Chapter 4

Orthogonality

Po-Ning Chen, Professor

Department of Electrical Engineering

National Chiao Tung University

Hsin Chu, Taiwan 30050, R.O.C.

4.1 Orthogonality of four subspaces

- The general relation of the four subspaces:

Answer:
- $\boldsymbol{C}(A)$ is perpendicular/orthogonal to $\boldsymbol{N}(A^{\mathsf{T}})$.
- $\boldsymbol{R}(A)$ is perpendicular/orthogonal to $\boldsymbol{N}(A)$.

- How the four subspaces are related to $A\boldsymbol{x} = \boldsymbol{b}$ system?

Answer:
- All $\boldsymbol{b}$'s in $A\boldsymbol{x} = \boldsymbol{b}$ gives $\boldsymbol{C}(A)$.
- All $\boldsymbol{x}$'s in $A\boldsymbol{x} = \boldsymbol{0}$ gives $\boldsymbol{N}(\boldsymbol{A})$.

- However, what if $A\boldsymbol{x} = \boldsymbol{b}$ has no solution? Can we find $\boldsymbol{x}$ such that $\|\boldsymbol{e}\|^2$ is minimized, where $\boldsymbol{e} = \boldsymbol{b} - A\boldsymbol{x}$?

Answer: Yes, we certainly can. This problem has a huge number of applications in reality.

4.1 Orthogonality of four subspaces

Definition (Orthogonality of subspaces): Two subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ are said to be *orthogonal* if

$$\boldsymbol{v}^{\mathsf{T}}\boldsymbol{w} = 0 \text{ for all } \boldsymbol{v} \in \boldsymbol{V} \text{ and all } \boldsymbol{w} \in \boldsymbol{W}.$$

- For orthogonal $\boldsymbol{V}$ and $\boldsymbol{W}$, $\boldsymbol{V} \cap \boldsymbol{W} = \{\mathbf{0}\}$ is a set consisting of one element $\mathbf{0}$, not an empty set.
- Note that $\mathbf{0}^{\mathsf{T}}\mathbf{0} = 0$; so, in a sense, $\mathbf{0}$ is orthogonal to itself.

  In fact, $\mathbf{0}$ is the only vector that is orthogonal to itself.

Exercise. Prove that $\boldsymbol{R}(A) \perp \boldsymbol{N}(A)$.

Answer. $\boldsymbol{R}(A)$ contains all vectors of the form $A^{\mathsf{T}}\boldsymbol{y}$, and all vectors $\boldsymbol{x}$ in $\boldsymbol{N}(A)$ satisfies $A\boldsymbol{x} = 0$. Hence,

$$\boldsymbol{x} \cdot A^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathsf{T}}A^{\mathsf{T}}\boldsymbol{y} = (A\boldsymbol{x})^{\mathsf{T}}\boldsymbol{y} = \mathbf{0}^{\mathsf{T}}\boldsymbol{y} = \mathbf{0}.$$

$\square$

4.1 Orthogonality of four subspaces

Exercise (Problem 9). Prove that $A^{\mathsf{T}}A$ has the same null space as A.

(Problem 9, Section 4.1) If $A^{\mathsf{T}}A\boldsymbol{x} = \mathbf{0}$ then $A\boldsymbol{x} = \mathbf{0}$. Reason: $A\boldsymbol{x}$ is in the nullspace of A^{T} and also in the _______ of A and those spaces are _______. *Conclusion: $A^{\mathsf{T}}A$ has the same nullspace as A. This key fact is repeated in the next section.*

Proof: It suffices to prove that $A^{\mathsf{T}}A\boldsymbol{x} = \mathbf{0}$ iff $A\boldsymbol{x} = \mathbf{0}$.
Note that $A\boldsymbol{x} = \mathbf{0}$ implying $A^{\mathsf{T}}A\boldsymbol{x} = \mathbf{0}$ trivially hold.

- $A^{\mathsf{T}}A\boldsymbol{x} = \mathbf{0} \implies A\boldsymbol{x} \in \boldsymbol{N}(A^{\mathsf{T}})$.
- By definition of the column space of A, $A\boldsymbol{x} \in \boldsymbol{C}(A)$.
- So, $A\boldsymbol{x} \perp A\boldsymbol{x}$, which implies $A\boldsymbol{x} = \mathbf{0}$.

$\square$

Alternative proof is that

- $B = A^{\mathsf{T}}A$. Hence, $\boldsymbol{R}(B) = \boldsymbol{R}(A)$, which implies $\boldsymbol{N}(B) = \boldsymbol{N}(A)$ because the **null space** is the **orthogonal complement** (see the next slide) of the **row space**. $\square$

4.1 Orthogonal complement

Definition (Orthogonal complement): The orthogonal complement of $\boldsymbol{V}$ contains of all vectors that are perpendicular to vectors in $\boldsymbol{V}$. It is denoted by $\boldsymbol{V}^{\perp}$ and is pronounced as "V perp."

- $\boldsymbol{R}(A)$ and $\boldsymbol{N}(A)$ are not just **orthogonal** to each other, but are **orthogonal complement** to each other!

$$\boldsymbol{N}(A)^{\perp} = \boldsymbol{C}(A^{\mathsf{T}}).$$

Example of orthogonal subspaces that are not mutually orthogonal complement.

1. $\{\mathbf{0}\}$ and xy-plane in $\Re^3$.
2. Two perpendicular lines in $\Re^3$.

4.1 Orthogonal complement

Important fact about orthogonal complements $\boldsymbol{V}$ and $\boldsymbol{V}^{\perp}$:

- Every vector $\boldsymbol{v}$ can be expressed as $\boldsymbol{v} = \boldsymbol{v}_r + \boldsymbol{v}_n$, where $\boldsymbol{v}_r \in \boldsymbol{V}$ and $\boldsymbol{v}_n \in \boldsymbol{V}^{\perp}$.
- $\boldsymbol{v}_r$ is the **projection** of $\boldsymbol{v}$ onto $\boldsymbol{V}$, and $\boldsymbol{v}_n$ is the **projection** of $\boldsymbol{v}$ onto $\boldsymbol{V}^{\perp}$.

Example.

1. Given $A_{m\times n}$, every $\boldsymbol{v}_{n\times 1}$ can be expressed as

$$\boldsymbol{v}_{n\times 1} = \boldsymbol{x}^{(r)}_{n\times 1} + \boldsymbol{x}^{(n)}_{n\times 1},$$

   where $A^{\mathsf{T}}_{n\times m}\boldsymbol{w}_{m\times 1} = \boldsymbol{x}^{(r)}_{n\times 1}$ for some $\boldsymbol{w}_{m\times 1} \in \Re^m$ and $A_{m\times n}\boldsymbol{x}^{(n)}_{n\times 1} = \mathbf{0}_{m\times 1}$.

2. Given $A_{m\times n}$, every $\boldsymbol{v}_{m\times 1}$ can be expressed as

$$\boldsymbol{v}_{m\times 1} = \boldsymbol{x}^{(c)}_{m\times 1} + \boldsymbol{x}^{(ln)}_{m\times 1},$$

   where $A_{m\times n}\boldsymbol{w}_{n\times 1} = \boldsymbol{x}^{(c)}_{m\times 1}$ for some $\boldsymbol{w}_{n\times 1} \in \Re^n$ and $A^{\mathsf{T}}_{n\times m}\boldsymbol{x}^{(ln)}_{m\times 1} = \mathbf{0}_{n\times 1}$.

Lemma (Uniqueness of projection): The projection $\boldsymbol{x}^{(r)}$ of a vector $\boldsymbol{v}$ onto the row space is unique.

