp (\vec{v} - \vec{v}')$.

Pythagorean Theorem: $\|\vec{e}'\|^2 = \|\vec{e} + (\vec{v} - \vec{v}')\|^2$
 $= \|\vec{e}\|^2 + \|\vec{v} - \vec{v}'\|^2$



This is a squared real number, has to be ≥ 0 .

So $\|\vec{e}'\|$ is smallest when $\|\vec{v} - \vec{v}'\|$ is as small as possible. This happens when $\|\vec{v} - \vec{v}'\| = 0$, or $\vec{v} = \vec{v}'$ (and then $\vec{e} = \vec{e}'$).

so smallest error $\|\vec{e}\| \iff \vec{e} \perp \vec{v}$ (orthogonal projection)

Now: What is orthogonal projection good for?

Linear regression (and more general curve fitting)

Example Can we find a line that goes through the data points $(-1, 3)$, $(0, 0)$, $(1, 1)$, and $(2, -3)$?

Intuition: Can always find a line through two points, but for four points? Probably not.

Solution: Turn this into a linear algebra problem: We want to find a line $y = C + Dx$ that contains the points, so:

$$\begin{aligned} C + (-1)D &= 3 \\ C + (0)D &= 0 \\ C + (1)D &= 1 \\ C + (2)D &= -3 \end{aligned} \longrightarrow \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 1 \\ -3 \end{bmatrix}$$

A

What we are really asking is: Is $\vec{b} = \begin{bmatrix} 3 \\ 0 \\ 1 \\ -3 \end{bmatrix}$ in the column space $C(A)$?

check = $\begin{bmatrix} 1 & -1 & | & 3 \\ 1 & 0 & | & 0 \\ 1 & 1 & | & 1 \\ 1 & 2 & | & -3 \end{bmatrix}$ $\xrightarrow[\text{Row 4 - Row 1}]{\text{Row 2 - Row 1, Row 3 - Row 1}}$ $\begin{bmatrix} 1 & -1 & | & 3 \\ 0 & 1 & | & -3 \\ 0 & 2 & | & -2 \\ 0 & 3 & | & -6 \end{bmatrix}$ $\xrightarrow[\text{Row 4 - 3 Row 2}]{\text{Row 1 + Row 2, Row 3 - 2 Row 2}}$ $\begin{bmatrix} 1 & 0 & | & 0 \\ 0 & 1 & | & -3 \\ 0 & 0 & | & 4 \\ 0 & 0 & | & 3 \end{bmatrix}$

So $\vec{b} = \begin{bmatrix} 3 \\ 0 \\ 1 \\ -3 \end{bmatrix}$ is not in $C(A)$

Inconsistent equations,
no solution

→ no line goes through
the four points

What we can do instead: Find a vector $\vec{p} = C \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + D \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$

in $C(A)$ that best approximates \vec{b} . The values of C and D for \vec{p} will give us the line $y = C + Dt$ that best fits the four data points.

Of course, \vec{p} should be the orthogonal projection of \vec{b} onto $C(A)$. Let's review how to find \vec{p} :

① \vec{p} in $C(A)$ means $\vec{p} = C \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + D \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = A \underset{\substack{\uparrow \\ C(A)}}{\widehat{x}} \begin{bmatrix} C \\ D \end{bmatrix}$

② $\vec{e} \perp \vec{b} - \vec{p}$ means $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot (\vec{b} - \vec{p}) = 0$
 $\begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} \cdot (\vec{b} - \vec{p}) = 0$

\vec{e} is the error, $\vec{b} - \vec{p}$

So $\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix}}_{A^T} (\vec{b} - A\hat{x}) = \vec{0} \rightarrow \boxed{(A^T A)\hat{x} = A^T \vec{b}}$

called the "normal equations" for $\hat{x} = \begin{bmatrix} C \\ D \end{bmatrix}$ (solution is the parameters C and D of the best fit line)

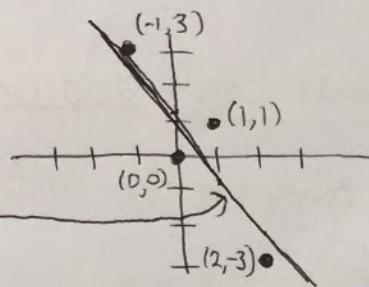
Specifically: $\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_{A} \hat{x} = \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} 3 \\ 0 \\ +1 \\ -3 \end{bmatrix}}_{\vec{b}}$

