Orthogonality

Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Linear Algebra

Outline

- Over-determined System $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
- Inner Product $(\boldsymbol{u}, \boldsymbol{v})$
- Vector Length $\|\boldsymbol{v}\|$
- Orthogonality $\boldsymbol{u} \perp \boldsymbol{v}$
- Orthogonal Complement $\mathbb{S}^{\perp}$
- Orthogonality of Fundamental Subspaces
- Projection
- Gram-Schmidt Process
- Function Approximation*

Over-determined System of Linear Equations

Definition (over-determined systems)

A system of linear equations is **over-determined** if there are more equations than unknowns.

For an over-determined system of linear equations

$$\begin{cases} a_{11}\,x_1 + \cdots + a_{1n}\,x_n = b_1 \\ \vdots \\ \vdots \\ \vdots \\ a_{m1}\,x_1 + \cdots + a_{mn}\,x_n = b_m \end{cases}$$

we have $m > n$.

Theorem (solving an over-determined system)

Let $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ be an over-determined system of linear equations. Suppose the columns of $\boldsymbol{A}$ are linearly independent. Exactly one of the following must be true.

1. *Unique solution*
2. *No solution*

From elimination to minimization

Let $\mathcal{L}$ be an over-determined system of linear equations.

- Elimination is often not good for solving $\mathcal{L}$
- Minimization always works

Let $\boldsymbol{Ax} = \boldsymbol{b}$ be an over-determined system of linear equations with m equations and n unknowns, and the rank of $\boldsymbol{A}$ be r.

- Elimination produces $m - r$ equations with 0 left sides.
- If any equation with 0 left side has a non-zero right side, a solution does not exist.

Definition (least-squares solution)

Let $\mathcal{L} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ be an over-determined system of linear equations. The sum of squared errors of $\mathcal{L}$ is

$$E(\boldsymbol{x}) = \sum_{i=1}^{m} (a_{i1}x_1 + \cdots + a_{in}x_n - b_i)^2$$

A least-squares solution of $\mathcal{L}$, denoted by $\hat{\boldsymbol{x}}$, minimizes $E(\boldsymbol{x})$

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} E(\boldsymbol{x})$$

If there exists an exact solution $\boldsymbol{x}_0$ of $\mathcal{L}$, then $\boldsymbol{x}_0$ must be a least-squares solution of $\mathcal{L}$ since $E(\boldsymbol{x}_0) = 0$ and

$$E(\boldsymbol{x}) \geq 0 = E(\boldsymbol{x}_0)$$

Inner Product

Definition (inner product)

Let $\mathbb{V}$ be a vector space and $\boldsymbol{u}, \boldsymbol{v}$ be vectors of $\mathbb{V}$. An inner product of $\boldsymbol{u}$ and $\boldsymbol{v}$, denoted by $(\boldsymbol{u}, \boldsymbol{v})$, is a scalar function with the following properties.

- Non-negativity

$$(\boldsymbol{u}, \boldsymbol{u}) \geq 0, \ (\boldsymbol{u}, \boldsymbol{u}) = 0 \ \Rightarrow \ \boldsymbol{u} = \boldsymbol{0}$$

- Linearity

$$(\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{u}, \boldsymbol{w}) + (\boldsymbol{v}, \boldsymbol{w}), \ (\boldsymbol{u}, c\boldsymbol{v}) = c(\boldsymbol{u}, \boldsymbol{v})$$

The dot product defined by

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} x_i y_i$$

is an inner product.

Definition (orthogonal vectors)

Let $\mathbb{V}$ be a vector space and $\boldsymbol{u}, \boldsymbol{v}$ be vectors of $\mathbb{V}$. Then $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if and only if $(\boldsymbol{u}, \boldsymbol{v}) = 0$.

That $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal is denoted by

$$\boldsymbol{u} \perp \boldsymbol{v}$$

Definition (length and distance)

Let $\mathbb{V}$ be a vector space and $\boldsymbol{x}, \boldsymbol{y}$ be vectors of $\mathbb{V}$.

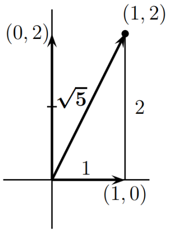
- The length of $\boldsymbol{x}$ is

$$\|\boldsymbol{x}\| = (\boldsymbol{x}, \boldsymbol{x})^{\frac{1}{2}} \text{ or equivalently } \|\boldsymbol{x}\|^2 = (\boldsymbol{x}, \boldsymbol{x})$$

- The distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\|\boldsymbol{x} - \boldsymbol{y}\|$$

Example (length of vector)

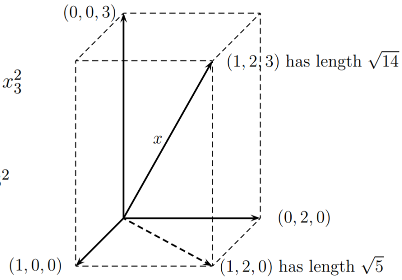


(a)

$$\|x\|^2 = x_1^2 + x_2^2 + x_3^2$$

$$5 = 1^2 + 2^2$$

$$14 = 1^2 + 2^2 + 3^2$$



(b)

Figure 3.1: The length of vectors (x_1, x_2) and (x_1, x_2, x_3).