Proof:

- Suppose $\boldsymbol{x}^{(r)}$ and $\tilde{\boldsymbol{x}}^{(r)}$ are both projections of $\boldsymbol{v}$.
- $\boldsymbol{x}^{(r)} \in \boldsymbol{R}(A)$ and $\tilde{\boldsymbol{x}}^{(r)} \in \boldsymbol{R}(A) \quad \Rightarrow \quad \boldsymbol{x}^{(r)} - \tilde{\boldsymbol{x}}^{(r)} \in \boldsymbol{R}(A)$
  (property of a vector space)
- $A\boldsymbol{v} = A\boldsymbol{x}^{(r)} + A\boldsymbol{x}^{(n)} = A\boldsymbol{x}^{(r)} = A\tilde{\boldsymbol{x}}^{(r)} \quad \Rightarrow \quad A(\boldsymbol{x}^{(r)} - \tilde{\boldsymbol{x}}^{(r)}) = \boldsymbol{0}$.
  $\Rightarrow \quad \boldsymbol{x}^{(r)} - \tilde{\boldsymbol{x}}^{(r)} \in \boldsymbol{N}(A)$
- $\boldsymbol{R}(A) \perp \boldsymbol{N}(A) \quad \Rightarrow \quad \boldsymbol{x}^{(r)} - \tilde{\boldsymbol{x}}^{(r)} = \boldsymbol{0}$. $\square$

$A_{m\times n}$ defines a **(projection, possibly many-to-one) function** mapping from $\Re^n$ to $\boldsymbol{R}(A)$!

4.1 Orthogonal complement

Lemma: $A_{m\times n}$ defines a **(one-to-one) function** mapping from $\boldsymbol{R}(A)$ to $\boldsymbol{C}(A)$!

Proof:

- $\boldsymbol{v} = A\boldsymbol{x}^{(r)} \in \boldsymbol{C}(A)$, where $\boldsymbol{x}^{(r)} \in \boldsymbol{R}(A)$. (I.e., $\boldsymbol{x}^{(r)}$ maps to $\boldsymbol{v} = A\boldsymbol{x}^{(r)}$.)
- if $A\boldsymbol{x}^{(r)} = A\tilde{\boldsymbol{x}}^{(r)} \quad \Rightarrow \quad \boldsymbol{x}^{(r)} = \tilde{\boldsymbol{x}}^{(r)}$. (I.e., the vector that maps to $\boldsymbol{v} = A\boldsymbol{x}^{(r)}$ is unique!) $\square$

- This lemma also conforms to that the dimensions of $\boldsymbol{C}(A)$ and $\boldsymbol{R}(A)$ should be the same.

- Since the inverse mapping from $\boldsymbol{C}(A)$ to $\boldsymbol{R}(A)$ exists, A contains an $r \times r$ invertible submatrix.

  Just pick up the r pivot columns and pivot rows to form the invertible submatrix!

4.1 How to find the projection?

- (Example 5) Find the projections of $\boldsymbol{x} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ onto $\boldsymbol{R}(A)$ and $\boldsymbol{N}(A)$, where $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$?

Answer: **Inner product with the orthonormal basis.**

– $\texttt{rref}(A) = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$.

– Orthonormal basis for $\boldsymbol{R}(A)$ is $\boldsymbol{b}^{(r)} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Then,

$$\boldsymbol{x}^{(r)} = (\boldsymbol{x} \cdot \boldsymbol{b}^{(r)})\boldsymbol{b}^{(r)} = 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

– Orthonormal basis for $\boldsymbol{N}(A)$ is $\boldsymbol{b}^{(n)} = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \end{bmatrix}$. Then,

$$\boldsymbol{x}^{(n)} = (\boldsymbol{x} \cdot \boldsymbol{b}^{(n)})\boldsymbol{b}^{(n)} = - \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

4.1 How to find the projection?

- (Problem 4.1B(d)) Find the projections of $\boldsymbol{x} = \begin{bmatrix} 6 \\ 4 \\ 5 \end{bmatrix}$ onto $\boldsymbol{R}(A)$ and $\boldsymbol{N}(A)$, where $A = \begin{bmatrix} 1 & -3 & -4 \end{bmatrix}$?

Answer: **Alternatively, solve coefficient $\boldsymbol{c}$ for linear system $B\boldsymbol{c} = \boldsymbol{x}$, where columns of B are bases for $\boldsymbol{N}(A)$ and $\boldsymbol{R}(A)$.**

– Basis for $\boldsymbol{R}(A)$ is $\begin{bmatrix} 1 \\ -3 \\ -4 \end{bmatrix}$.

– Basis for $\boldsymbol{N}(A)$ is $\begin{bmatrix} 3 & 4 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$.

– Then, Gauss-Jordan gives

$$\begin{bmatrix} 3 & 4 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & -4 \end{bmatrix} \boldsymbol{c} = \begin{bmatrix} 6 \\ 4 \\ 5 \end{bmatrix} \Rightarrow \boldsymbol{c} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \Rightarrow \begin{cases} \boldsymbol{x}^{(n)} = \begin{bmatrix} 3 & 4 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \\ 1 \end{bmatrix} \\ \boldsymbol{x}^{(r)} = \begin{bmatrix} 1 \\ -3 \\ -4 \end{bmatrix} \begin{bmatrix} -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix} \end{cases}$$

4.2 Projections

Definition (Projection matrix): For a linear subspace $\mathbb{V}$, we can project any vector $\boldsymbol{b}$ onto this subspace.

Such a projection $\boldsymbol{p}$ can be obtained through *projection matrix* P, i.e.,

$$\boldsymbol{p} = P\boldsymbol{b}.$$

- If $\boldsymbol{b} \in \mathbb{V}$, then its projection to $\mathbb{V}$ is itself.
- Hence, $P\boldsymbol{p} = \boldsymbol{p}$ and $P^2\boldsymbol{p} = P\boldsymbol{p} = \boldsymbol{p}$ and $\ldots$ and $P^n\boldsymbol{p} = \boldsymbol{p}$.
- Let $\boldsymbol{b}_i$ be a vector with all components being 0 except the ith component, which is 1. Then, we have

$$P(P\boldsymbol{b}_i) = P\boldsymbol{b}_i.$$

  By collecting all the $\boldsymbol{b}_i$ to form the identity matrix, we obtain

$$P(PI) = PI \quad \Longrightarrow \quad P^2 = P.$$

  The projection matrix P satisfies $P^2 = P$.

4.2 Projections

Examples of the projection matrices.

- Onto the the z-axis, $P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

- Onto the the xy-plane, $P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Since z-axis and xy-plane are orthogonal complement, we have

$$P_1\boldsymbol{b} + P_2\boldsymbol{b} = \boldsymbol{b} = I\boldsymbol{b} \;\Rightarrow\; (P_1 + P_2 - I)\boldsymbol{b} = \mathbf{0} \text{ for every } \boldsymbol{b}.$$

$\Rightarrow$ $P_1 + P_2 = I$ for projection matrices of orthogonal complement subspaces

4.2 Orthogonality and projections

- $\boldsymbol{p}$ is the projection of $\boldsymbol{b}$ onto a subspace $\mathbb{V}$.

  Then, $\boldsymbol{p}$ and $\boldsymbol{b} - \boldsymbol{p}$ are orthogonal.

Proof:

- Let $\mathbb{V}^{\perp}$ be the orthogonal complement of $\mathbb{V}$.
- Let P and $P^{\perp}$ be the projection matrices of $\mathbb{V}$ and $\mathbb{V}^{\perp}$, respectively.
- Then,

$$
\begin{aligned}
& P\boldsymbol{b} + P^{\perp}\boldsymbol{b} = \boldsymbol{p} + \boldsymbol{p}^{\perp} = \boldsymbol{b} \\
\Rightarrow\ & \boldsymbol{p}^{\perp} = \boldsymbol{b} - \boldsymbol{p} \\
\Rightarrow\ & \boldsymbol{p}^{\perp} \cdot \boldsymbol{p} = \boldsymbol{0} \text{ by definition of orthogonal complement}
\end{aligned}
$$

4.2 How to determine the projection based on orthogonal vectors?

Suppose the subspace $\mathbb{V}$ has rank r and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r$ are linearly independent vectors that span $\mathbb{V}$.

