1. Find the projection \( p \) of \( b = \begin{bmatrix} 1 \\ -4 \\ 1 \end{bmatrix} \) onto the line through \( \begin{bmatrix} 4 \\ 4 \\ 2 \end{bmatrix} \) and the projection matrix \( P \) so that \( p = Pb \). Draw a picture showing how \( b, p, \) and the line are related geometrically.

Using \( a = \begin{bmatrix} 1 \\ -4 \\ 1 \end{bmatrix} \), \( a \cdot a = 16 + 16 + 2 = 36 \), \( a \cdot b = 4 - 16 + 2 = -10 \).

\[
p = a \frac{a \cdot b}{a \cdot a} = \frac{-10}{36} \begin{bmatrix} -10 \\ -10 \\ -5 \end{bmatrix}
\]

\[
P = a a^T = \frac{1}{36} \begin{bmatrix} 16 & 16 & 8 \\ 16 & 16 & 8 \\ 8 & 8 & 4 \end{bmatrix} = \begin{bmatrix} \frac{4}{9} & \frac{4}{9} & \frac{2}{9} \\ \frac{4}{9} & \frac{4}{9} & \frac{2}{9} \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}
\]

Draw a picture and note that the vector \( p - b \) is perpendicular to the line.

2. Suppose we want to find the projection \( p \) of \( b \) onto \( V = \text{span}\{a_1, a_2, \ldots, a_n\} \) where the \( a_i \)'s are linearly dependent. Why will our usual formula \( p = A(A^T A)^{-1} A^T b \) fail? What can we do to get around this problem and still find \( p \)?

As stated in class, \( A^T A \) is invertible if and only if the columns of \( A \) are linearly independent. Since the columns of \( A \) are linearly dependent, \( A^T A \) is not invertible, so calculating \( (A^T A)^{-1} \) clearly fails.

The way to get around the problem is to find use GE to find the pivot columns, then create \( A' \) with only the pivot columns. A theorem says that the pivot columns together form a basis for the column space, so the columns of \( A' \) are linearly independent and we can calculate a projection point. Furthermore, the columns of \( A' \) span \( \text{col}(A) \), \( (\text{col}(A') = \text{col}(A)) \) so the projection point using \( A' \) must also be the closest point in \( \text{col}(A) \) to \( b \), so we get the answer we were originally looking for.

3. Explain why the solution(s) of the normal equations solves the least squares problem, i.e., why the solution makes the error as small as possible.

The key fact that we must use is that given the correct projection point \( p \) that is closest to \( b \), the error vector \( e = p - b \) is orthogonal to the column space of \( A \). Last week's homework showed that the co-kernel of \( A \) is orthogonal to the column space of \( A \), and some simple dimensional analysis shows that the co-kernel is the largest space orthogonal to the column space of \( A \), so it must be the orthogonal complement of \( A \). (or alternatively, this fact is listed as a theorem in the book) Thus, the error vector \( e \) is in the co-kernel, or \( e \in \text{Nul}(A^T) \).

Since the error vector is \( e = (p - b) = (Ax - b) \) for some \( x \) that makes \( e \perp \text{col}(A) \), we
know

\[ A^T e = 0 \]
\[ A^T (Ax - b) = 0 \]
\[ A^T Ax - A^T b = 0 \]
\[ A^T Ax = A^T b \]

So then a solution \( x \) to the normal equation \( A^T Ax = A^T b \) gives the unique \( p = Ax \) such that \( (p - b) \perp \text{col}(A) \).