**Example 39.** Find the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$, where

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

**Solution.** First, $A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$ and $A^T\boldsymbol{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$.

Hence, the normal equations $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$ take the form $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}\hat{\boldsymbol{x}} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$.

Solving, we immediately find $\hat{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

**Check.** The error $A\hat{\boldsymbol{x}} - \boldsymbol{b} = \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ is indeed orthogonal to $\operatorname{col}(A)$. Because $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} = 0$ and $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = 0$.

**Definition 40.** The **(orthogonal) projection** $\hat{\boldsymbol{b}}$ of a vector $\boldsymbol{b}$ onto a subspace $Y$ is the vector in $Y$ closest to $\boldsymbol{b}$.

| |
|---|
| The (orthogonal) projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\operatorname{col}(A)$ is $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$.<br>Here, $\hat{\boldsymbol{x}}$ is a least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ (i.e. $\hat{\boldsymbol{x}}$ solves $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$). |

**Why?** Why is $A\hat{\boldsymbol{x}}$ the projection of $\boldsymbol{b}$ onto $\operatorname{col}(A)$?

Because, for a least squares solution $\hat{\boldsymbol{x}}$, $A\hat{\boldsymbol{x}} - \boldsymbol{b}$ is as small as possible (and any element in $\operatorname{col}(A)$ is of the form $A\boldsymbol{x}$ for some $\boldsymbol{x}$).

**Note.** This is a recipe for computing any orthogonal projection! That's because every subspace $Y$ can be written as $\operatorname{col}(A)$ for some choice of the matrix $A$ (take, for instance, $A$ so that its columns are a basis for $Y$).

**Example 41.** What is the orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto $\operatorname{span}\left\{ \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right\}$?

**Solution.** This is the same question as: what is the projection of $\boldsymbol{b}$ onto $\operatorname{col}(A)$, with $A$ and $\boldsymbol{b}$ as in the previous example.

The projection of $\boldsymbol{b}$ onto $\operatorname{Col}(A)$ is $A\hat{\boldsymbol{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

**Example 42. (extra homework)**

(a) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ onto $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\}$?

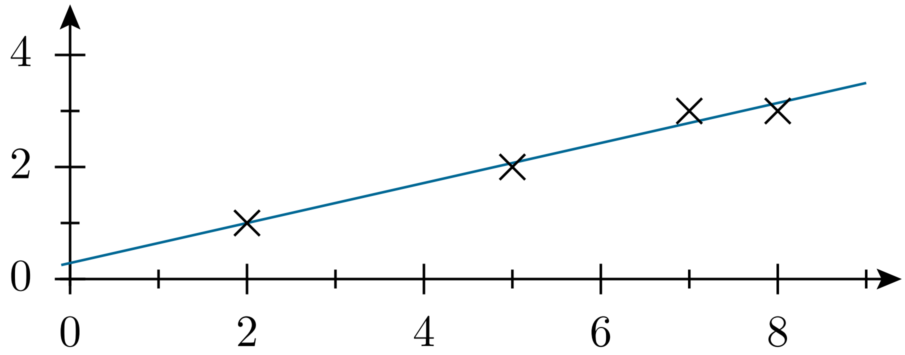
(b) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ onto $\operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$?

**Solution. (final answer only)** The projections are $\left(\frac{11}{6}, \frac{1}{3}, \frac{7}{6}\right)^T$ and $\left(\frac{3}{2}, 0, \frac{3}{2}\right)^T$.


Armin Straub
straub@southalabama.edu

Experimental data: $(x_i, y_i)$

Wanted: parameters $a, b$ such that $y_i \approx a + b x_i$ for all $i$




This approximation should be so that $\mathrm{SS_{res}} = \underbrace{\sum_i [y_i - (a + b x_i)]^2}_{\text{residual sum of squares}}$ is as small as possible.

**Example 43.** Determine the line that best fits the data points $(2,1), (5,2), (7,3), (8,3)$.

**Solution.** We need to determine the values $a, b$ for the best-fitting line $y = a + bx$.

If there was a line that fit the data perfectly, then:

$$\begin{aligned} a + 2b &= 1 \qquad (2,1) \\ a + 5b &= 2 \qquad (5,2) \\ a + 7b &= 3 \qquad (7,3) \\ a + 8b &= 3 \qquad (8,3) \end{aligned}$$

In matrix form, this is: $\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{bmatrix}}_{\text{design matrix } X} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}}_{\text{observation vector } \boldsymbol{y}}$ (writing the points as $(x_i, y_i)$)

Using our points, these equations become $\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$. [This system is inconsistent (as expected).]

We compute a least squares solution.

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}, \qquad X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}.$$

Solving the normal equations $\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$, we find $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$.

Hence, the least squares line is $y = \frac{2}{7} + \frac{5}{14} x$.

The plot above shows our points together with this line. It does look like a very good fit!

Armin Straub
straub@southalabama.edu