$\Rightarrow$ projection $\boldsymbol{p}$ (of $\boldsymbol{b}$) can be written as:

$$\boldsymbol{p} = \hat{x}_1\boldsymbol{a}_1 + \cdots + \hat{x}_r\boldsymbol{a}_r = \underbrace{\begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_r \end{bmatrix}}_{=A} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_r \end{bmatrix} = A\hat{\boldsymbol{x}}.$$

$\Rightarrow$ $(\boldsymbol{b} - \boldsymbol{p}) \in \mathbb{V}^{\perp}$ gives

$$\begin{cases} \boldsymbol{a}_1^{\mathsf{T}}(\boldsymbol{b} - \boldsymbol{p}) = 0 \\ \boldsymbol{a}_2^{\mathsf{T}}(\boldsymbol{b} - \boldsymbol{p}) = 0 \\ \vdots \\ \boldsymbol{a}_r^{\mathsf{T}}(\boldsymbol{b} - \boldsymbol{p}) = 0 \end{cases} \equiv \; A^{\mathsf{T}}(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = \boldsymbol{0} \;\Longrightarrow\; A^{\mathsf{T}}A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b}$$

This concludes (if $A^{\mathsf{T}}A$ is invertible):

$$\boxed{\hat{\boldsymbol{x}} = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\boldsymbol{b} \quad \text{and} \quad \boldsymbol{p} = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\boldsymbol{b} \quad \text{and} \quad P = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}.}$$

4.2 How to determine the projection based on orthogonal vectors?

Example. Find the projection of $\boldsymbol{b}$ onto a one-dimensional subspace $\mathbb{V}$ with $\boldsymbol{a} \in \mathbb{V}$ and $\boldsymbol{a} \neq \boldsymbol{0}$.

Answer: We know $A = \begin{bmatrix} \boldsymbol{a} \end{bmatrix}$. Hence,

$$\boldsymbol{p} = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\boldsymbol{b} = \boldsymbol{a}(\boldsymbol{a}^{\mathsf{T}}\boldsymbol{a})^{-1}\boldsymbol{a}^{\mathsf{T}}\boldsymbol{b} = \left(\frac{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{a}}\right)\boldsymbol{a} = \left(\frac{\boldsymbol{a} \cdot \boldsymbol{b}}{\boldsymbol{a} \cdot \boldsymbol{a}}\right)\boldsymbol{a}$$

$\square$

- $\boxed{P = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \text{ implies } P^{\mathsf{T}} = P.}$

  The projection matrix onto a space $\mathbb{V}$ can always be represented as the form $A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$, where the columns of A span $\mathbb{V}$. So, $P^{\mathsf{T}} = P$ is always true for a projection matrix.

4.2 Projection matrix and its column space

Tip: From slide 3-64, $P_{m\times m} = B_{m\times r}C_{r\times m}$ (and the rank of all matrices is r). Then,

$$\boldsymbol{C}(P_{m\times m}) = \boldsymbol{C}(B_{m\times r}) \quad \text{and} \quad \boldsymbol{R}(P_{m\times m}) = \boldsymbol{R}(C_{r\times m}).$$

Hence,

$\{\boldsymbol{p} : \boldsymbol{p} = P\boldsymbol{b} \text{ for some } \boldsymbol{b} \in \Re^m\}$ is the column space of $P = A(A^\mathsf{T}A)^{-1}A^\mathsf{T}$ and also A.

and

$\left\{\boldsymbol{p} : \boldsymbol{p} = P^\perp\boldsymbol{b} = (I - P)\boldsymbol{b} \text{ for some } \boldsymbol{b} \in \Re^m\right\}$ is the left nullspace space of $P = A(A^\mathsf{T}A)^{-1}A^\mathsf{T}$ and also A.

What if $A^\mathsf{T}A$ is not invertible?

4.2 How to determine the projection onto column space?

Suppose $\boldsymbol{C}(A_{m\times n})$ has rank r.

Note that if $r < n$, then $(A^{\mathsf{T}}A)_{n\times n}$ is not invertible!

This gives:

$$\hat{\boldsymbol{x}} = (\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}}\boldsymbol{b} \quad \text{and} \quad \boldsymbol{p} = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}}\boldsymbol{b} \quad \text{and} \quad P = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}}.$$

where $\underline{A}$ is formed by the r linearly independent (usually, pivot) columns of A, and the matrix P projects $\boldsymbol{b}$ to the column space of A (equivalently, $\underline{A}$).

4.2 How to determine the projection onto column space?

Example. $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Find the projection matrix onto $\boldsymbol{C}(A)$.

Answer:

- First, we note the column space of A is the xy-plane.
- Taking $\underline{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ gives $P = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.
- Taking $\underline{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$ gives $P = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.
- Taking $\underline{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$ gives $P = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

$\square$

4.2 How to determine the projection onto column space?

Example. $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Find $\hat{\boldsymbol{x}}$ that minimizes $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2$.

Answer:

- $\underline{A} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \implies P = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$.

- Then, $A\hat{\boldsymbol{x}} = P\boldsymbol{b} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$, i.e., $\hat{x}_1 + \hat{x}_2 = \frac{1}{3}$. $\square$

(Problem 21, Section 4.3) Continue from the previous example. Which of the four spaces contains the error vector $\boldsymbol{e}$ $(= \boldsymbol{b} - A\hat{\boldsymbol{x}})$? Which contains $\boldsymbol{p}$ $(= \boldsymbol{b} - \boldsymbol{e})$? Which contains $\hat{\boldsymbol{x}}$? What is the nullspace of A?

Answer: It can be derived that

$$\begin{cases} \boldsymbol{C}(A) = \{\boldsymbol{b} \in \Re^3 : b_1 = b_2 = b_3\} \\ \boldsymbol{R}(A) = \{\boldsymbol{b} \in \Re^2 : b_1 = b_2\} \\ \boldsymbol{N}(A) = \{\boldsymbol{x} \in \Re^2 : x_1 + x_2 = 0\} \\ \boldsymbol{N}(A^{\mathrm{T}}) = \{\boldsymbol{x} \in \Re^3 : x_1 + x_2 + x_3 = 0\} \end{cases}$$

So, $\boldsymbol{e} \in \boldsymbol{N}(A^{\mathrm{T}})$, $\boldsymbol{p} \in \boldsymbol{C}(A)$, one of $\hat{\boldsymbol{x}}$ in $\boldsymbol{R}(A)$ (but $\hat{\boldsymbol{x}}$ can never in $\boldsymbol{N}(A)$ since $A\hat{\boldsymbol{x}} = P\boldsymbol{b} \neq \boldsymbol{0}$), and $\boldsymbol{N}(A)$ is a line in $\Re^2$. $\square$

- **Final note.** When $n = r$, the answer reduces to what has been provided in the textbook, i.e., $\hat{\boldsymbol{x}}$ in $\boldsymbol{R}(A) = \Re^n$, and $\boldsymbol{N}(A) = \{\boldsymbol{0}\}$.

4.3 Least square approximations

Question: Find the solution of $A\hat{\boldsymbol{x}} = \boldsymbol{b}$ given that $\boldsymbol{b} \in \boldsymbol{C}(A)$.

Answer:

- The solution can be found using methods taught before (e.g., Gauss-Jordan).
- $\boldsymbol{e} = \boldsymbol{b} - A\hat{\boldsymbol{x}}$ satisfies $\|\boldsymbol{e}\|^2 = 0$.

Question: Find $\boldsymbol{x}$ such that $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2$ is minimized (where $\boldsymbol{b}$ is not necessarily in $\boldsymbol{C}(A)$.)

Answer:

- This is to find $\hat{\boldsymbol{x}}$ such that $A\hat{\boldsymbol{x}} = P\boldsymbol{b}$, where P is the projection matrix of $\boldsymbol{C}(A)$. The solution will minimize $\|\boldsymbol{e}\|^2 = \|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2$.
- If A is invertible, then $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. So, the problem is equivalent to:

$$A\hat{\boldsymbol{x}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b} \Leftrightarrow \boxed{A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}}$$

- If A is not invertible, then $P = \underline{A}(\underline{A}^{\mathrm{T}}\underline{A})^{-1}\underline{A}^{\mathrm{T}}$. So, the problem is equivalent to:

$$A\hat{\boldsymbol{x}} = \underline{A}(\underline{A}^{\mathrm{T}}\underline{A})^{-1}\underline{A}^{\mathrm{T}}\boldsymbol{b} \Leftrightarrow \boxed{\underline{A}^{\mathrm{T}}A\hat{\boldsymbol{x}} = \underline{A}^{\mathrm{T}}\boldsymbol{b}} \Leftrightarrow \boxed{A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}}$$

Example. $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Find $\hat{\boldsymbol{x}}$ that minimizes $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2$.