$\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ -8 \end{bmatrix} \leftarrow \text{"normal equations"}$

Solve: $\begin{bmatrix} 4 & 2 & | & 1 \\ 2 & 6 & | & -8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & | & -4 \\ 4 & 2 & | & 1 \end{bmatrix} \xrightarrow[\text{-4 Row 1}]{\text{Row 2}} \begin{bmatrix} 1 & 3 & | & -4 \\ 0 & -10 & | & 17 \end{bmatrix} \rightarrow$

$\begin{bmatrix} 1 & 3 & | & -4 \\ 0 & 1 & | & -17/10 \end{bmatrix} \xrightarrow[\text{-3 Row 2}]{\text{Row 1}} \begin{bmatrix} 1 & 0 & | & 11/10 \\ 0 & 1 & | & -17/10 \end{bmatrix} \rightarrow \begin{matrix} C = 11/10 \\ D = -17/10 \end{matrix}$

So best fit line is $y = \frac{11}{10} - \frac{17}{10}x$



What is the "error" we have minimized here?

It is $\|\vec{e}\|$, which is $\|\vec{p} - \vec{b}\|$. Or, we are minimizing

$\|\vec{e}\|^2 = \|\vec{p} - \vec{b}\|^2 = \|A\hat{x} - \vec{b}\|^2 \leftarrow \text{"least squares"}$

$\|A\hat{x} - \vec{b}\|^2 = \left\| \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \\ 1 \\ -3 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} C - D - 3 \\ C \\ C + D - 1 \\ C + 2D + 3 \end{bmatrix} \right\|^2$

$$= (C-D-3)^2 + C^2 + (C+D-1)^2 + (C+2D+3)^2$$

squared difference between the line's y-values and the actual data points' y-values - We could also minimize this sum-of-squared-errors function using calculus:

Find C, D , such that $\frac{\partial}{\partial C} (\text{error function}) = 0$

$$\frac{\partial}{\partial D} (\text{error function}) = 0$$

Least-square lines in general: Find the best-fit line for the data points $(t_1, b_1), (t_2, b_2), \dots, (t_n, b_n)$.

Assume the t 's are all different (so think of the b 's as measurements that you make at different times).

We want a line $y = C + Dt$. At first we try to solve:

$$\begin{aligned} C + t_1 D &= b_1 \\ C + t_2 D &= b_2 \\ &\vdots \\ C + t_n D &= b_n \end{aligned} \quad \rightarrow \quad \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$\nwarrow \quad \nearrow$
 $A \quad \hat{x} \quad \vec{b}$

Usually this is impossible, no solution, so instead project \vec{b} onto $C(A)$ using the normal equations:

$$\begin{array}{l} A\hat{x} = \vec{b} \\ \text{No solution} \end{array} \xrightarrow[\text{by } A^T]{\text{Multiply}} (A^T A)\hat{x} = A^T \vec{b}$$

This is just a 2×2 matrix, so system isn't difficult to solve.

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_n \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix} = \begin{bmatrix} n & t_1 + t_2 + \dots + t_n \\ t_1 + t_2 + \dots + t_n & t_1^2 + t_2^2 + \dots + t_n^2 \end{bmatrix}$$

$$\text{Also: } A^T \vec{b} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ t_1 & t_2 & \dots & t_n \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} b_1 + b_2 + \dots + b_n \\ t_1 b_1 + t_2 b_2 + \dots + t_n b_n \end{bmatrix}$$

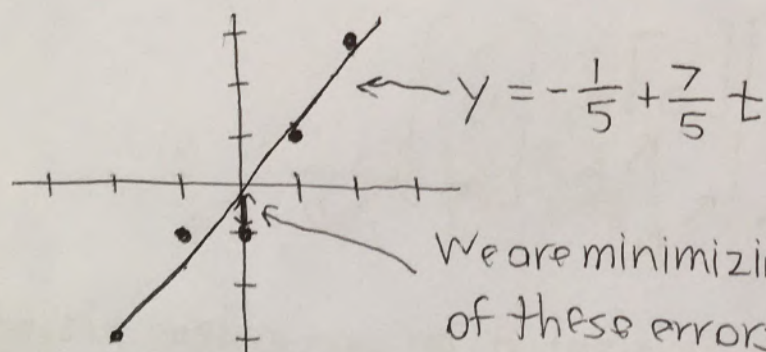
Example: Fit the data points $(-2, -3), (-1, -1), (0, -1), (1, 1), (2, 3)$.