Theorem (Pythagoras theorem)

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be vectors of $\mathbb{R}^2$ and $\boldsymbol{x} \perp \boldsymbol{y}$. Then

$$\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 = \|\boldsymbol{x} - \boldsymbol{y}\|^2$$

We have

$$\begin{aligned} \boldsymbol{x} \perp \boldsymbol{y} &\Leftrightarrow (\boldsymbol{x}, \boldsymbol{y}) = 0 \Leftrightarrow x_1 y_1 + x_2 y_2 = 0 \\ &\Leftrightarrow (x_1^2 + x_2^2) + (y_1^2 + y_2^2) = (x_1 - y_1)^2 + (x_2 - y_2)^2 \\ &\Leftrightarrow \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 = \|\boldsymbol{x} - \boldsymbol{y}\|^2 \end{aligned}$$

Example (Pythagoras)

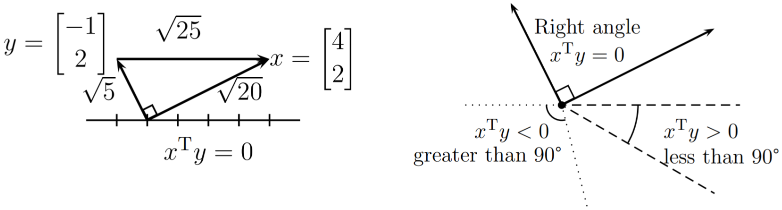


Figure 3.2: A right triangle with $5 + 20 = 25$. Dotted angle 100°, dashed angle 30°.

Orthogonal Sets and Spaces

Definition (orthogonal sets)

Let $\mathcal{U}$ and $\mathcal{W}$ be sets of vectors of space $\mathbb{V}$. $\mathcal{U}$ and $\mathcal{W}$ are orthogonal sets if every vector of $\mathcal{U}$ is orthogonal to every vector of $\mathcal{W}$.

Examples

$$\{[0\ 0\ 1]^T\} \perp \{[1\ 0\ 0]^T, [0\ 1\ 0]^T\}$$
$$\{[1\ 0\ 0]^T\} \perp \{[0\ 1\ 0]^T\}$$
$$\{[1\ 0\ 0]^T, [0\ 0\ 1]^T\} \not\perp \{[0\ 1\ 0]^T, [0\ 0\ 1]^T\}$$

- orthogonal spanning sets
- orthogonal bases
- orthogonal subspaces

Lemma (orthogonal fundamental subspaces)

Let $\boldsymbol{A}$ be a matrix of order $m \times n$.

- *Row space is orthogonal to nullspace*

$$\mathbb{N}(\boldsymbol{A}) \perp \mathbb{C}\left(\boldsymbol{A}^T\right)$$

- *Column space is orthogonal to left nullspace*

$$\mathbb{N}(\boldsymbol{A}^T) \perp \mathbb{C}\left(\boldsymbol{A}\right)$$

$$\begin{aligned}
\boldsymbol{x} \in \mathbb{N}(\boldsymbol{A}) \;\Rightarrow\; & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \;\Rightarrow\; \boldsymbol{a}_{i:}\boldsymbol{x} = 0,\ \forall i \\
\Rightarrow\; & \left(\boldsymbol{x}, \boldsymbol{a}_{i:}^T\right) = 0,\ \forall i \;\Rightarrow\; \left(\boldsymbol{x}, \sum_i c_i \boldsymbol{a}_{i:}^T\right) = 0,\ \forall c_i \\
\Rightarrow\; & \boldsymbol{x} \perp \left(\sum_i c_i \boldsymbol{a}_{i:}^T\right),\ \forall c_i \;\Rightarrow \boldsymbol{x} \perp \mathbb{C}\left(\boldsymbol{A}^T\right)
\end{aligned}$$

Just a spanning set

Let $\mathcal{U}$ and $\mathcal{W}$ be sets of vectors of space $\mathbb{V}$.

- Suppose $\mathcal{U}$ is a subspace. $\mathcal{W} \perp \mathcal{U}$ if and only if $\mathcal{W}$ is orthogonal to a spanning set (e.g. a basis) of $\mathcal{U}$.
- Suppose both $\mathcal{U}$ and $\mathcal{W}$ are subspaces. $\mathcal{W} \perp \mathcal{U}$ if and only if spanning sets of $\mathcal{W}$ and $\mathcal{U}$ are orthogonal.

