6.5: least-squares problems.

In this section, we will study *inconsistent* systems of linear equations and learn how to find the 'best possible solutions' of such systems. The necessity of 'solving' inconsistent systems arises in the computation of least squares regression lines, as shown below.

1. Least Squares Motivation

Suppose we have three data points on a graph and we want to find a line through them. Unless the data points are already on a line, there is not a line (a linear function!) that will go through all three points. To compromise, we will search for the line $y = c_0 + c_1\, x$ that 'best' fits the data.

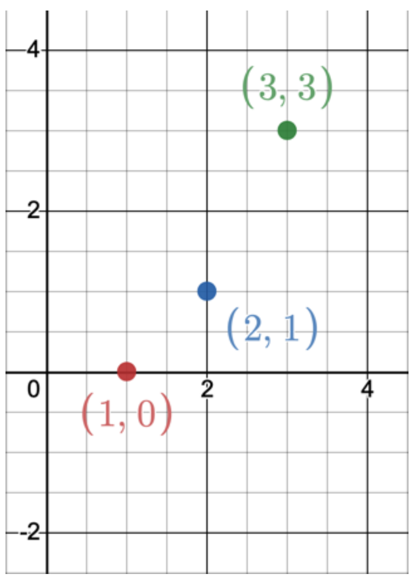
How can we write this as a system of equations? And how can we write this in matrix form?

$0 = c_0 + c_1 \cdot 1$

$1 = c_0 + c_1 \cdot 2$

$3 = c_0 + c_1 \cdot 3$

$$\Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$



$$\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 3 & 3 \end{array}\right] \xrightarrow[\text{reduce}]{\text{row}} \underbrace{\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & -1/2 \end{array}\right]}_{\text{inconsistent system!}}$$

$\Rightarrow$ no line $y = c_0 + c_1 x$ goes through all 3 points.

The points are not collinear (along the same line), so the system is inconsistent. Although it is impossible to find $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{b}$, we look for an $\mathbf{x}$ that minimizes the norm of the error $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$. The smaller the norm of $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$, the better the approximation. The solution $\mathbf{x} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$ of this minimization problem results in the **least squares regression line** $y = c_0 + c_1\, x$.

To get started on this problem, we need one big piece of information from a section we didn't cover, Section 6.3 which discussed 'Orthogonal Projections.'

Theorem 1.1. The Best Approximation Theorem: *Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then, $\hat{\mathbf{y}}$ is the closest points in W to $\mathbf{y}$ in the sense that*

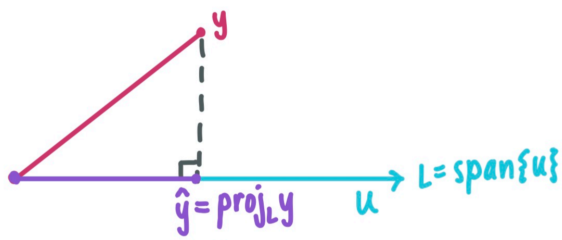
$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$$

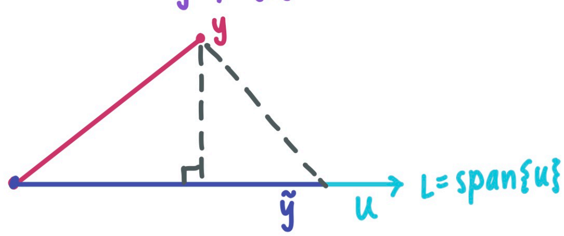
for all $\mathbf{v}$ in W which are distinct from $\hat{\mathbf{y}}$.

In essence, this tells us that the orthogonal projection $\hat{\mathbf{y}}$ in W is the 'best' approximation of $\mathbf{y}$. The statement in the theorem $\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$ states:

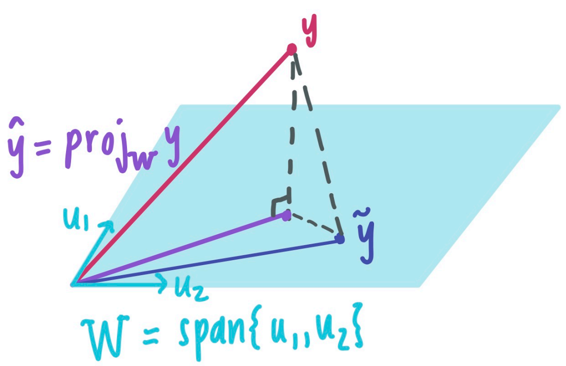
distance B/W y & $\hat{y}$ < distance Between y & any other vector in W

We see illustrations of this below in 2D and 3D.



