Columbia University
in the City of New York

Math Methods for Political Science

Lecture 6: Orthogonality and Least Squares

Thibault Vatter <tv2233@columbia.edu>
Department of Statistics, Columbia University

September 27, 2017

Outline

Inner product

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

Definition 1 (Inner product)

The **inner product** (or **dot product**) is $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \sum_{i=1}^{n} u_i v_i$.

Example 1 (Inner product)

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \implies \mathbf{u} \cdot \mathbf{v} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32.$$

Theorem 1 (Inner product)

Let $\mathbf{w} \in \mathbb{R}^n$ and $c \in \mathbb{R}$, then

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w} = \mathbf{u}\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}$
3. $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$
4. $\mathbf{u} \cdot \mathbf{u} \geq 0$ *and* $\mathbf{u} \cdot \mathbf{u} = 0 \iff \mathbf{u} = \mathbf{0}$

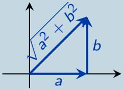
Remark: 2. and 3. $\implies \left(\sum_{i=1}^{p} c_i \mathbf{u}_i\right) \cdot \mathbf{v}_i = \sum_{i=1}^{p} c_i (\mathbf{u}_i \cdot \mathbf{v}_i)$

Length

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

Definition 2 (Length and unit vectors)

- $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ is the **length** (or **norm**) of $\mathbf{v}$.
- If $\|\mathbf{u}\| = 1$, then $\mathbf{u}$ is a **unit vector**.

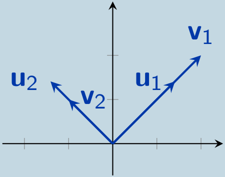


Theorem 2 (Length of scalar multiple and normalization)

- $\|c\mathbf{v}\| = |\, c \,| \, \|\mathbf{v}\|$
- $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$ *is a unit vector obtained by* **normalizing** $\mathbf{v}$.

Example 2 (Length and unit vectors)

- $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$
- $\|\mathbf{v}_1\| = \sqrt{8}$ and $\|\mathbf{v}_2\| = \sqrt{2}$
- $\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$



Distance

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

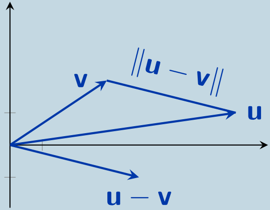
Definition 3 (Distance)

The **distance** between $\mathbf{u}$ and $\mathbf{v}$ is $\text{dist}\,(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$.

Example 3 (Distance)

- $\mathbf{u} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$
- $\mathbf{u} - \mathbf{v} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$

  $\implies \text{dist}\,(\mathbf{u}, \mathbf{v}) = \sqrt{17}$



Orthogonal vectors

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

Definition 4 (Orthogonal vectors)

$\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

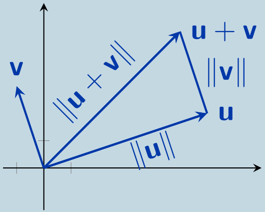
Remark: $\mathbf{0}$ is orthogonal to any $\mathbf{u}$.

Theorem 3 (The Pythagorean theorem)

$\mathbf{u}$ and $\mathbf{v}$ are orthogonal

$$\iff$$

$$\|\mathbf{u} - \mathbf{v}\| = \|\mathbf{u}\| + \|\mathbf{v}\|$$



Outline

Orthogonal complement

Let W be a subspace of $\mathbb{R}^n$.

Definition 5 (Orthogonal complement)

- $\mathbf{z} \in \mathbb{R}^n$ is **orthogonal to W** if $\mathbf{z} \cdot \mathbf{u} = 0 \ \forall \mathbf{u} \in W$.
- The **orthogonal complement** is $W^\perp = \{\mathbf{z} \in \mathbb{R}^n \mid \mathbf{z} \text{ is orthogonal to } W\}$.

