- New concepts:
  - Projection review, projection onto subspaces
  - Best approximation
  - LSQ through projection
  - LSQ solution by normal equations

- Orthogonal projection of $\boldsymbol{v} \in \mathbb{R}^m$ along direction $\boldsymbol{q} \in \mathbb{R}^m$, $\|\boldsymbol{q}\| = 1$

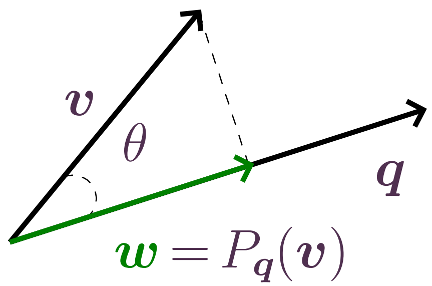


Figure 1. Orthogonal projection operation $P_{\boldsymbol{q}}$.

- $\boldsymbol{w} = (\|\boldsymbol{v}\| \cos\theta)\, \boldsymbol{q} = \left(\|\boldsymbol{v}\| \dfrac{\boldsymbol{q} \cdot \boldsymbol{v}}{\|\boldsymbol{q}\|\, \|\boldsymbol{v}\|} \right) \boldsymbol{q} = (\boldsymbol{q}^T \boldsymbol{q}) \boldsymbol{q} = \boldsymbol{q}(\boldsymbol{q}^T \boldsymbol{v}) = (\boldsymbol{q}\boldsymbol{q}^T)\boldsymbol{v} \Rightarrow$
- Projection matrix $\boldsymbol{P}_{\boldsymbol{q}} = \boldsymbol{q}\boldsymbol{q}^T$ ($\|\boldsymbol{q}\| = 1$)

- Simple example: projection in $\mathbb{R}^3$ onto x_1x_2-plane of $\boldsymbol{v} \in \mathbb{R}^3$

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \boldsymbol{I}\boldsymbol{v} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 + v_3\boldsymbol{e}_3$$

- The projection of $\boldsymbol{v}$ onto x_1x_2 plane is $\boldsymbol{w} = v_1\boldsymbol{e}_1 + v_2\boldsymbol{e}_2 = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}$
- Projection is linear mapping, hence $\boldsymbol{w} = \boldsymbol{P}_{12}\,\boldsymbol{v}$

$$\boldsymbol{P}_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \boldsymbol{e}_1\boldsymbol{e}_1^T + \boldsymbol{e}_2\boldsymbol{e}_2^T$$

- Recall definition of orthonormal vectors
- Columns of $\boldsymbol{Q} = [\ \boldsymbol{q}_1 \ \ \dots \ \ \boldsymbol{q}_n\ ] \in \mathbb{R}^{m\times n}$ are orthonormal if

$$\boldsymbol{Q}^T\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1^T \\ \vdots \\ \boldsymbol{q}_n^T \end{bmatrix} [\ \boldsymbol{q}_1 \ \ \dots \ \ \boldsymbol{q}_n\ ] = \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{q}_1 & \boldsymbol{q}_1^T\boldsymbol{q}_2 & \dots & \boldsymbol{q}_1^T\boldsymbol{q}_n \\ \boldsymbol{q}_2^T\boldsymbol{q}_1 & \boldsymbol{q}_2^T\boldsymbol{q}_2 & \dots & \boldsymbol{q}_2^T\boldsymbol{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_n^T\boldsymbol{q}_1 & \boldsymbol{q}_n^T\boldsymbol{q}_2 & \dots & \boldsymbol{q}_n^T\boldsymbol{q}_n \end{bmatrix} = \boldsymbol{I}_n$$

- Consider $C(\boldsymbol{Q})$ with $\boldsymbol{Q} = [\ \boldsymbol{q}_1 \ \ \dots \ \ \boldsymbol{q}_n\ ] \in \mathbb{R}^{m\times n}$ orthonormal. The matrix

