Theorem 4.2.9 (Best Approximation)**.** Let V be a subspace of $\mathbb{R}^n$, $\mathbf{b}$ be any vector in $\mathbb{R}^n$, and $\mathbf{p} = \cancel{\text{proj}_V(\mathbf{b})}$ proj of b onto V. Then $\mathbf{p}$ is the closest point in V to $\mathbf{b}$, in the sense that

$$\|\mathbf{b}-\mathbf{p}\| < \|\mathbf{b}-\mathbf{v}\|$$

for all other $\mathbf{v}$ in V.

Example 4.2.10. Let $\boxed{W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}}$ and find the closest point in W to $\mathbf{y}$ where

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \qquad \text{and } \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

$$[u_1\ u_2]\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \left[\begin{array}{cc|c} 2 & -2 & 1 \\ 5 & 1 & 2 \\ -1 & 1 & 3 \end{array}\right]$$

(see `https://www.geogebra.org/m/hybndwvh`)

Closest point is $p = A\vec{x}$ where $A = [u_1\ u_2]$ and $\hat{x}$ is soln

$$A^TA\hat{x} = A^T\vec{b}$$

① find A^TA and A^Tb

$$A = \begin{bmatrix} 2 & -2 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} \qquad A^T = \begin{bmatrix} 2 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 2 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -2 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & -4 \\ -4 & 6 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 2 & 1 & -1 \\ -2 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

② Solve $\boxed{A^TA}\hat{x} = \boxed{A^Tb}$

$$\begin{bmatrix} 6 & -4 \\ -4 & 6 \end{bmatrix}\hat{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 6 & -4 & 1 \\ -4 & 6 & 3 \end{array}\right] \sim \quad \begin{array}{l} \hat{x}_1 = 0.9 \\ \hat{x}_2 = 1.1 \end{array}$$

③ $$\vec{p} = A\hat{x} = \begin{bmatrix} 2 & -2 \\ 2 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 0.9 \\ 1.1 \end{bmatrix} = \begin{bmatrix} -0.4 \\ 2 \\ 0.2 \end{bmatrix}$$

Example 4.2.11. The distance from a point $\mathbf{u}$ in $\mathbb{R}^n$ to a subspace V is defined as the distance from $\mathbf{u}$ to the nearest point in V. Find the distance from $\mathbf{u}$ to $V = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ where

$$\mathbf{u} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \qquad \text{and } \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

(see `https://www.geogebra.org/m/p9q2n95j`)

① $A^TA = \begin{bmatrix} 30 & 0 \\ 0 & 6 \end{bmatrix}$

$A^Tb = \begin{bmatrix} 15 \\ -21 \end{bmatrix}$

$$\begin{bmatrix} 30 & 0 \\ 0 & 6 \end{bmatrix} \hat{x} \begin{bmatrix} 15 \\ -21 \end{bmatrix}$$

② $\left[\begin{array}{cc|c} 30 & 0 & 15 \\ 0 & 6 & -21 \end{array}\right] \rightsquigarrow \begin{aligned} \hat{x}_1 &= 1/2 \\ \hat{x}_2 &= \frac{-21}{6} = \frac{-7}{2} \end{aligned}$

③ $\vec{p} = A\hat{x} = \begin{bmatrix} 5 & 1 \\ -2 & 2 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1/2 \\ -7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$

distance from $\vec{u}$ to V is $\|\vec{u} - \vec{p}\|$

$$\left\| \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix} \right\|$$

$$\sqrt{0^2 + 3^2 + 6^2}$$

$$= \sqrt{45}$$

4.3 Least squares

4.3. Key Ideas

- The least squares solution $\widehat{\mathbf{x}}$ minimizes $E = \|A\mathbf{x} - \mathbf{b}\|^2$. This is the sum of squares of the errors in the m equations ($m > n$).
- The best $\widehat{\mathbf{x}}$ comes from the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$.
- To fit m points by a line $b = C + Dt$, the normal equations give C and D.
- The heights of the best line are $\mathbf{p} = (p_1, \ldots, p_m)$. The vertical distances to the data points are the errors $\mathbf{e} = (e_1, \ldots, e_m)$.
- If we try to fit m points by a combination of $n < m$ functions, the m equations $A\mathbf{x} = \mathbf{b}$ are generally unsolvable. The n equations $A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}$ give the least squares solution – the combination with the smallest mean square error.

