Suppose \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \) are the columns of \( \mathbf{A} \) and suppose \( X = [x_1, x_2, \ldots, x_n] \). Then \( \mathbf{A} = (\mathbf{v}_1 | \mathbf{v}_2 | \ldots | \mathbf{v}_n) \) and \( AX = (x_1 | x_2 | \ldots | x_n)^T \). Hence \( AX = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \ldots + x_n \mathbf{v}_n \).

iff \( b = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \ldots + x_n \mathbf{v}_n \)

iff \( b \) is a linear combination of the columns of \( \mathbf{A} \).

iff \( b \) is in the column space of \( \mathbf{A} \).

The least-squares solution for \( AX = b \) is the projection of \( b \) onto \( \mathbf{W} \).

This is the \textit{least-squares} solution, it is the best approximate solution of \( AX = b \).

We could find the least-squares solution by calculating \( \text{proj}_W b \) and then solving \( AX = \text{proj}_W b \). But there is an easier way.

The exact equation is \( A \T X = (A \T b) \).

The least-squares solution for \( AX = b \)

\[ AX = \text{proj}_W b \text{ (by definition of “least-squares”) \} \}

\[ b-AX = b - \text{proj}_W b \text{ is \perp to the column space of } A. \] revealing the closest point to \( b \) in \( \mathbf{W} \).

\[ v_i \cdot (b-AX) = 0 \text{ for each column } v_i. \]

\[ \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} (b-AX) = \mathbf{0}. \]

\[ \Rightarrow \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} = A \T (b-AX) = \mathbf{0}. \]

\[ \Rightarrow A \T (b-AX) = \mathbf{0}. \]

\[ \Rightarrow A \T b - A \T AX = \mathbf{0}. \]

\[ \Rightarrow A \T AX = A \T b. \quad \text{exact equation} \]

To solve \( A \T AX = (A \T b) \), first find \( A \T b \) and \( A \T A \).

**Least squares approximation**

Note. For column vectors \( \mathbf{u}, \mathbf{v} \): the dot product \( \mathbf{v} \cdot \mathbf{u} = \) the matrix product \( \mathbf{v}^T \mathbf{u} \).

Let \( \mathbf{A} \) be a matrix.

Let \( \mathbf{W} = \) the column space of \( \mathbf{A} = \) the space spanned by the columns of \( \mathbf{A} \).

**THEOREM.** \( AX = b \) has a solution iff \( b \) is in the column space of \( \mathbf{A} \).

**PROOF.** Suppose \( \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \) are the columns of \( \mathbf{A} \) and suppose \( X = [x_1, x_2, \ldots, x_n] \).

\[ AX = (x_1 | x_2 | \ldots | x_n)^T = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \ldots + x_n \mathbf{v}_n. \]

Hence \( AX \)

iff \( b = x_1 \mathbf{v}_1 + x_2 \mathbf{v}_2 + \ldots + x_n \mathbf{v}_n \)

iff \( b \) is a linear combination of the columns of \( \mathbf{A} \).

iff \( b \) is in the column space of \( \mathbf{A} \).

Suppose \( b \) is in the column space \( \mathbf{W} \) of \( \mathbf{A} \). Thus \( AX = b \) has no solution.

**The error vector** of an approximate solution to \( AX = b \) is the difference \( e = b - AX \) between the desired value \( b \) and approximate value \( AX \) found. The least-squares solution has the smallest error, i.e., \( ||e|| \) is minimum.

**Approximating functions**

Suppose we know the values \( f(t_1), f(t_2), f(t_3) \) of an otherwise unknown function \( f(t) \). Suppose we wish to approximate it as a linear combination \( af_1(t) + bf_2(t) + cf_3(t) \) of three known functions \( f_1, f_2, f_3 \).

Thus we wish to find the \( X = [a, b, c]^T \) such that \( af_1(t_1) + bf_2(t_1) + cf_3(t_1) = f(t_1) \)

\[ af_1(t_2) + bf_2(t_2) + cf_3(t_2) = f(t_2) \]

\[ af_1(t_3) + bf_2(t_3) + cf_3(t_3) = f(t_3) \]

\[ \vdots \]

\[ af_1(t_n) + bf_2(t_n) + cf_3(t_n) = f(t_n) \]