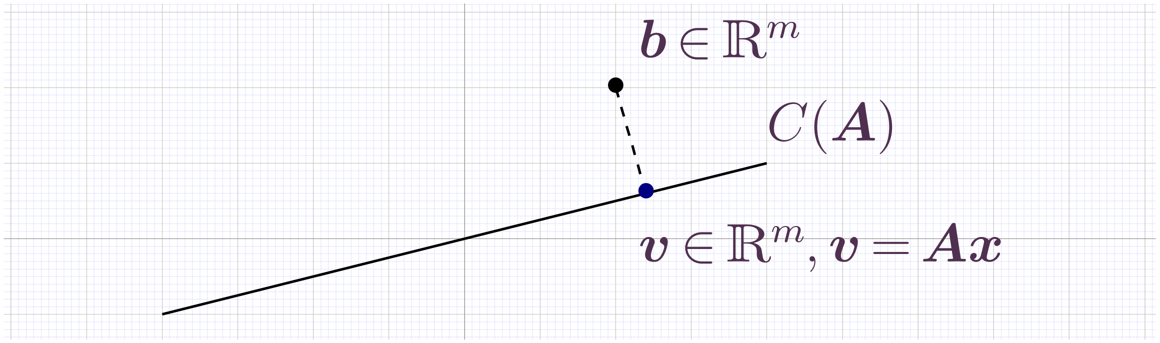
$$\boldsymbol{P}_{\boldsymbol{Q}} = \boldsymbol{Q}\boldsymbol{Q}^T = [\ \boldsymbol{q}_1 \ \ \dots \ \ \boldsymbol{q}_n\ ] \begin{bmatrix} \boldsymbol{q}_1^T \\ \vdots \\ \boldsymbol{q}_n^T \end{bmatrix} = \boldsymbol{q}_1\boldsymbol{q}_1^T + \dots + \boldsymbol{q}_n\boldsymbol{q}_n^T$$

is the standard matrix of the orthogonal projection of a vector in $\mathbb{R}^m$ onto the subspace spanned by $\{\ \boldsymbol{q}_1 \ \ \dots \ \ \boldsymbol{q}_n\ \}$, namely $C(\boldsymbol{Q})$.

- Consider approximating $\boldsymbol{b} \in \mathbb{R}^m$ by linear combination of n vectors, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$
- Make approximation error $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{v} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$ as small as possible

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2$$

Error is measured in the 2-norm $\Rightarrow$ the *least squares problem* (LSQ)



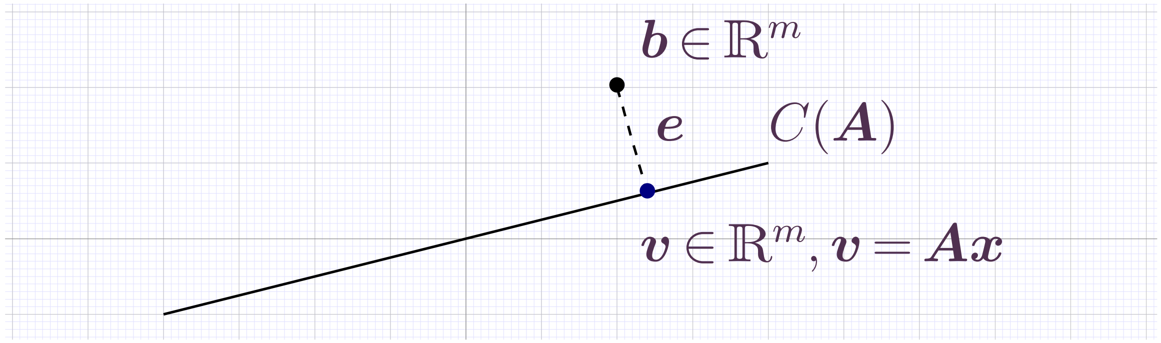
- Solution is the projection of $\boldsymbol{b}$ onto $C(\boldsymbol{A})$

$$\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{A}, \boldsymbol{P}_{C(\boldsymbol{A})} = \boldsymbol{Q}\boldsymbol{Q}^T, \boldsymbol{v} = (\boldsymbol{Q}\boldsymbol{Q}^T)\boldsymbol{b}$$

- The vector $\boldsymbol{x}$ is found by back-substitution from

$$\boldsymbol{v} = (\boldsymbol{Q}\boldsymbol{Q}^T)\boldsymbol{b} = (\boldsymbol{Q}\boldsymbol{R})\boldsymbol{x} \Rightarrow \boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^T\boldsymbol{b}.$$

- The best approximant is found when the error vector $\boldsymbol{e}$ is orthogonal to $C(\boldsymbol{A})$



$$\boldsymbol{e} \perp C(\boldsymbol{A}) \Rightarrow \boldsymbol{A}^T\boldsymbol{e} = 0 \Rightarrow \boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = 0 \Rightarrow (\boldsymbol{A}^T\boldsymbol{A})\,\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$$

- The system $(\boldsymbol{A}^T\boldsymbol{A})\,\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$ is *the normal system* of the LSQ problem
- Note that $\boldsymbol{A}^T\boldsymbol{A} \in \mathbb{R}^{n \times n}$