Suppose

$$\mathbf{dim}\,\mathcal{U} = 3, \ \mathbf{dim}\,\mathcal{W} = 2$$

It suffices to check 3 basis vectors (instead of every vector) of $\mathcal{U}$ against 2 basis vectors of $\mathcal{W}$.

Example (orthogonal subspaces)

- ✓ the z axis and the x–y plane
- ✓ the x axis and the y axis
- × the x–z plane and the y–z plane

Lemma (orthogonal ⇒ linearly independent)

Let $\mathcal{U} = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ be a set of non-zero vectors that are mutually orthogonal. Then $\mathcal{U}$ is linearly independent.

Proof.

Suppose $\sum_{i=1}^{n} c_i \boldsymbol{v}_i = \boldsymbol{0}$. For any j

$$\begin{aligned}\left(\boldsymbol{v}_j, \sum_{i=1}^{n} c_i \boldsymbol{v}_i\right) = 0 \;&\Rightarrow\; \sum_{i=1}^{n} c_i \left(\boldsymbol{v}_j, \boldsymbol{v}_i\right) = 0 \\ &\Rightarrow\; c_j \left(\boldsymbol{v}_j, \boldsymbol{v}_j\right) = 0 \\ &\Rightarrow\; c_j = 0\end{aligned}$$

So $\mathcal{U}$ is linearly independent. □

Definition (orthogonal complement)

Let $\mathbb{S}$ be a subspace of space $\mathbb{V}$. The orthogonal complement of $\mathbb{S}$ is the set of all vectors orthogonal to $\mathbb{S}$.

- Notation for orthogonal complement

$$\mathbb{S}^{\perp} = \{\boldsymbol{v} \mid \boldsymbol{v} \perp \mathbb{S}\}$$

- Orthogonal complement is maximal. For any $\mathbb{T} \perp \mathbb{S}$

$$\boldsymbol{v} \in \mathbb{T} \;\Rightarrow\; \boldsymbol{v} \perp \mathbb{S} \;\Rightarrow\; \boldsymbol{v} \in \mathbb{S}^{\perp}$$

so

$$\mathbb{T} \subset \mathbb{S}^{\perp}$$

An orthogonal complement is a subspace

Let $\mathbb{S}$ be a subspace of space $\mathbb{V}$. $\mathbb{S}^\perp$ is a subspace of $\mathbb{V}$.

Proof.

For any $\boldsymbol{u} \in \mathbb{S}$, scalars c_1, c_2 and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{S}^\perp$

$$(\boldsymbol{u}, c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1(\boldsymbol{u}, \boldsymbol{v}_1) + c_2(\boldsymbol{u}, \boldsymbol{v}_2) = 0$$

$$\Rightarrow \; (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) \perp \mathbb{S}$$

$$\Rightarrow \; c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 \in \mathbb{S}^\perp$$

Thus $\mathbb{S}^\perp$ is a subspace. $\square$

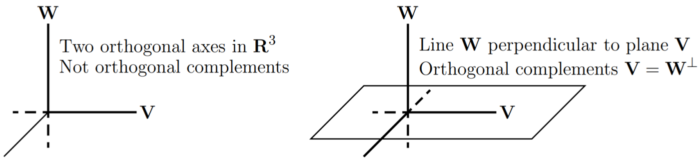


Figure 3.3: Orthogonal complements in $\mathbf{R}^3$: a plane and a line (not two lines).

Theorem (fundamental theorem part II)

Let $\boldsymbol{A}$ *be a matrix of order* $m \times n$.

- *Nullspace is the orthogonal complement of row space*

$$\left(\mathbb{C}\left(\boldsymbol{A}^T\right)\right)^{\perp} = \mathbb{N}(\boldsymbol{A})$$

- *Left nullspace is the orthogonal complement of column space*

$$\left(\mathbb{C}\left(\boldsymbol{A}\right)\right)^{\perp} = \mathbb{N}(\boldsymbol{A}^T)$$

$$\begin{aligned}
\boldsymbol{v} \in \left(\mathbb{C}\left(\boldsymbol{A}^T\right)\right)^{\perp} \;&\Leftrightarrow\; \boldsymbol{v} \perp \mathbb{C}\left(\boldsymbol{A}^T\right) \;\Leftrightarrow\; \boldsymbol{v} \perp \boldsymbol{a}_{i:}^T,\ \forall i \\
&\Leftrightarrow\; \boldsymbol{a}_{i:}\boldsymbol{v} = 0,\ \forall i \\
&\Leftrightarrow\; \boldsymbol{A}\boldsymbol{v} = \boldsymbol{0} \\
&\Leftrightarrow\; \boldsymbol{v} \in \mathbb{N}(\boldsymbol{A})
\end{aligned}$$

Theorem (sum of dimensions)

Let $\mathbb{S}$ be a subspace of $\mathbb{V}$.

$$\textbf{\textit{dim}}\,\mathbb{S} + \textbf{\textit{dim}}\,\mathbb{S}^\perp = \textbf{\textit{dim}}\,\mathbb{V}$$

Proof.