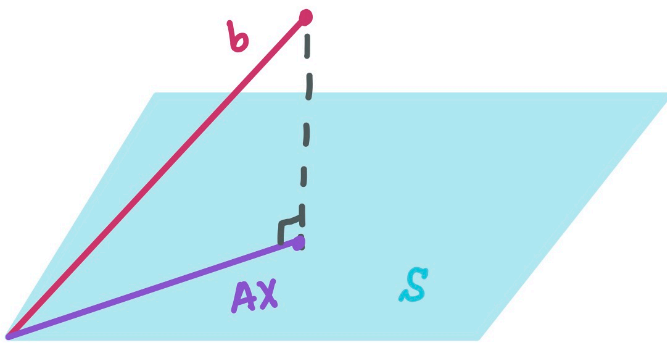


* orthogonal projections are the "Best approximation" of vector y on to some space W.



With the motivating problem for this section, we are trying to find the vector $\mathbf{x} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$ that minimizes $\|\mathbf{Ax} - \mathbf{b}\|$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^m$.

We are attempting to find the vector $\mathbf{b}$ that is as close as possible to the vector $\mathbf{Ax}$ in the space $S = \text{span}\{\text{columns of } \mathbf{A}\}$, or the range of the matrix $\mathbf{A}$. From the Best Approximation Theorem, we know that the desired vector is the projection of $\mathbf{b}$ onto S.



$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

$$A \cdot x = b$$

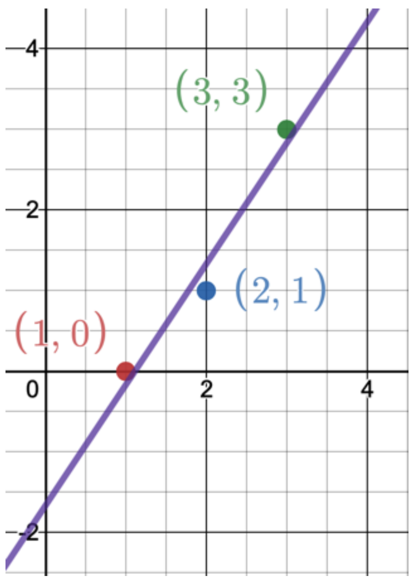
Without getting into some of the details we glossed over in Chapter 4... we have the following property based on orthogonality $\mathbf{A^T}(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}$. Manipulating this equation gives us the following:

$$\mathbf{A^T}(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}$$
$$\mathbf{A^T Ax} - \mathbf{A^T b} = \mathbf{0}$$
$$\boxed{\mathbf{A^T Ax} = \mathbf{A^T b}}$$

The solution of the least squares problem comes down to solving the $n \times n$ system $\mathbf{A^T Ax} = \mathbf{A^T b}$ for the vector $\mathbf{x}$. These equations are known as the **normal equations** of the least squares problem $\mathbf{Ax} = \mathbf{b}$.

Example 1.1. *Find the solution of the least squares problem* $\mathbf{Ax} = \mathbf{b}$ *defined below. Using that solution, find the equation of the line noted below.*

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$



normal eqns : $A^T A x = A^T \cdot b$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} 1+1+1 & 1+2+3 \\ 1+2+3 & 1+4+9 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

$$= \begin{bmatrix} 0+1+3 \\ 0+2+9 \end{bmatrix}$$

$$A^T \cdot b = \begin{bmatrix} 4 \\ 11 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \end{bmatrix}$$

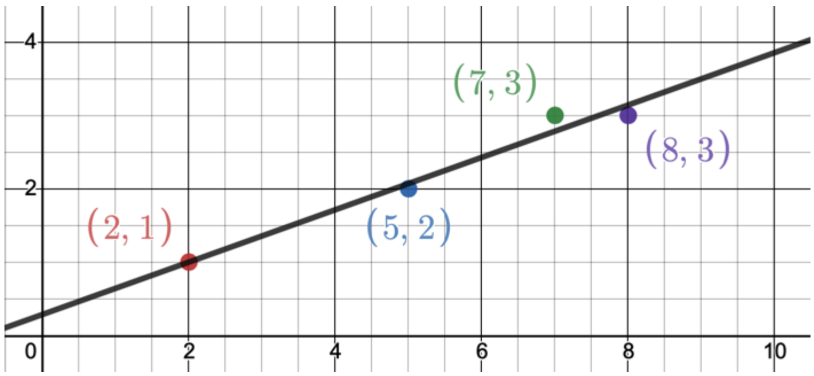
$$\begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 4 \\ 11 \end{bmatrix}$$

$$= \frac{1}{6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 11 \end{bmatrix}$$

$$\begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 11 \end{bmatrix} = \boxed{\begin{bmatrix} -5/3 \\ 3/2 \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}}$$

so, $y = -\frac{5}{3} + \frac{3}{2} \cdot x$
'best' fits the line through the data pts