Answer:

- $\underline{A} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \implies P = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$.
- Hence,
  - $\boxed{A\hat{\boldsymbol{x}} = \underline{A}(\underline{A}^{\mathsf{T}}\underline{A})^{-1}\underline{A}^{\mathsf{T}}\boldsymbol{b}}$ is equivalent to $\hat{x}_1 + \hat{x}_2 = \frac{1}{3}$.
  - $\boxed{\underline{A}^{\mathsf{T}}A\hat{\boldsymbol{x}} = \underline{A}^{\mathsf{T}}\boldsymbol{b}}$ is equivalent to $3\hat{x}_1 + 3\hat{x}_2 = 1$.
  - $\boxed{A^{\mathsf{T}}A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b}}$ is equivalent to $3\hat{x}_1 + 3\hat{x}_2 = 1$.

So the above three systems are equivalent. $\square$

4.3 Least square approximations

Example. $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Find $\hat{\boldsymbol{x}}$ that minimizes $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2$.

Answer:

- $\underline{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \implies P = \underline{A}(\underline{A}^{\mathrm{T}}\underline{A})^{-1}\underline{A}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
- Hence,
  - $\boxed{A\hat{\boldsymbol{x}} = \underline{A}(\underline{A}^{\mathrm{T}}\underline{A})^{-1}\underline{A}^{\mathrm{T}}\boldsymbol{b}}$ is equivalent to $\begin{cases} \hat{x}_1 + \hat{x}_2 = 0 \\ \hat{x}_3 = 1 \end{cases}$.
  - $\boxed{\underline{A}^{\mathrm{T}}A\hat{\boldsymbol{x}} = \underline{A}^{\mathrm{T}}\boldsymbol{b}}$ is equivalent to $\begin{cases} \hat{x}_1 + \hat{x}_2 = 0 \\ \hat{x}_3 = 1 \end{cases}$.
  - $\boxed{A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}}$ is equivalent to $\begin{cases} \hat{x}_1 + \hat{x}_2 = 0 \\ \hat{x}_3 = 1 \end{cases}$.

So the above three systems are equivalent. □

Normal equations $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ gives us the **least square approximation**.

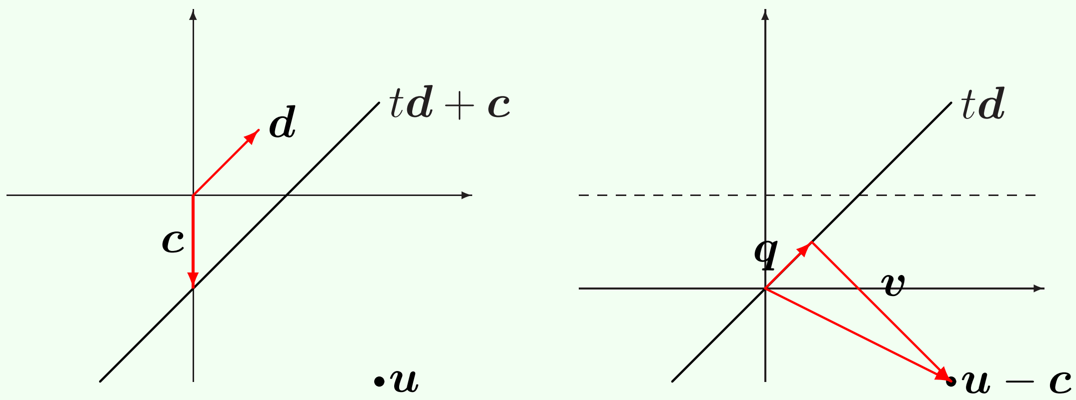
Applications of least square approximation

- *Fitting a straight line:* To find the line that minimizes square errors to a group of points.

  – A line (excluding $x_1 = \text{constant}$) on $\Re^2$ can be parameterized via t under fixed D and C as

  $$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = t \begin{bmatrix} 1 \\ D \end{bmatrix} + \begin{bmatrix} 0 \\ C \end{bmatrix} = t\boldsymbol{d} + \boldsymbol{c} = \begin{bmatrix} t \\ C + tD \end{bmatrix}.$$

  – The distance of a point $\boldsymbol{u} = \begin{bmatrix} t \\ b \end{bmatrix}$ to this line can be calculated by $\|\boldsymbol{v}\|$, where $\boldsymbol{v} = (\boldsymbol{u} - \boldsymbol{c}) - \boldsymbol{q}$, and $\boldsymbol{q}$ is the projection of $\boldsymbol{u} - \boldsymbol{c}$ onto the line $t\boldsymbol{d}$.



- So

$$\boldsymbol{q} = \boldsymbol{d}(\boldsymbol{d}^{\mathsf{T}}\boldsymbol{d})^{-1}\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{u}-\boldsymbol{c}).$$

Then

$$\begin{aligned}
\|\boldsymbol{v}\|^2 &= \|(\boldsymbol{u}-\boldsymbol{c})-\boldsymbol{q}\|^2 = \|(\boldsymbol{u}-\boldsymbol{c})-\boldsymbol{d}(\boldsymbol{d}^{\mathsf{T}}\boldsymbol{d})^{-1}\boldsymbol{d}^{\mathsf{T}}(\boldsymbol{u}-\boldsymbol{c})\|^2 \\
&= \left\| \frac{1}{1+D^2}\begin{bmatrix} D^2 & -D \\ -D & 1 \end{bmatrix} \underbrace{\begin{bmatrix} t-0 \\ b-C \end{bmatrix}}_{=(\boldsymbol{u}-\boldsymbol{c})} \right\|^2 = \left\| \underbrace{(b-C-tD)\begin{bmatrix} -\frac{D}{1+D^2} \\ \frac{1}{1+D^2} \end{bmatrix}}_{=\boldsymbol{v}} \right\|^2 \\
&= \frac{(b-C-tD)^2}{1+D^2} = \frac{1}{1+D^2}\left\| b - \begin{bmatrix} 1 & t \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} \right\|^2
\end{aligned}$$

- We wish to determine $\boldsymbol{d}$ and $\boldsymbol{c}$ such that for $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$,

$$\sum_{i=1}^{k}\|\boldsymbol{v}_i\|^2 = \frac{1}{1+D^2}\left\| \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}}_{\boldsymbol{b}} - \underbrace{\begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_k \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\hat{\boldsymbol{x}}} \right\|^2$$

is minimized.

4.3 Least square approximations

(Problem 1, Section 4.3) Find the best straight line that fits points

$$\begin{bmatrix} t \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \end{bmatrix} \text{ and } \begin{bmatrix} 4 \\ 20 \end{bmatrix}.$$

Give the errors for each point.

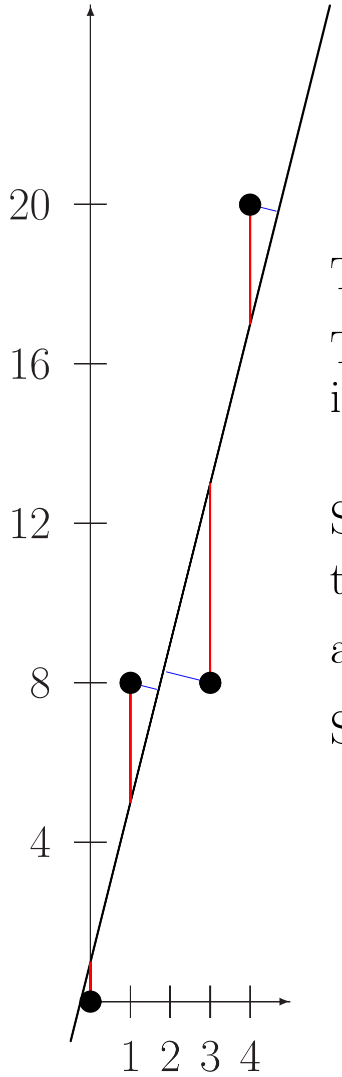
Answer:

- $\sum_{i=1}^{4} \|\boldsymbol{v}_i\|^2 = \frac{1}{1+D^2} \left\| \underbrace{\begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}}_{\boldsymbol{b}} - \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} C \\ D \end{bmatrix}}_{\hat{\boldsymbol{x}}} \right\|^2$

- Solving $A^{\mathsf{T}} A \hat{\boldsymbol{x}} = A^{\mathsf{T}} \boldsymbol{b}$, i.e., $\begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix} \hat{\boldsymbol{x}} = \begin{bmatrix} 36 \\ 112 \end{bmatrix}$, yields $\hat{\boldsymbol{x}} = \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$.