Let $\mathbf{dim}\,\mathbb{S} = r$, $\mathbf{dim}\,\mathbb{S}^\perp = k$, $\mathbf{dim}\,\mathbb{V} = n$. Let $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$ be a basis of $\mathbb{S}$, and $\mathcal{B}' = \{\boldsymbol{v}'_1, \ldots, \boldsymbol{v}'_k\}$ be a basis of $\mathbb{S}^\perp$.

1. $\mathcal{B} \cup \mathcal{B}'$ is a linearly independent set of $\mathbb{V}$, so $r + k \leq n$
2. Augment $\mathcal{B}$ a basis of $\mathbb{V}$, say $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$, and ensure $\boldsymbol{v}_j \perp \mathbf{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{j-1})$ for $j = r+1, \ldots, n$. Let $\mathbb{W} = \mathbf{span}(\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\})$. Then $\mathbf{dim}\,\mathbb{W} = n - r$.

$$\mathbb{W} \perp \mathbb{S} \;\Rightarrow\; \mathbb{W} \subset \mathbb{S}^\perp \Rightarrow \mathbf{dim}\,\mathbb{W} \leq \mathbf{dim}\,\mathbb{S}^\perp \Rightarrow n - r \leq k$$
$$\Rightarrow r + k \geq n$$

Hence $r + k = n$. □

Example (sum of dimensions)

Let $\boldsymbol{A}$ be a matrix of order $m \times n$ with rank r.

- We have
$$\mathbb{N}(\boldsymbol{A}) = \mathbb{C}\left(\boldsymbol{A}^T\right)^{\perp}$$
- We also have
$$\mathbf{dim}\,\mathbb{C}\left(\boldsymbol{A}^T\right) + \mathbf{dim}\,\mathbb{N}(\boldsymbol{A}) = r + (n - r) = n$$
- Thus
$$\mathbf{dim}\,\mathbb{C}\left(\boldsymbol{A}^T\right) + \mathbf{dim}\,\mathbb{C}\left(\boldsymbol{A}^T\right)^{\perp} = \mathbf{dim}\,\mathbb{R}^n$$

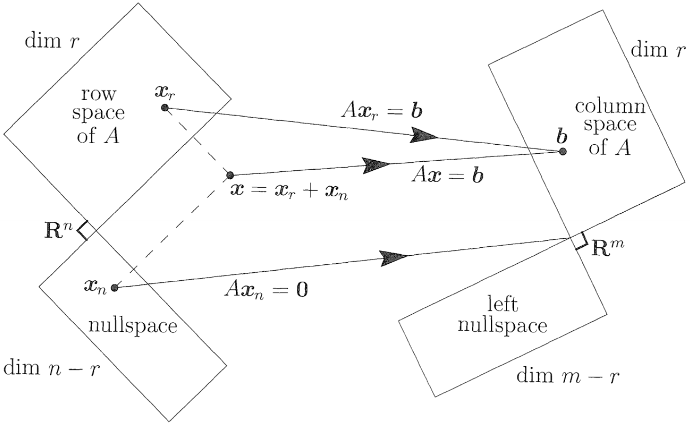


Figure 3.4: The true action $Ax = A(x_{\text{row}} + x_{\text{null}})$ of any m by n matrix.

Projection

Definition (projection as the closest point)

Let $\mathbb{V}$ be a space, $\boldsymbol{b}$ be a vector of $\mathbb{V}$, and $\mathbb{S}$ be a subspace of $\mathbb{V}$. The projection of $\boldsymbol{b}$ to $\mathbb{S}$ is the vector $\boldsymbol{p} \in \mathbb{S}$ with the shortest distance to $\boldsymbol{b}$

$$\boldsymbol{p} = \arg\min_{\boldsymbol{v} \in \mathbb{S}} \|\boldsymbol{b} - \boldsymbol{v}\|$$

- Projection minimizes the length of error vector

$$\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{v}$$

- Note

$$\min_{\boldsymbol{v} \in \mathbb{S}} \|\boldsymbol{b} - \boldsymbol{v}\|^2 \neq \min_{\boldsymbol{v} \in \mathbb{S}} \|\boldsymbol{b} - \boldsymbol{v}\|$$

$$\arg\min_{\boldsymbol{v} \in \mathbb{S}} \|\boldsymbol{b} - \boldsymbol{v}\|^2 = \arg\min_{\boldsymbol{v} \in \mathbb{S}} \|\boldsymbol{b} - \boldsymbol{v}\|$$

- Dealing with $\|\boldsymbol{b} - \boldsymbol{v}\|^2$ is easier than $\|\boldsymbol{b} - \boldsymbol{v}\|$.

min AND arg min

Let $f(\boldsymbol{x})$ be a (multi-variate) function of $\boldsymbol{x}$.