Example 1.2. *Find the equation* $y = c_0 + c_1 x$ *of the least-squares line that best fits the data points* $(2,1),(5,2),(7,3),(8,3)$. *First, write out the system in matrix form* $\mathbf{A}\mathbf{x} = \mathbf{b}$, *then determine the normal equations for the least squares problem.*



$c_0 + c_1 \cdot 2 = 1$

$c_0 + c_1 \cdot 5 = 2$

$c_0 + c_1 \cdot 7 = 3$

$c_0 + c_1 \cdot 8 = 3$

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} c_0 \\ c_1 \end{bmatrix}}_{x} = \underbrace{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}}_{b}$$

$\Rightarrow$ normal eqns:

$$A^T A \cdot x = A^T b$$

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

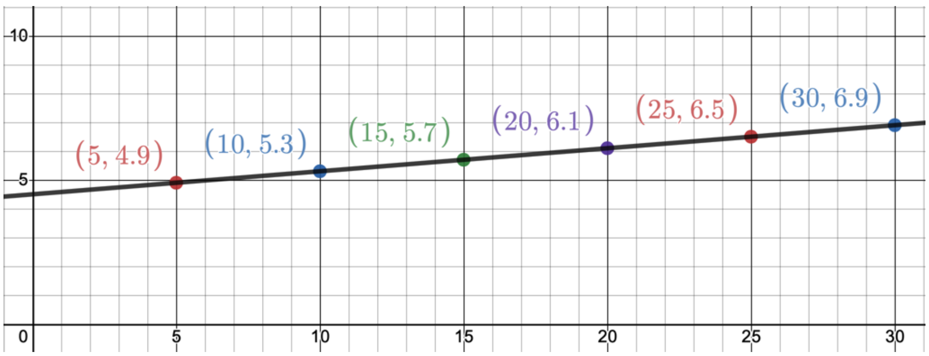
$$\begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \cdot \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \cdot \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

$\boxed{\Rightarrow y = \frac{2}{7} + \frac{5}{14} x \text{ is the 'Best' fit for a line that goes through all 4 data points.}}$

Example 1.3. *The table below show the world population for six different years.*

Year	*1985*	*1990*	*1995*	*2000*	*2005*	*2010*
Population (in billions)	*4.9*	*5.3*	*5.7*	*6.1*	*6.5*	*6.9*

Let $x = 5$ represent the year 1985. Find the least squares regression quadratic polynomial $y = c_0 + c_1 x + c_2 x^2$ for the data and use the model to estimate the population for the year 2020.



$\boxed{y = 4.5 + .08X}$

$$4.9 = c_0 + c_1 \cdot 5 + c_2 \cdot (5)^2$$
$$5.3 = c_0 + c_1 \cdot 10 + c_2 \cdot (10)^2$$
$$5.7 = c_0 + c_1 \cdot 15 + c_2 (15)^2$$
$$6.1 = c_0 + c_1 \cdot 20 + c_2 \cdot (20)^2$$
$$6.5 = c_0 + c_1 \cdot 25 + c_2 \cdot (25)^2$$
$$6.9 = c_0 + c_1 \cdot 30 + c_2 \cdot (30)^2$$

$$\Rightarrow \begin{bmatrix} 1 & 5 & 25 \\ 1 & 10 & 100 \\ 1 & 15 & 225 \\ 1 & 20 & 400 \\ 1 & 25 & 625 \\ 1 & 30 & 900 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 4.9 \\ 5.3 \\ 5.7 \\ 6.1 \\ 6.5 \\ 6.9 \end{bmatrix}$$

$$A \cdot x = b$$

normal eqns: $A^T A x = A^T b$

$$\begin{bmatrix} 6 & 105 & 2{,}275 \\ 105 & 2{,}275 & 55{,}125 \\ 2{,}275 & 55{,}125 & 1{,}421{,}875 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 35.4 \\ 654.5 \\ 14{,}647.5 \end{bmatrix}$$

$$\boxed{\Rightarrow \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 4.5 \\ .08 \\ 0 \end{bmatrix}}$$

· this says line of best fit is linear! $\boxed{y = 4.5 + .08x} + 0 \cdot x^2$

2020 → $x=40$ w/ our data.

$y = 4.5 + .08(40) = 7.7$

so, in 2020, the model predicts 7.7 billion people for population.