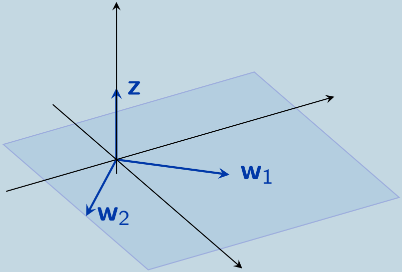
Theorem 4 (Orthogonal complement)

- *Let* $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ *be s.t. Span* $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\} = W$, *then* $\mathbf{z} \in W^\perp \iff \mathbf{z} \cdot \mathbf{v}_i = 0 \ \forall i$.
- $W^\top$ *is a subspace of* $\mathbb{R}^n$.
- *If A is an $m \times n$ matrix, then* $(Row\,A)^\perp = Nul\,A$ *and* $(Col\,A)^\perp = Nul\,A^\top$.

Orthogonal complement cont'd

Example 4 (Orthogonal complement)

Let $\mathbf{z} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{w}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$. Then if $L = \text{Span}\ \{\mathbf{z}\}$ and $W = \text{Span}\ \{\mathbf{w}_1, \mathbf{w}_2\}$, we have $L = W^\perp$ and $W = L^\perp$.



Orthogonal set and basis

Columbia University in the City of New York

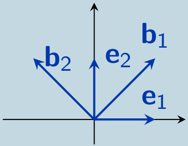
Let $\mathbf{v}_1, \cdots, \mathbf{v}_p \in \mathbb{R}^n$ and $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$.

Definition 6 (Orthogonal set and basis)

- S is an **orthogonal set** if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for $i \neq j$.
- An **orthogonal basis** is a basis that is also an orthogonal set.

Example 5 (Orthogonal set and basis)

- The standard basis is orthogonal.
- $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is another orthogonal basis of $\mathbb{R}^2$



Theorem 5 (Orthogonal set)

If $v_i \neq \mathbf{0}$ $\forall i$, then S is linearly independent.

Corollary 1 (Orthogonal basis)

If $v_i \neq \mathbf{0}$ $\forall i$, then S is an orthogonal basis for Span S.

Orthogonal set and basis

Theorem 6 (Coordinates in an orthogonal basis)

Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ *be an orthogonal basis of* $W \subseteq \mathbb{R}^n$, *then* $\forall \mathbf{y} \in W$,

$$\mathbf{y} = \sum_{i=1}^{p} c_i \mathbf{v}_i \iff c_i = \frac{\mathbf{y} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \, \forall i.$$

Example 6 (Coordinates in an orthogonal basis)

Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$.

- $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{v}_3 = \mathbf{v}_2 \cdot \mathbf{v}_3 = 0 \implies \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis of $\mathbb{R}^3$.
- $$\begin{aligned} \mathbf{y} \cdot \mathbf{v}_1 &= 11 & \mathbf{y} \cdot \mathbf{v}_1 &= -12 & \mathbf{y} \cdot \mathbf{v}_1 &= -33 \\ \mathbf{v}_1 \cdot \mathbf{v}_1 &= 11 & \mathbf{v}_2 \cdot \mathbf{v}_1 &= 6 & \mathbf{v}_3 \cdot \mathbf{v}_1 &= 33/2 \end{aligned}$$
  $$\implies \mathbf{y} = \frac{11}{11}\mathbf{v}_1 + \frac{-12}{6}\mathbf{v}_1 + \frac{-33}{33/2}\mathbf{v}_1 = \mathbf{v}_1 - 2\mathbf{v}_2 - 2\mathbf{v}_3.$$

Orthonormal set and basis

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

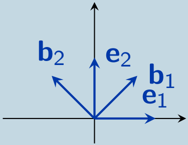
Let $\mathbf{u}_1, \cdots, \mathbf{u}_p \in \mathbb{R}^n$ and $S = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

Definition 7 (Orthonormal set and basis)

- S is an **orthonormal set** if is orthogonal and $\|\mathbf{u}_i\| = 1 \ \forall i$.
- An **orthonormal basis** is a basis that is also orthonormal.