- The minimum value of $f(\boldsymbol{x})$ is denoted by

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

- A value of $\boldsymbol{x}$ that minimizes $f(\boldsymbol{x})$ is denoted by

$$\arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

- Let $f^* = \min_{\boldsymbol{x}} f(\boldsymbol{x})$ and $\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$.

$$f(\boldsymbol{x}^*) = f^*$$

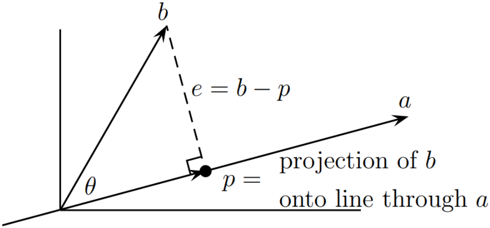


Figure 3.5: The projection p is the point (on the line through a) closest to b.

Orthogonality condition of projection

Let $\boldsymbol{p}$ be the projection of $\boldsymbol{b}$ to $\mathbb{S}$.

- We have
$$(\boldsymbol{b}-\boldsymbol{p}) \perp \mathbb{S}$$
- In particular, since $\boldsymbol{p} \in \mathbb{S}$, we have
$$(\boldsymbol{b}-\boldsymbol{p}) \perp \boldsymbol{p}$$

PROJECTION TO A VECTOR AND PROJECTION MATRIX

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be vectors of space $\mathbb{V}$.

- The projection of $\boldsymbol{b}$ to $\boldsymbol{a}$ is

$$\boldsymbol{p} = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\,\boldsymbol{a}$$

- Projection is left multiplication by a projection matrix

$$\boldsymbol{p} = \boldsymbol{a}\,\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}\,\boldsymbol{b} = \boldsymbol{P}\,\boldsymbol{b}$$

where

$$\boldsymbol{P} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T\boldsymbol{a}}$$

Proof.

Let $\boldsymbol{p} = \boldsymbol{a}x$. By the orthogonality condition

$$\begin{aligned}(\boldsymbol{p}, \boldsymbol{b} - \boldsymbol{p}) = 0 &\Rightarrow \boldsymbol{a}^T(\boldsymbol{b} - \boldsymbol{a}x) = 0 \\ &\Rightarrow (\boldsymbol{a}^T\boldsymbol{a})x = \boldsymbol{a}^T\boldsymbol{b} \\ &\Rightarrow x = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}}\end{aligned}$$

Hence

$$\begin{aligned}\boldsymbol{p} &= \boldsymbol{a}x \\ &= \boldsymbol{a}\,\frac{\boldsymbol{a}^T\boldsymbol{b}}{\boldsymbol{a}^T\boldsymbol{a}} \\ &= \boldsymbol{P}\,\boldsymbol{b}\end{aligned}$$

□

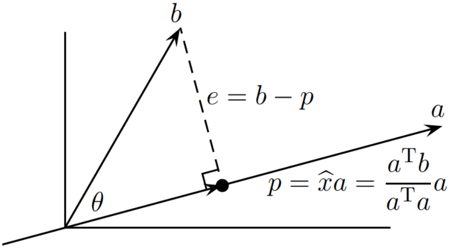


Figure 3.7: The projection p of b onto a, with $\cos\theta = \frac{Op}{Ob} = \frac{a^{\mathrm{T}}b}{\|a\|\|b\|}$.

Note that

$$Op = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\|\boldsymbol{a}\|}$$

Least-squares Solution of Over-determined System

Squared Errors and Least-Squares Solution

Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a system of m equations and n unknowns.

- **Error** of equation i

$$(\boldsymbol{a}_{i:}\boldsymbol{x} - b_i)$$

- Sum of squared errors

$$E(\boldsymbol{x}) = \sum_{i=1}^{m} (\boldsymbol{a}_{i:}\boldsymbol{x} - b_i)^2$$

- Least-squares solution

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} E(\boldsymbol{x})$$

Note

$$E(\boldsymbol{x}) = \|\boldsymbol{Ax} - \boldsymbol{b}\|^2$$

Example (least-squares solution)

Consider

$$\begin{cases} x = 2 \\ 2x = 3 \end{cases}$$

Re-write it as

$$\begin{cases} x - 2 = 0 \\ 2x - 3 = 0 \end{cases}$$

An x incurs an error of $(x-2)$ for the first equation and $(2x-3)$ for the second equation. Hence, a least-squares solution is

$$\begin{aligned} \hat{x} &= \arg\min_x \; E(x) \\ &= \arg\min_x \; \left((x-2)^2 + (2x-3)^2\right) \end{aligned}$$