Example 4.3.1. Find the closest line $y = a + bx$ to the points $(0, 6)$, $(1, 0)$ and $(2, 0)$. (see `https://www.geogebra.org/m/aqtkdpbm`)

assume $y = a + bx$
plug in points

$$\left.\begin{aligned} 6 &= a + b(0) \\ 0 &= a + b(1) \\ 0 &= a + b(2) \end{aligned}\right\} \text{must be true for line to exist}$$

system of eqns

$$\begin{array}{ccc} a & b & \\ \end{array}$$
$$\left[\begin{array}{cc|c} 1 & 0 & 6 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \end{array}\right] \leftarrow \text{inconsistent!}$$

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \vec{x} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \text{ has no soln!}$$

Solve $A^T A \hat{x} = A^T b$ instead (to get closest thing to $(6,0,0)$ in $\mathrm{Col}(A)$)

$A^T A = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}$ $A^T b = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$

find $\hat{x}$ $\left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right] \rightsquigarrow \begin{aligned} a &= 5 \\ b &= -3 \end{aligned}$ — so $y = 5 - 3x$

Observation 4.3.2. When $A\mathbf{x} = \mathbf{b}$ has no solution, solve $A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}$ instead. In this example, the heights of the best line are $\mathbf{p} = (p_1, \ldots, p_m)$. The vertical distances to the data points are the errors $\mathbf{e} = (e_1, \ldots, e_m)$.

Example 4.3.3. Find the parabola $y = \cancel{a + bx + x^2}$ $y = a + bx + cx^2$ through the points $(0, 6)$, $(1, 0)$ and $(2, 0)$. (see `https://www.geogebra.org/m/aqtkdpbm`)

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 6 \\ 1 & 1 & 1 & 0 \\ 1 & 2 & 4 & 0 \end{array}\right] \sim \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \left[\begin{array}{ccc|c} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -9 \\ 0 & 0 & 1 & 3 \end{array}\right] \end{array}$$

consistent

$\Rightarrow$ there is a parabola thru these pts

$y = a + bx + cx^2$

$$\boxed{y = 6 - 9x + 3x^2}$$

Example 4.3.4. Find the plane through the points $(1,1,1)$, $(1,1,3)$ and $(1,2,3)$. Use the fact that a plane in $\mathbb{R}^3$ has the equation

$$x = ay + bz + c.$$

(see `https://www.geogebra.org/m/rgespmyf`)

Example 4.3.5. Find a plane that best fits $(1, 1, 1)$, $(2, 2, 1)$, $(1, 1, 3)$ and $(1, 2, 3)$. Use the fact that a plane in $\mathbb{R}^3$ has the equation

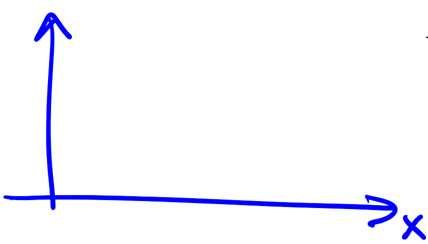
$$x = ay + bz + c.$$

(see `https://www.geogebra.org/m/rgespmyf`)

Definition 4.3.6. For an $m \times n$ matrix A and a vector $\mathbf{b}$, we cannot always get the error $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ down to zero. When $\mathbf{e}$ is zero, $\mathbf{x}$ is an exact solution to $A\mathbf{x} = \mathbf{0}$. *When the length of* $\mathbf{e}$ *is as small as possible,* $\widehat{\mathbf{x}}$ *is a* **least squares solution**. This happens when $\widehat{\mathbf{x}}$ is a solution to $A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}$. T

In English...

Example 4.3.7. You're taking average temperatures each month for the winter and spring months of 2020. You find the following data



Month	Average Temp C°	
Jan	3	(1,3)
Feb	4	(2,4)
Mar	4	(3,4)
Apr	5	(4,5)
May	5	(5,5)

After sketching a scatterplot, you guess that a curve of the form $y = a + b\sqrt{x}$ will fit the data best. Find an equation of this type that best fits the data. How well does it fit?
(see `https://www.geogebra.org/m/maf7nqrv`)

$y = a + b\sqrt{x}$

$$3 = a + \sqrt{1}\,b$$
$$4 = a + \sqrt{2}\,b$$
$$4 = a + \sqrt{3}\,b$$
$$5 = a + \sqrt{4}\,b$$
$$5 = a + \sqrt{5}\,b$$

$$A = \begin{bmatrix} 1 & \sqrt{1} \\ 1 & \sqrt{2} \\ 1 & \sqrt{3} \\ 1 & \sqrt{4} \\ 1 & \sqrt{5} \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 5 & 8.38 \\ 8.38 & 15 \end{bmatrix} \qquad A^Tb = \begin{bmatrix} 21 \\ 36.76 \end{bmatrix}$$

$$A^TA\hat{x} = A^Tb \rightarrow \begin{array}{l} a = 1.46 \\ b = 1.64 \end{array}$$

so $y = 1.46 + 1.64\sqrt{x}$ minimizes error

Observation 4.3.8. The least squares solution $\widehat{\mathbf{x}}$ minimizes $E = \frac{1}{n}\|\mathbf{e}\|^2 = \frac{1}{n}\|A\mathbf{x} - \mathbf{b}\|^2$. This is the sum of squares of the errors in the m equations $(m > n)$.