- $\begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \end{bmatrix} = \begin{bmatrix} -\frac{D}{1+D^2} \\ \frac{1}{1+D^2} \end{bmatrix} (\boldsymbol{b} - A\hat{\boldsymbol{x}})^{\mathsf{T}} = \begin{bmatrix} -\frac{4}{1+4^2} \\ \frac{1}{1+4^2} \end{bmatrix} (\boldsymbol{b} - A\hat{\boldsymbol{x}})^{\mathsf{T}} = \begin{bmatrix} -\frac{4}{17} \\ \frac{1}{17} \end{bmatrix} \begin{bmatrix} -1 & 3 & -5 & 3 \end{bmatrix}$.

4.3 Least square approximations



The red color line segment is $b - (C + tD)$ (the direct difference).

The blue color line segment is the true error,
i.e., the projection of the direct difference
onto the vector perpendicular to the line $C + tD$.

Since the minimization of the two yields the same result,
the textbook defines the error to be $\boxed{e_i = b_i - (C + t_i D)}$,
and the projection to be $\boxed{p_i = C + t_i D}$.

See Figure 4.9 on page 227 or the next page.

(Problem 1, Section 4.3) With $b = 0, 8, 8, 0$ at $t = 0, 1, 3, 4$, set up and solve the normal equations $A^{\mathsf{T}} A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b}$. For the best straight line in Figure 4.9a, find its four heights p_i and four errors e_i. What is the minimum value $E = e_1^2+e_2^2+e_3^2+e_4^2$.

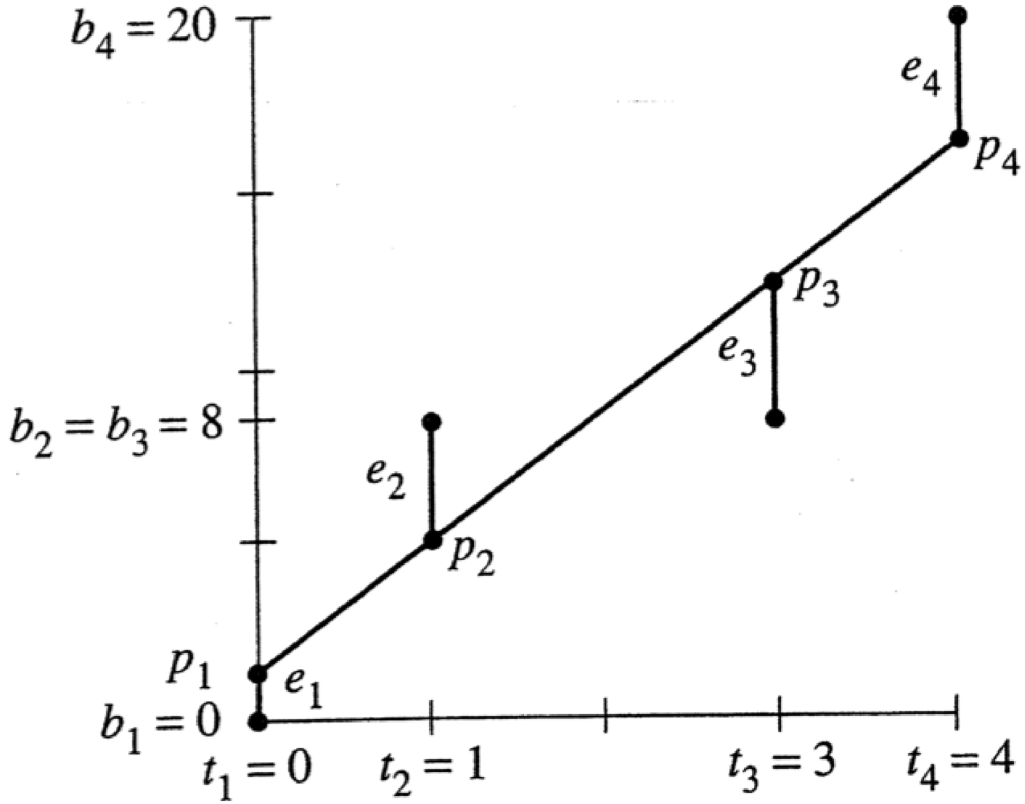


Figure 4.9:

- Do not confuse with that $\boldsymbol{b} - A\hat{\boldsymbol{x}} = \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ -5 \\ 3 \end{bmatrix}$ is perpendicular to the columns of $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$, i.e., $\boldsymbol{e}$ is perpendicular to the column space of A.
- In the above, $\|\boldsymbol{e}\|$ is the shortest distance of vector $\boldsymbol{b}$ to the column space of A.

4.3 Least square approximations

- For simplicity, in this textbook, we will directly take the points on the target equation and find the best fit curve.

  Example. Find the best straight $b = C$ (respectively, $b = C + tD + t^2E$) that fits points

$$\begin{bmatrix} t \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \end{bmatrix} \text{ and } \begin{bmatrix} 4 \\ 20 \end{bmatrix}.$$

  Answer. Form the equations as follows.

$$\begin{cases} 0 \ = C|_{t=0} \\ 8 \ = C|_{t=1} \\ 8 \ = C|_{t=3} \\ 20 = C|_{t=4} \end{cases} \implies \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [C] = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}$$

  respectively,

$$\begin{cases} 0 \ = C + 0 \cdot D + 0 \cdot E \\ 8 \ = C + 1 \cdot D + 1 \cdot E \\ 8 \ = C + 3 \cdot D + 9 \cdot E \\ 20 = C + 4 \cdot D + 16 \cdot E \end{cases} \implies \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}$$

$\square$

4.3 Least square approximations

Question: Find $\boldsymbol{x}$ such that $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2$ is minimized (where $\boldsymbol{b}$ is not necessarily in $\boldsymbol{C}(A)$.)

Answer:

- This is to find $\hat{\boldsymbol{x}}$ satisfying the normal equations $\boxed{A^{\mathsf{T}} A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b}}$.

Calculus interpretation: Suppose all components are real.

- We wish to minimize $E \triangleq \|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2 = \sum_{i=1}^{m}(b_i - \begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,n}\end{bmatrix}\hat{\boldsymbol{x}})^2$; hence, the solution satisfies

$$\begin{aligned}\frac{\partial E}{\partial \hat{x}_j} &= \sum_{i=1}^{m} 2(-a_{i,j})(b_i - \begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,n}\end{bmatrix}\hat{\boldsymbol{x}}) \\ &= -2\begin{bmatrix} a_{1,j} & a_{2,j} & \cdots & a_{m,j}\end{bmatrix}(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = 0.\end{aligned}$$

  I.e., $\boxed{A^{\mathsf{T}}(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = 0}$, which coincides with the **normal equations**.

Summary: The derivative of $\|\boldsymbol{b} - A\boldsymbol{x}\|^2$ with respect to $\boldsymbol{x}$ is $-2A^{\mathsf{T}}(\boldsymbol{b} - A\boldsymbol{x})$.

Example 4.3B. Find the parabola $C + Dt + Et^2$ that comes closet (least square error) to five points $\begin{bmatrix} t \\ b \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$.

Example.