Example 7 (Orthonormal set and basis)

- The standard basis is orthonormal.
- $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ is another orthogonal basis of $\mathbb{R}^2$



Theorem 7 (Coordinates in an orthonormal basis)

Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ be an orthonormal basis of $W \subseteq \mathbb{R}^n$, then $\forall \mathbf{y} \in W$,

$$\mathbf{y} = \sum_{i=1}^{p} c_i \mathbf{u}_i \iff c_i = \mathbf{y} \cdot \mathbf{u}_i \, \forall i.$$

Orthogonal matrices

Let U be an $m \times n$ matrix.

Theorem 8 (Orthonormal columns)

1. *U has orthonormal columns* $\iff U^\top U = I$.
2. *If U has orthonormal columns and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then*
   a) $\|U\mathbf{x}\| = \|\mathbf{x}\|$,
   b) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.

Let U be an $n \times n$ matrix.

Definition 8 (Orthogonal matrix)

U is **orthogonal** if $U^{-1} = U^\top$.

Theorem 9 (Orthogonal matrix)

The following statements are equivalents:

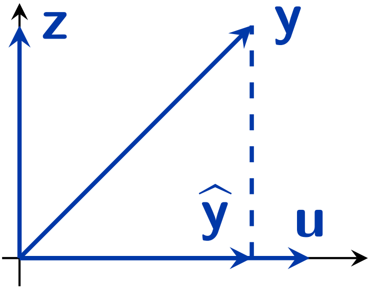
1. *U is an orthogonal matrix.*
2. *U has orthonormal columns.*
3. *U has orthonormal rows.*

Outline

The general idea

Let $\mathbf{y}, \mathbf{u} \in \mathbb{R}^n$ and assume that we want

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}, \text{ with } \begin{cases} \widehat{\mathbf{y}} = \alpha \mathbf{u} \text{ for } \alpha \in \mathbb{R}, \\ \mathbf{z} \cdot \mathbf{u} = 0 \end{cases} .$$



$$\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}} \implies \mathbf{z} \cdot \mathbf{u} = 0 \iff \alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \text{ and } \widehat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$$

Orthogonal projection

Columbia University in the City of New York

Let $\mathbf{y} \in \mathbb{R}^n$, $W \subseteq \mathbb{R}^n$ a subspace, $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ a basis for W.

Definition 9 (Orthogonal projection)

The **orthogonal projection** $\text{proj}_W \mathbf{y}$ is s.t. $\mathbf{y} = \text{proj}_W \mathbf{y} + \mathbf{z}$ with

1. $\text{proj}_W \mathbf{y} \in W$
2. $\mathbf{z} = \mathbf{y} - \text{proj}_W \mathbf{y} \in W^\perp$

Remark: we often use $\widehat{\mathbf{y}}$ to denote $\text{proj}_W \mathbf{y}$.

Theorem 10 (Orthogonal projection)

1. $\widehat{\mathbf{y}}$ *is unique.*
2. $\|\mathbf{y} - \widehat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \ \forall \mathbf{v} \in W^\perp, \mathbf{v} \neq \widehat{\mathbf{y}}$ *(**best approximation**),*
3. $\mathcal{B}$ *orthogonal* $\implies \widehat{\mathbf{y}} = \sum_{i=1}^p \frac{(\mathbf{y} \cdot \mathbf{v}_i)}{(\mathbf{v}_i \cdot \mathbf{v}_i)} \mathbf{v}_i,$
4. $\mathcal{B}$ *orthonormal* $\implies \widehat{\mathbf{y}} = \sum_{i=1}^p (\mathbf{y} \cdot \mathbf{v}_i) \mathbf{v}_i.$