Example (2 equations and 1 unknown)

Let $\boldsymbol{a}x = \boldsymbol{b}$ be an over-determined system with 2 equations and 1 unknown

$$\mathcal{L} : \begin{cases} a_1 x &= b_1 \\ a_2 x &= b_2 \end{cases}$$

The sum of squared errors is

$$E(x) = (a_1 x - b_1)^2 + (a_2 x - b_2)^2$$

A least-squares solution minimizes $E(x)$. By calculus

$$\hat{x} = \arg\min_x E(x) \;\Rightarrow\; \left.\frac{dE(x)}{dx}\right|_{x=\hat{x}} = 0$$

$$\Rightarrow\; 2[(a_1\hat{x} - b_1)a_1 + (a_2\hat{x} - b_2)a_2] = 0 \;\Rightarrow\; \hat{x} = \frac{a_1 b_1 + a_2 b_2}{a_1^2 + a_2^2}$$

Lemma (least-squares = projection)

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be vectors in space $\mathbb{V}$. Let $\hat{x}$ be the least-squares solution of $\boldsymbol{a}x = \boldsymbol{b}$ and $\boldsymbol{p}$ be the projection of $\boldsymbol{b}$ to $\boldsymbol{a}$.

$$\boldsymbol{p} = \boldsymbol{a}\hat{x}$$

Proof.

$$\hat{x} = \arg\min_{x} \|\boldsymbol{a}x - \boldsymbol{b}\|^2 = \arg\min_{x} \|\boldsymbol{b} - \boldsymbol{a}x\|^2$$

$$\Rightarrow \quad \boldsymbol{a}\hat{x} = \arg\min_{\boldsymbol{v}=\boldsymbol{a}x} \|\boldsymbol{b} - \boldsymbol{v}\|^2$$

Also

$$\boldsymbol{p} = \arg\min_{\boldsymbol{v}\in\textbf{span}(\boldsymbol{a})} \|\boldsymbol{b} - \boldsymbol{v}\|^2 = \arg\min_{\boldsymbol{v}=\boldsymbol{a}x} \|\boldsymbol{b} - \boldsymbol{v}\|^2$$

Hence

$$\boldsymbol{p} = \boldsymbol{a}\hat{x}$$

□

Example (2 equations and 1 unknown)

The least-squares solution of $\boldsymbol{a}x = \boldsymbol{b}$ is

$$\hat{x} = \frac{a_1 b_1 + a_2 b_2}{a_1^2 + a_2^2} = \frac{\boldsymbol{a}^T \boldsymbol{b}}{\boldsymbol{a}^T \boldsymbol{a}}$$

The projection of $\boldsymbol{b}$ to $\boldsymbol{a}$ is

$$\boldsymbol{p} = \boldsymbol{P}\boldsymbol{b} = \frac{\boldsymbol{a}\boldsymbol{a}^T}{\boldsymbol{a}^T \boldsymbol{a}} \boldsymbol{b} = \boldsymbol{a} \frac{\boldsymbol{a}^T \boldsymbol{b}}{\boldsymbol{a}^T \boldsymbol{a}}$$

Hence

$$\boldsymbol{p} = \boldsymbol{a}\hat{x}$$

Theorem (least-squares solution and projection)

Let $\hat{\boldsymbol{x}}$ *be a least-squares solution of* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *and* $\boldsymbol{p}$ *be the projection of* $\boldsymbol{b}$ *to* $\mathbb{C}(\boldsymbol{A})$*.*

$$\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}}$$

Proof.

$$\begin{aligned} \hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = \arg\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|^2 \\ \Rightarrow \quad \boldsymbol{A}\hat{\boldsymbol{x}} &= \arg\min_{\boldsymbol{v}=\boldsymbol{A}\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{v}\|^2 \end{aligned}$$

Also

$$\boldsymbol{p} = \arg\min_{\boldsymbol{v}\in\mathbb{C}(\boldsymbol{A})} \|\boldsymbol{b} - \boldsymbol{v}\|^2 = \arg\min_{\boldsymbol{v}=\boldsymbol{A}\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{v}\|^2$$

Hence $\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}}$. □

Theorem (normal equation)

Let $\hat{\boldsymbol{x}}$ *be a least-squares solution of* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

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

Proof.