- Write down the five equations:

$$\begin{cases} C + D \cdot (-2) + E(-2)^2 & = 0 \\ C + D \cdot (-1) + E(-1)^2 & = 0 \\ C + D \cdot (0) + E(0)^2 & = 1 \\ C + D \cdot (1) + E(1)^2 & = 0 \\ C + D \cdot (2) + E(2)^2 & = 0 \end{cases} \Longrightarrow \underbrace{\begin{bmatrix} 1 & (-2) & (-2)^2 \\ 1 & (-1) & (-1)^2 \\ 1 & (0) & (0)^2 \\ 1 & (1) & (1)^2 \\ 1 & (2) & (2)^2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} C \\ D \\ E \end{bmatrix}}_{\hat{\boldsymbol{x}}} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{b}}$$

- Since $A\hat{\boldsymbol{x}} = \boldsymbol{b}$ has no solution, we turn to the normal equations $A^{\mathsf{T}}A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b}$. The solution will minimize $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\|^2$. $\square$

4.4 Orthogonal bases and Gram-Schmidt

- The problem of solving $A\boldsymbol{x} = \boldsymbol{b}$ (or minimizing $\|\boldsymbol{b} - A\boldsymbol{x}\|^2$) will become easy if the **columns** of A are **orthogonal**.
- The reason for the above claim is obvious since $A^{\mathsf{T}}A$ becomes **diagonal** matrix! The problem then reduces to:

$$\boxed{A^{\mathsf{T}}A\boldsymbol{x} = \boldsymbol{x} = A^{\mathsf{T}}\boldsymbol{b} \quad \text{if} \quad A^{\mathsf{T}}A = I}.$$

Definition (Orthonormal basis): The vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ are **orthonormal** if

$$\boldsymbol{q}_i^{\mathsf{T}}\boldsymbol{q}_j = \begin{cases} 0 & \text{when } i \neq j \ (\textbf{orthogonality}) \\ 1 & \text{when } i = j \ (\textbf{unit vector}) \end{cases}$$

- $Q = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix}$ is sometimes called *orthonormal matrix.*

Example (Orthonormal matrix). Every permutation matrix (e.g., $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$) is a orthonormal matrix. □

An important fact about the square orthonormal basis.

- The square orthonormal matrix Q satisfies $Q^{\mathsf{T}} = Q^{-1}$.

  Proof: A direct consequence of $Q^{\mathsf{T}}Q = I$. $\square$

Example (Reflection matrix). Another interesting example for the orthonormal matrix Q is when

$$\boxed{\begin{aligned} Q &= I - 2\boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} \\ &= (I - \boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}) - \boldsymbol{q}\boldsymbol{q}^{\mathsf{T}} \\ &= P^{\perp} - P \end{aligned}}$$

Tip: Why is it called a **reflection matrix**?

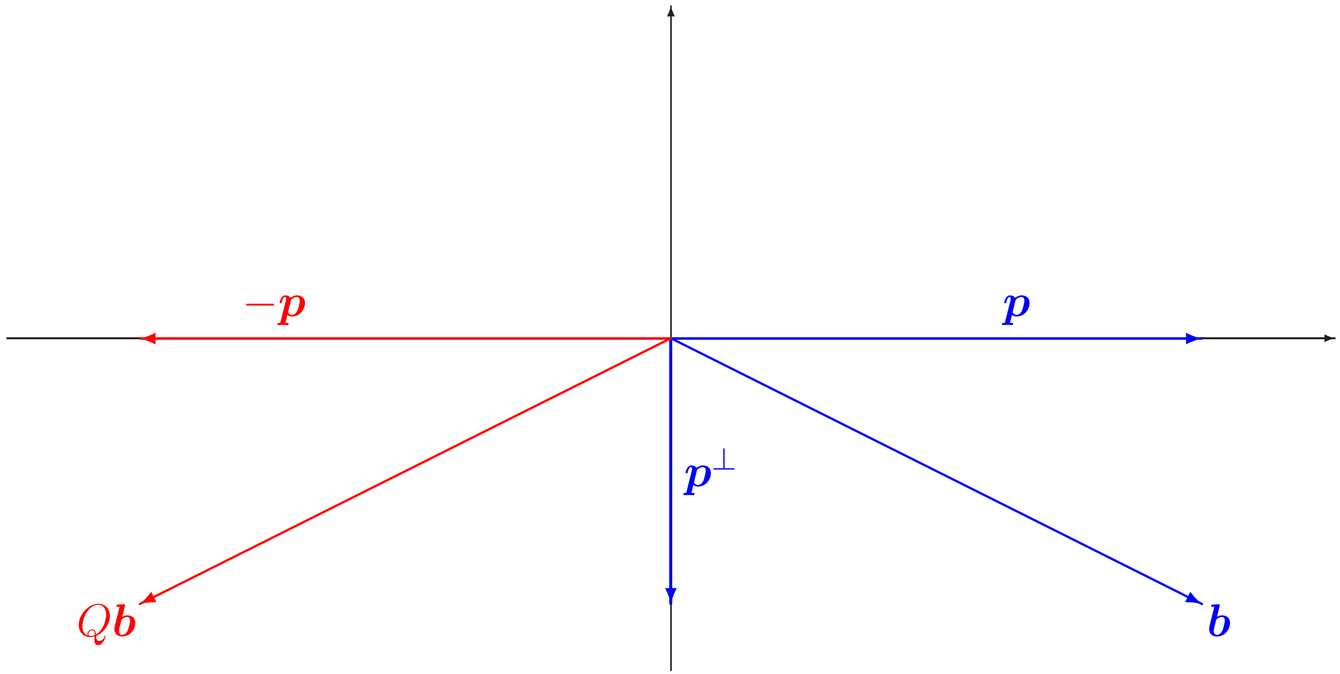
- The projection matrix on to a subspace spanned by a unit vector $\boldsymbol{q}$ is

$$P = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} = \boldsymbol{q}(\boldsymbol{q}^{\mathsf{T}}\boldsymbol{q})\boldsymbol{q}^{\mathsf{T}} = \boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}.$$

- The project matrix $P^{\perp}$ on the the orthogonal complement of the space spanned by $\boldsymbol{q}$ is

$$P^{\perp} = I - \boldsymbol{q}\boldsymbol{q}^{\mathsf{T}}.$$

- So, the reflection matrix $Q\boldsymbol{b} = P^{\perp}\boldsymbol{b} - P\boldsymbol{b} = \boldsymbol{p}^{\perp} - \boldsymbol{p}$; hence, $Q\boldsymbol{b}$ is the "mirror" of $\boldsymbol{b}$ with respect to the orthogonal complement of the subspace spanned by $\boldsymbol{q}$.



4.4 Properties of orthonormal matrix mapping

1. **Length invariance**: $\boxed{\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|}$
2. **Angle invariance**: $\boxed{(Q\boldsymbol{x})^{\mathsf{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}}$

   Proof: We only need to prove the second property, which can be easily seen by
   $$(Q\boldsymbol{x})^{\mathsf{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathsf{T}}Q^{\mathsf{T}}Q\boldsymbol{y} = \boldsymbol{x}^{\mathsf{T}}I\boldsymbol{y} = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}.$$
   $\square$

With A replaced by Q, we have

$$\hat{\boldsymbol{x}} = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\boldsymbol{b} = Q^{\mathsf{T}}\boldsymbol{b} \quad \text{and} \quad \boldsymbol{p} = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\boldsymbol{b} = QQ^{\mathsf{T}}\boldsymbol{b}$$
$$\text{and} \quad P = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} = QQ^{\mathsf{T}}.$$

Further assuming that Q is a square matrix, we have $QQ^{\mathsf{T}} = I$ and

$$\hat{\boldsymbol{x}} = Q^{\mathsf{T}}\boldsymbol{b} = \begin{bmatrix} \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b} \\ \vdots \\ \boldsymbol{q}_n^{\mathsf{T}}\boldsymbol{b} \end{bmatrix} \quad \text{and} \quad \boldsymbol{p} = QQ^{\mathsf{T}}\boldsymbol{b} = \boldsymbol{b} \quad \text{and} \quad P = QQ^{\mathsf{T}} = I.$$

Example. $Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$ rotates $\boldsymbol{b}$ counterclockwisely by degree θ.

4.4 Gram-Schmidt procedure

Observations

- Under the assumption that Q is not necessarily a square matrix (and hence, possibly, $\boldsymbol{p} \neq \boldsymbol{b}$) the error

$$\begin{aligned} \boldsymbol{e} &= \boldsymbol{b} - Q\hat{\boldsymbol{x}} = \boldsymbol{b} - \hat{x}_1\boldsymbol{q}_1 - \hat{x}_2\boldsymbol{q}_2 - \cdots - \hat{x}_n\boldsymbol{q}_n \\ &= \boldsymbol{b} - (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b})\boldsymbol{q}_1 - (\boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b})\boldsymbol{q}_2 - \cdots - (\boldsymbol{q}_n^{\mathsf{T}}\boldsymbol{b})\boldsymbol{q}_n \\ &= \boldsymbol{b} - \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}) - \boldsymbol{q}_2(\boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b}) - \cdots - \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathsf{T}}\boldsymbol{b}) \\ &= \boldsymbol{b} - \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathsf{T}} \\ \boldsymbol{q}_2^{\mathsf{T}} \\ \vdots \\ \boldsymbol{q}_n^{\mathsf{T}} \end{bmatrix} \boldsymbol{b} = \boldsymbol{b} - Q_{m\times n}Q_{n\times m}^{\mathsf{T}}\boldsymbol{b} = \left(I_{m\times m} - Q_{m\times n}Q_{n\times m}^{\mathsf{T}}\right)\boldsymbol{b} \end{aligned}$$

  is perpendicular to $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n$ (i.e., the column space of Q).