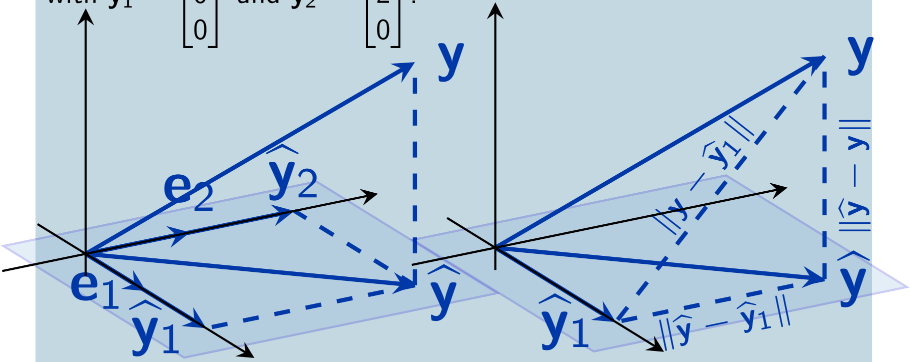
Remark: for 3. and 4., we sometimes write $\widehat{\mathbf{y}} = \sum_{i=1}^p \widehat{\mathbf{y}}_i$.
Remark: for 4., we can rewrite $\widehat{\mathbf{y}} = UU^\top \mathbf{y}$ for $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{bmatrix}$.

Orthogonal projection cont'd

Example 8 (Orthogonal projection)

Let $W = \text{Span}\,\{\mathbf{e}_1, \mathbf{e}_2\}$ and $\mathbf{y} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$, then $\widehat{\mathbf{y}} = \widehat{\mathbf{y}}_1 + \widehat{\mathbf{y}}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$

with $\widehat{\mathbf{y}}_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$ and $\widehat{\mathbf{y}}_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$.



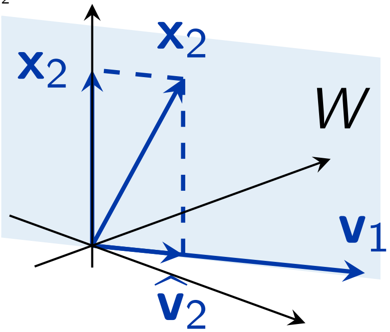
Outline

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

The general idea

Let $W = \text{Span}\ \{\mathbf{x}_1, \mathbf{x}_2\}$ where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.

1. $\mathbf{v}_1 = \mathbf{x}_1$
2. $\mathbf{v}_2 = \mathbf{x}_2 - \underbrace{\text{proj}_{\mathbf{v}_1} \mathbf{x}_2}_{\widehat{\mathbf{v}}_2} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$
3. $\{\mathbf{v}_1, \mathbf{v}_2\}$ is lin. ind. and $\dim W = 2 \implies$ orthogonal basis



The Gram-Schmidt process

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Theorem 11 (Gram-Schmidt)

Given $\{\mathbf{x}_1, \cdots, \mathbf{x}_p\}$ *a basis of a nonzero subspace* $W \subseteq \mathbb{R}^n$*, define*

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}.
\end{aligned}$$

Then $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ *is an orthogonal basis of* W *and* $W_k = \mathit{Span}\,\{\mathbf{v}_1, \cdots, \mathbf{v}_k\} = \mathit{Span}\,\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}\ \forall k$.

Remark: $\mathbf{v}_k = \mathbf{x}_k - \mathrm{proj}_{W_k} \mathbf{x}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\mathbf{x}_k \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$.
Remark: an orthonormal basis is obtained with $\mathbf{u}_k = \mathbf{v}_k / \|\mathbf{v}_k\|$.

The Gram-Schmidt process

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Example 9 (Gram-Schmidt)

Let $\mathbf{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$ ($\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2, \}$ lin. ind.

$\Longrightarrow$ $W = \text{Span}\ \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2, \}$ is a subspace of $\mathbb{R}^4$).

1. $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{Span}\ \{\mathbf{v}_1\}$.

2a. $\mathbf{v}_2 = \mathbf{x}_2 - \text{proj}_{W_1}\mathbf{x}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix}$.