The projection of $\boldsymbol{b}$ to $\mathbb{C}(\boldsymbol{A})$ is $\boldsymbol{A}\hat{\boldsymbol{x}}$. It follows that the error vector $(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}})$ is orthogonal to $\mathbb{C}(\boldsymbol{A})$.

$$\begin{aligned}
(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) \perp \mathbb{C}(\boldsymbol{A}) &\Rightarrow (\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) \perp \boldsymbol{a}_i, \ \forall i \\
&\Rightarrow \boldsymbol{a}_i^T(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = 0, \ \forall i \\
&\Rightarrow \boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{0} \\
&\Rightarrow \boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}
\end{aligned}$$

□

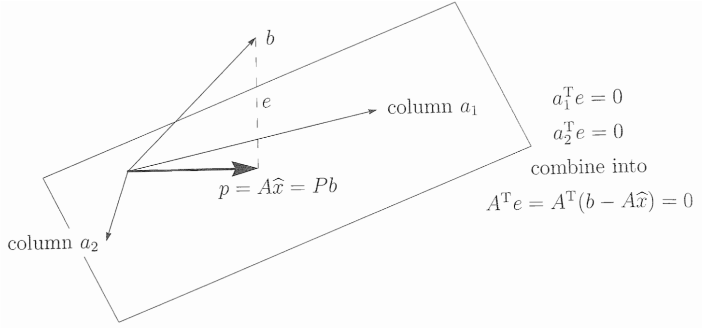


Figure 3.8: Projection onto the column space of a 3 by 2 matrix.

Unique least-squares solution

Let $\boldsymbol{A}$ be a matrix of order $m \times n$ with linearly independent columns and $m > n$. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has unique least-squares solution

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\,\boldsymbol{b}$$

Proof.

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \Rightarrow \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$$

So $\mathbb{N}(\boldsymbol{A}^T\boldsymbol{A}) = \mathbb{N}(\boldsymbol{A})$ and $\mathbf{dim}\,\mathbb{N}(\boldsymbol{A}^T\boldsymbol{A}) = \mathbf{dim}\,\mathbb{N}(\boldsymbol{A}) = 0$. Thus $\mathbf{rank}(\boldsymbol{A}^T\boldsymbol{A}) = n$ and $\left(\boldsymbol{A}^T\boldsymbol{A}\right)$ is invertible. Hence the normal equation $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$ has unique solution

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\,\boldsymbol{b}$$

□

Example (unique least-squares solution)

Find a least-squares solution of an over-determined system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{bmatrix}, \ \boldsymbol{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b} = \begin{bmatrix} 13 & -5 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Theorem (projection to column space)

Let $\boldsymbol{A}$ be a matrix with linearly independent columns.

- *The projection of any $\boldsymbol{b}$ to column space $\mathbb{C}(\boldsymbol{A})$ is*

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

- *The projection matrix is*

$$\boldsymbol{P} = \boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T$$

The least-squares solution of $\boldsymbol{Ax} = \boldsymbol{b}$ is

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{b}$$

The projection of $\boldsymbol{b}$ to $\mathbb{C}(\boldsymbol{A})$ is

$$\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}} = \underbrace{\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T}_{\text{projection matrix}}\boldsymbol{b}$$

Fitting Data

REGRESSION

- Given a data set $\{(t_1, y_1), \ldots, (t_m, y_m)\}$
- Find a function $\hat{y} = f(t)$ to fit the data set

- For example, we may assume

$$\hat{y} = f(t) = c + d\,t$$

- The parameters c and d are decided by minimizing the error between data and function, i.e. between y_i and $f(t_i)$

Linear Fitting Function

We assume

$$\hat{y} = f(t) = c + d\,t$$

- The difference (error) between y_i and $f(t_i)$ is

$$y_i - f(t_i) = y_i - (c + d\,t_i)$$

- Ideally, we want c and d such that

$$y_i = f(t_i) = c + d\,t_i,\ i = 1, \ldots, m$$

- We are solving a system of 2 unknowns (for the parameters c and d) and m equations (for the data points)

A system of linear equations

The equations to satisfy can be written as

$$\begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

It can be represented by $\boldsymbol{Ax} = \boldsymbol{b}$ where

$$\boldsymbol{A} = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \ \boldsymbol{x} = \begin{bmatrix} c \\ d \end{bmatrix}, \ \boldsymbol{b} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

SOLUTION BY NORMAL EQUATION

Consider $\boldsymbol{Ax} = \boldsymbol{b}$ arising from fitting a data set to a function.

- Over-determined if the number of data points is more than the number of parameters
- Look for a least-squares solution

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

- For data $\{(t_1, y_1), \ldots, (t_m, y_m)\}$ and function $f(t) = c + dt$

$$\begin{bmatrix} 1 & \ldots & 1 \\ t_1 & \ldots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} \begin{bmatrix} \hat{c} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} 1 & \ldots & 1 \\ t_1 & \ldots & t_m \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} m & \sum_{i=1}^m t_i \\ \sum_{i=1}^m t_i & \sum_{i=1}^m t_i^2 \end{bmatrix} \begin{bmatrix} \hat{c} \\ \hat{d} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m y_i \\ \sum_{i=1}^m t_i y_i \end{bmatrix}$$

Example (fitting a linear function)

Fit data set $\{(-1,1),(1,1),(2,3)\}$ to a linear function.