- If $\|\boldsymbol{e}\| \neq 0$, then $\boldsymbol{b}$ cannot be **completely** represented by Q.
  - This means that $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ is not enough to span the **space** of interest.
  - We can then add a new vector $\boldsymbol{q}_{n+1} = \frac{\boldsymbol{e}}{\|\boldsymbol{e}\|}$ to complete the **basis** for this **space**.

- This leads to the basic idea of the **Gram-Schmidt procedure**.

4.4 Gram-Schmidt procedure

Gram-Schmidt procedure is used to identify the dimension spanned by a sequence of vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$ and find an orthonormal basis for this space.

(Modified) Gram-Schmidt procedure

Initialization: A sequence of (non-zero) vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$.

1. $\boldsymbol{q}_1 = \frac{\boldsymbol{e}_1}{\|\boldsymbol{e}_1\|}$, where $\boxed{\boldsymbol{e}_1 = \boldsymbol{b}_1}$.

2. $\boldsymbol{q}_2 = \frac{\boldsymbol{e}_2}{\|\boldsymbol{e}_2\|}$, where $\boxed{\boldsymbol{e}_2 = \boldsymbol{b}_2 - (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_2)\boldsymbol{q}_1}$, if $\|\boldsymbol{e}_2\|$ is not zero.

3. $\boldsymbol{q}_3 = \frac{\boldsymbol{e}_3}{\|\boldsymbol{e}_3\|}$, where $\boxed{\boldsymbol{e}_3 = \boldsymbol{b}_3 - (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_3)\boldsymbol{q}_1 - (\boldsymbol{q}_2^\mathsf{T}\boldsymbol{b}_3)\boldsymbol{q}_2} = \boldsymbol{b}_3 - \boldsymbol{q}_1(\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_3) - \boldsymbol{q}_2(\boldsymbol{q}_2^\mathsf{T}\boldsymbol{b}_3) = \boldsymbol{b}_3 - \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{q}_1^\mathsf{T} \\ \boldsymbol{q}_2^\mathsf{T}\end{bmatrix}\boldsymbol{b}_3 = \left(I_{m\times m} - Q_{m\times 2}Q_{2\times m}^\mathsf{T}\right)\boldsymbol{b}_3$, if $\|\boldsymbol{e}_3\|$ is not zero.

4. $\cdots$

5. Repeat the above procedure until $\boldsymbol{b}_n$ is examined.

Note that by the basic property of projections, $\boldsymbol{e}_2$ is perpendicular to $\boldsymbol{q}_1$, and $\boldsymbol{e}_3$ is perpendicular to $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$, and etc. See the first observation in the previous slide.

4.4 QR decomposition

- Gram-Schmidt procedure actually changes $A = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \end{bmatrix}$ into QR.

$$
\begin{aligned}
A &= \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_3 & \cdots \end{bmatrix}_{m\times n} \\
&= \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 + (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_2)\boldsymbol{q}_1 & \boldsymbol{e}_3 + (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_3)\boldsymbol{q}_1 + (\boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b}_3)\boldsymbol{q}_2 & \cdots \end{bmatrix}_{m\times n} \\
&= \begin{bmatrix} \|\boldsymbol{e}_1\|\boldsymbol{q}_1 & \|\boldsymbol{e}_2\|\boldsymbol{q}_2 + (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_2)\boldsymbol{q}_1 & \|\boldsymbol{e}_3\|\boldsymbol{q}_3 + (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_3)\boldsymbol{q}_1 + (\boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b}_3)\boldsymbol{q}_2 & \cdots \end{bmatrix}_{m\times n} \\
&= \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 & \cdots \end{bmatrix}_{m\times r} \begin{bmatrix} \|\boldsymbol{e}_1\| & \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_2 & \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_3 & \cdots \\ 0 & \|\boldsymbol{e}_2\| & \boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b}_3 & \cdots \\ 0 & 0 & \|\boldsymbol{e}_3\| & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}_{r\times n} \\
&= {\color{red}\begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 & \cdots \end{bmatrix}_{m\times r} \begin{bmatrix} \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_1 & \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_2 & \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_3 & \cdots \\ 0 & \boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b}_2 & \boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b}_3 & \cdots \\ 0 & 0 & \boldsymbol{q}_3^{\mathsf{T}}\boldsymbol{b}_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}_{r\times n}}
\end{aligned}
$$

where the last red-colored step follows, for example,

$$
\|\boldsymbol{e}_3\| = (\|\boldsymbol{e}_3\|\boldsymbol{q}_3^{\mathsf{T}})\boldsymbol{q}_3 = \boldsymbol{e}_3^{\mathsf{T}}\boldsymbol{q}_3 = (\boldsymbol{b}_3^{\mathsf{T}} - (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_3)\boldsymbol{q}_1^{\mathsf{T}} - (\boldsymbol{q}_2^{\mathsf{T}}\boldsymbol{b}_3)\boldsymbol{q}_2^{\mathsf{T}})\boldsymbol{q}_3 = \boldsymbol{b}_3^{\mathsf{T}}\boldsymbol{q}_3.
$$

4.4 QR decomposition

- Can we decompose or factorize A into QR such that Q is an orthonormal matrix and R in an upper triangular matrix?

Answer: Yes, by Gram-Schmidt procedure.

Notably, the upper triangular R is often taken to be a square matrix when $m \geq n$ in, e.g., MATLAB. Hence,

$$A_{m\times n} = Q_{m\times n} R_{n\times n}.$$

What if $r < n$? Suppose $r = 2 < n = 3$, and

$$\begin{cases} \boldsymbol{e}_1 = (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_1)\boldsymbol{q}_1 = \boldsymbol{b}_1 \neq \boldsymbol{0} \Rightarrow \boldsymbol{q}_1 = \boldsymbol{e}_1/\|\boldsymbol{e}_1\| \\ \boldsymbol{e}_2 = (\boldsymbol{q}_3^\mathsf{T}\boldsymbol{b}_2)\boldsymbol{q}_2 = \boldsymbol{b}_2 - (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_2)\boldsymbol{q}_1 = \boldsymbol{0} \Rightarrow \boldsymbol{q}_2 \text{ ??} \\ \boldsymbol{e}_3 = (\boldsymbol{q}_3^\mathsf{T}\boldsymbol{b}_3)\boldsymbol{q}_3 = \boldsymbol{b}_3 - (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_3)\boldsymbol{q}_1 \neq \boldsymbol{0} \Rightarrow \boldsymbol{q}_3 = \boldsymbol{e}_3/\|\boldsymbol{e}_3\| \end{cases} \Rightarrow \begin{cases} \boldsymbol{b}_1 = (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_1)\boldsymbol{q}_1 \\ \boldsymbol{b}_2 = (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_2)\boldsymbol{q}_1 \\ \boldsymbol{b}_3 = (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_3)\boldsymbol{q}_1 + (\boldsymbol{q}_3^\mathsf{T}\boldsymbol{b}_3)\boldsymbol{q}_3 \end{cases}$$

$$\begin{aligned} \Rightarrow A = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_3 \end{bmatrix}_{m\times 3} &= \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_3 \end{bmatrix}_{m\times 2} \begin{bmatrix} \boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_1 & \boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_2 & \boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_3 \\ 0 & 0 & \boldsymbol{q}_3^\mathsf{T}\boldsymbol{b}_3 \end{bmatrix}_{2\times 3} \\ &= \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix}_{m\times 3} \begin{bmatrix} \boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_1 & \boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_2 & \boldsymbol{q}_1^\mathsf{T}\boldsymbol{b}_3 \\ 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{q}_3^\mathsf{T}\boldsymbol{b}_3 \end{bmatrix}_{3\times 3} \end{aligned}$$