2b. (optional scaling) $\mathbf{v}_2' = 4\mathbf{v}_2$.

3. $W_2 = \text{Span}\ \{\mathbf{v}_1, \mathbf{v}_2'\}$.

4. $\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{W_2}\mathbf{x}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2' = \begin{bmatrix}0\\-2/3\\1/3\\1/3\end{bmatrix}$.

5. $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal basis of W.

Outline

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$.

Definition 10 (Least-squares solution)

A **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is $\widehat{\mathbf{x}} \in \mathbb{R}^n$ s.t.

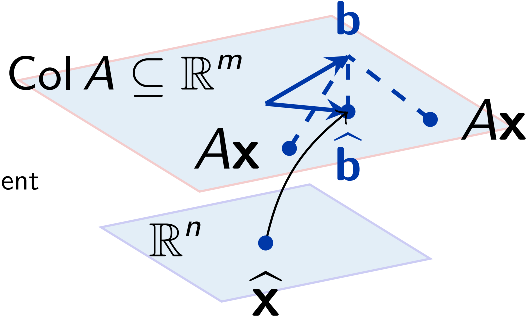
$$\|\mathbf{b} - A\widehat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\| \ \forall \mathbf{x} \in \mathbb{R}^n.$$

Let $\widehat{\mathbf{b}} = A\widehat{\mathbf{x}} \in \operatorname{Col} A$, then

Thm 10 $\implies \widehat{\mathbf{b}} = \operatorname{proj}_{\operatorname{Col} A}\mathbf{b}$

$\implies A\mathbf{x} = \widehat{\mathbf{b}}$ consistent

with $\widehat{\mathbf{x}}$ the solution



Normal equations

Let A be an $m \times n$ matrix with columns $\mathbf{a}_1, \cdots, \mathbf{a}_n$ & $\mathbf{b} \in \mathbb{R}^m$, then

$$\widehat{\mathbf{b}} = A\widehat{\mathbf{x}} = \text{proj}_{\text{Col}\,A}\mathbf{b} \iff \mathbf{b} - \widehat{\mathbf{b}} \in (\text{Col}\,A)^\perp \iff \mathbf{a}_j \cdot (\mathbf{b} - A\widehat{\mathbf{x}})\,\forall j$$
$$\iff A^\top(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0} \iff A^\top A\widehat{\mathbf{x}} = A^\top\mathbf{b}$$

Definition 11 (Normal equations)

The **normal equations** for $A\mathbf{x} = \mathbf{b}$ are $A^\top A\mathbf{x} = A^\top\mathbf{b}$.

Theorem 12 (Normal equations)

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ is equal to the (nonempty) set of solutions of $A^\top A\mathbf{x} = A^\top\mathbf{b}$.
Furthermore, the following statements are equivalent:

1. $A\mathbf{x} = \mathbf{b}$ *has a unique least-squares solution* $\forall\mathbf{b} \in \mathbb{R}^m$.
2. *The columns of A are linearly independent.*
3. $A^\top A$ *is invertible.*

Remark: 1.-2.-3. $\implies \widehat{\mathbf{x}} = (A^\top A)^{-1}A^\top\mathbf{b}$.

Normal equations cont'd

Example 10 (Normal equations I)

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} \implies A^\top A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \text{ and } A^\top \mathbf{b} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

Then $A^\top A\mathbf{x} = A^\top \mathbf{b}$ becomes $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$. Since $A^\top A$ is invertible, we have

$$\begin{aligned} \widehat{\mathbf{x}} &= (A^\top A)^{-1} A^\top \mathbf{b} \\ &= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} \end{aligned}$$

Normal equations cont'd

Example 11 (Normal equations II)

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix} \implies A^\top A = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \text{ and}$$

$$A^\top \mathbf{b} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}. \text{Then } \begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \overset{\text{RREF}}{\Longrightarrow}$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \implies \widehat{\mathbf{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$