$$\text{Here, } A^T A = \begin{bmatrix} 5 & -2-1+0+1+2 \\ -2-1+0+1+2 & 4+1+0+1+4 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}$$

Nice! This matrix is diagonal because the columns of A are \perp .

$$A^T \vec{b} = \begin{bmatrix} -3-1-1+1+3 \\ 6+1+0+1+6 \end{bmatrix} = \begin{bmatrix} -1 \\ 14 \end{bmatrix}$$

We need to solve the normal equations: $\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} -1 \\ 14 \end{bmatrix} \rightarrow \begin{matrix} C = -1/5 \\ D = 7/5 \end{matrix}$



We are minimizing the sum of the squares of these errors.

$$\begin{aligned} \text{Error} &= \|\vec{e}\| = \|A\hat{x} - \vec{b}\| = \left\| \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1/5 \\ 7/5 \end{bmatrix} - \begin{bmatrix} -3 \\ -1 \\ -1 \\ 1 \\ 3 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -3 - (-3) \\ -8/5 + 1 \\ -1/5 + 1 \\ 6/5 - 1 \\ 13/5 - 3 \end{bmatrix} \right\| \\ &= \left\| \begin{bmatrix} 0 \\ -3/5 \\ 4/5 \\ 1/5 \\ -2/5 \end{bmatrix} \right\| = \sqrt{0 + \frac{9}{25} + \frac{16}{25} + \frac{1}{25} + \frac{4}{25}} = \sqrt{\frac{30}{25}} = \sqrt{\frac{6}{5}} \quad (\text{Not too big}) \end{aligned}$$

Why do we require t 's to be different?

This guarantees that columns of A aren't multiples of each other (so they are independent).

Then $A^T A$ is invertible, and there is only one solution for \hat{x} .

We can also find best fits using other types of functions:

Try to fit our original data set to a parabola. Do

$(-1, 3), (0, 0), (1, 1), (2, -3)$ fit on $y = C + Dx + Ex^2$?

Try to solve:

$$\begin{aligned} C - D + E &= 3 \\ C + 0D + 0E &= 0 \\ C + D + E &= 1 \\ C + 2D + 4E &= -3 \end{aligned} \quad \rightsquigarrow \quad \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 1 \\ -3 \end{bmatrix}$$

$A \quad \hat{x} \quad \vec{b}$

Probably no solution still, so try to project onto $C(A)$ instead:

Same normal equations $A^T A \hat{x} = A^T \vec{b}$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 2 \\ 1 & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ -8 \\ -8 \end{bmatrix}$$

Solve:

$$\left[\begin{array}{ccc|c} 4 & 2 & 6 & 1 \\ 2 & 6 & 8 & -8 \\ 6 & 8 & 18 & -8 \end{array} \right] \xrightarrow[\substack{\text{Row 2} - \frac{1}{2}\text{Row 1} \\ \text{Row 3} - \frac{3}{2}\text{Row 1}}]{\text{Row 2} - \frac{1}{2}\text{Row 1}} \left[\begin{array}{ccc|c} 4 & 2 & 6 & 1 \\ 0 & 5 & 5 & -17/2 \\ 0 & 5 & 9 & -19/2 \end{array} \right] \longrightarrow \left[\begin{array}{ccc|c} 4 & 2 & 6 & 1 \\ 0 & 5 & 5 & -17/2 \\ 0 & 0 & 4 & -1 \end{array} \right]$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 27/20 \\ 0 & 1 & 0 & -29/20 \\ 0 & 0 & 1 & -1/4 \end{array} \right] \longleftarrow \left[\begin{array}{ccc|c} 1 & 1/2 & 0 & 5/8 \\ 0 & 1 & 0 & -29/20 \\ 0 & 0 & 1 & -1/4 \end{array} \right] \longleftarrow \left[\begin{array}{ccc|c} 1 & 1/2 & 3/2 & 1/4 \\ 0 & 1 & 1 & -17/10 \\ 0 & 0 & 1 & -1/4 \end{array} \right]$$

Best fit parabola: $y = \frac{27}{20} - \frac{29}{20}t - \frac{1}{4}t^2$

Compare with line: $y = \frac{11}{10} - \frac{17}{10}t$

Parabola should have less error since $C(A)$ bigger \rightarrow easier to get a good approximation.