4.4 QR decomposition

Example. $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

$$\begin{cases} \boldsymbol{e}_1 = \boldsymbol{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{q}_1 = \dfrac{\boldsymbol{e}_1}{\|\boldsymbol{e}_1\|} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ \boldsymbol{e}_2 = \boldsymbol{b}_2 - (\boldsymbol{q}_1^\top \boldsymbol{b}_2)\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{q}_2 \text{ ??} \\ \boldsymbol{e}_3 = \boldsymbol{b}_3 - (\boldsymbol{q}_1^\top \boldsymbol{b}_3)\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{q}_3 = \frac{\boldsymbol{e}_3}{\|\boldsymbol{e}_3\|} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \end{cases}$$

4.4 QR decomposition

Choose $\boldsymbol{q}_2$ to be a unit vector orthogonal to both $\boldsymbol{q}_1$ and $\boldsymbol{q}_3$. Then

$$\Rightarrow A = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_3 \end{bmatrix}_{3\times 3} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_3 \end{bmatrix}_{3\times 2} \begin{bmatrix} \boldsymbol{q}_1^\intercal \boldsymbol{b}_1 & \boldsymbol{q}_1^\intercal \boldsymbol{b}_2 & \boldsymbol{q}_1^\intercal \boldsymbol{b}_3 \\ 0 & 0 & \boldsymbol{q}_3^\intercal \boldsymbol{b}_3 \end{bmatrix}_{2\times 3}$$

$$\text{(Answer 1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}_{3\times 2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}_{2\times 3}$$

$$= \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix}_{m\times 3} \begin{bmatrix} \boldsymbol{q}_1^\intercal \boldsymbol{b}_1 & \boldsymbol{q}_1^\intercal \boldsymbol{b}_2 & \boldsymbol{q}_1^\intercal \boldsymbol{b}_3 \\ 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{q}_3^\intercal \boldsymbol{b}_3 \end{bmatrix}_{3\times 3}$$

$$\text{(Answer 2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}_{3\times 3} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{3\times 3}$$

$$\text{(Answer 3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{3\times 3} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}_{3\times 3}$$

$\square$

Example. $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$

$$\begin{cases} \boldsymbol{e}_1 = \boldsymbol{b}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{q}_1 = \dfrac{\boldsymbol{e}_1}{\|\boldsymbol{e}_1\|} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \boldsymbol{e}_2 = \boldsymbol{b}_2 - (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_2)\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \boldsymbol{q}_2 \; ?? \\ \boldsymbol{e}_3 = \boldsymbol{b}_3 - (\boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_3)\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \boldsymbol{q}_3 = \frac{\boldsymbol{e}_3}{\|\boldsymbol{e}_3\|} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{cases}$$

$$\begin{aligned} \Rightarrow A = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_3 \end{bmatrix}_{2\times 3} &= \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_3 \end{bmatrix}_{2\times 2} \begin{bmatrix} \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_1 & \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_2 & \boldsymbol{q}_1^{\mathsf{T}}\boldsymbol{b}_3 \\ 0 & 0 & \boldsymbol{q}_3^{\mathsf{T}}\boldsymbol{b}_3 \end{bmatrix}_{2\times 3} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{2\times 2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}_{2\times 3} \end{aligned}$$

$\square$

4.4 QR decomposition

- In most cases $A_{m\times n}$ has rank $r = n$; hence, $A_{m\times n} = Q_{m\times n}R_{n\times n}$ and R is invertible. This simplifies

$$A^{\mathsf{T}}A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b}$$

to

$$A^{\mathsf{T}}A\hat{\boldsymbol{x}} = A^{\mathsf{T}}\boldsymbol{b} \quad \Leftrightarrow \quad R^{\mathsf{T}}Q^{\mathsf{T}}QR\hat{\boldsymbol{x}} = R^{\mathsf{T}}Q\boldsymbol{b} \quad \Leftrightarrow \quad R\hat{\boldsymbol{x}} = Q\boldsymbol{b},$$

which, by the upper triangularity of R, can be solved directly by **back substitution**!

4.4 (Modified) Gram-Schmidt procedure

The Matlab program to do the QR decomposition (or Gram-Schmidt process)

```
                               % Modified Gram-Schmidt
R=zeros(n)                     % Initialization
for j=1:n                      %
    v=A(:,j);                  % v = b_j
    for i=1:j-1                % The columns up to j-1 of Q are now q_1,...,q_{j-1}.
        R(i,j)=Q(:,i)'*v;      % r_{i,j} = q_i^T b_j  (m multiplications)  (12a)
        v=v-R(i,j)*Q(:,i)      % Perform v = v - (q_i^T b_j) q_i  (m multiplications)  (12b)
    end                        % v = e_j is now perpendicular to all q_1,...,q_{j-1}.
    R(j,j)=norm(v);            % Set diagonal entry of R.  (m multiplications)  (11)
    Q(:,j)=v/R(j,j);           % Normalization to obtain q_j
end
```

(See Problem 28) In total, we need

$$m \cdot \sum_{j=1}^{n} 1 + 2m \cdot \sum_{j=1}^{n} \sum_{i=1}^{j-1} 1 = mn + m(n-1)n = mn^2 \text{ multiplications.}$$

(Problem 28, Section 4.4) Where are the mn^2 multiplications in equaitons (11) and (12)?

A direct way to do QR-decomposition is:

```
[Q,R]=qr(A)
```

4.4 Gram-Schmidt procedure

The **"modified" Gram-Schmidt procedure** that we just introduced is numerically more stable than the **"unmodified" (original) Gram-Schmidt procedure** below (because the division of the former is only performed at the end).

Initialization: A sequence of (non-zero) vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$.

1. $\boldsymbol{c}_1 = \boldsymbol{b}_1$
2. $\boldsymbol{c}_2 = \boldsymbol{b}_2 - \dfrac{\boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{b}_2}{\boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{c}_1} \boldsymbol{c}_1 \quad \left(= \boldsymbol{b}_2 - \boldsymbol{c}_1 \dfrac{\boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{b}_2}{\boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{c}_1} = \boldsymbol{b}_2 - \boldsymbol{c}_1 (\boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{c}_1)^{-1} \boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{b}_2 \right)$
3. $\boldsymbol{c}_3 = \boldsymbol{b}_3 - \dfrac{\boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{b}_3}{\boldsymbol{c}_1^{\mathsf{T}} \boldsymbol{c}_1} \boldsymbol{c}_1 - \dfrac{\boldsymbol{c}_2^{\mathsf{T}} \boldsymbol{b}_3}{\boldsymbol{c}_2^{\mathsf{T}} \boldsymbol{c}_2} \boldsymbol{c}_2 \quad \left(= \boldsymbol{b}_3 - \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 \end{bmatrix} \left(\begin{bmatrix} \boldsymbol{c}_1^{\mathsf{T}} \\ \boldsymbol{c}_2^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{c}_1^{\mathsf{T}} \\ \boldsymbol{c}_2^{\mathsf{T}} \end{bmatrix} \boldsymbol{b}_3 \right)$
4. $\cdots$
5. Repeat the above procedure until $\boldsymbol{b}_n$ is examined.

4.4 Naming convention

- The textbook uses $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \ldots$ (uppercase letters) to denote $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3, \ldots$ obtained in the (original) Gram-Schmidt procedure in the previous slide. I personally do not like using the uppercase boldface letters to denote vectors. So I change it to $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3, \ldots$.
  - You may need to know this naming convention in order to solve those problems at the end of the section. For example, Problem 22.

(Problem 22, Section 4.4) Find orthogonal vectors $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ by Gram-Schmidt from

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{c} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}.$$

4.4 Naming convention

- (Definition from the textbook) **Orthogonal matrix**: Q is called the *orthogonal matrix* when Q is square and its columns are orthonormal bases of the column space of Q.
  - This naming convection does cause a confusion. It should be called *orthonormal square matrix*.
  - So when being asked "Whether Q^{-1} is an orthogonal matrix when Q is?" The answer should be YES according to the definition in the textbook. (See *Problem 20*.)

(Problem 20, Section 4.4) True or false (give an example in either case):
(a) Q^{-1} is an orthogonal matrix when Q is an orthogonal matrix.
(b) If Q (3 by 2) has orthonormal columns then $\|Q\boldsymbol{x}\|$ always equals to $\boldsymbol{x}$.