It leads to an over-determined system $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ where

$$\boldsymbol{A}=\begin{bmatrix}1 & -1\\ 1 & 1\\ 1 & 2\end{bmatrix},\ \boldsymbol{x}=\begin{bmatrix}c\\ d\end{bmatrix},\ \boldsymbol{b}=\begin{bmatrix}1\\ 1\\ 3\end{bmatrix}$$

We have

$$\boldsymbol{A}^T\boldsymbol{A}=\begin{bmatrix}3 & 2\\ 2 & 6\end{bmatrix},\ \boldsymbol{A}^T\boldsymbol{b}=\begin{bmatrix}5\\ 6\end{bmatrix}$$

Thus, the least-squares solution is

$$\hat{\boldsymbol{x}}=\begin{bmatrix}\hat{c}\\ \hat{d}\end{bmatrix}=(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}=\begin{bmatrix}\frac{9}{7}\\ \frac{4}{7}\end{bmatrix}$$

Among all lines, $f(t)=\hat{c}+\hat{d}t$ minimizes $\sum_{i=1}^{m}(y_i-f(t_i))^2$.

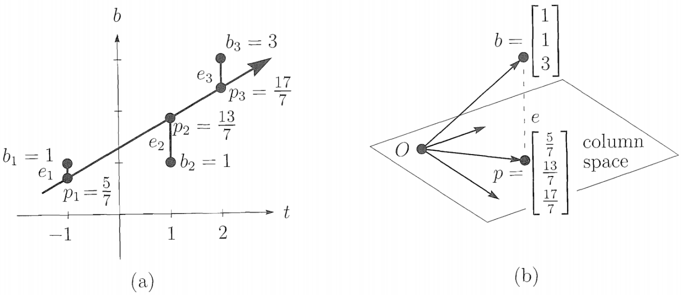


Figure 3.9: Straight-line approximation matches the projection p of b.

Orthonormal Basis

Definition (orthonormal vectors)

A group of vectors are **orthonormal** if

- the vectors are orthogonal
- every vector is a unit vector (of length 1)

- A set with orthonormal vectors is orthonormal
- For an orthonormal set $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$

$$(\boldsymbol{q}_i, \boldsymbol{q}_j) = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

- A basis with orthonormal vectors is orthonormal

SIMPLIFICATION WITH ORTHONORMAL VECTORS

Suppose $\boldsymbol{A}$ has linearly independent column vectors.

- The projection matrix to $\mathbb{C}(\boldsymbol{A})$ is

$$\boldsymbol{P} = \boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T$$

- If the column vectors are orthonormal, we have $\left(\boldsymbol{A}^T\boldsymbol{A}\right) = \boldsymbol{I}$ and

$$\boldsymbol{P} = \boldsymbol{A}\boldsymbol{A}^T = \sum_{j=1}^{n} \boldsymbol{a}_j\boldsymbol{a}_j^T$$

- Note

$$\boldsymbol{P} = \sum_{j=1}^{n} \boldsymbol{P}_j$$

where $\boldsymbol{P}_j = \boldsymbol{a}_j\boldsymbol{a}_j^T$ is the matrix for projection to $\boldsymbol{a}_j$.

Vector representation with orthonormal basis

Let $\mathcal{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ be an orthonormal basis of space $\mathbb{V}$. The representation of a vector $\boldsymbol{x} \in \mathbb{V}$ with $\mathcal{Q}$ is

$$[\boldsymbol{x}_{\mathcal{Q}}] = \begin{bmatrix} (\boldsymbol{q}_1, \boldsymbol{x}) \\ \vdots \\ (\boldsymbol{q}_n, \boldsymbol{x}) \end{bmatrix}$$

Proof.

Suppose $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{q}_i$.

$$(\boldsymbol{q}_j, \boldsymbol{x}) = \left(\boldsymbol{q}_j, \sum_{i=1}^{n} x_i \boldsymbol{q}_i\right) = \sum_{i=1}^{n} x_i \left(\boldsymbol{q}_j, \boldsymbol{q}_i\right) = \sum_{i=1}^{n} x_i \delta_{ij}$$
$$= x_j$$

□

Inner product with an orthonormal basis

Let $\mathcal{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$ be an orthonormal basis of space $\mathbb{V}$. The inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ of $\mathbb{V}$ is the dot product of the representation of $\boldsymbol{x}$ and $\boldsymbol{y}$ with $\mathcal{Q}$

$$(\boldsymbol{x}, \boldsymbol{y}) = [\boldsymbol{x}_{\mathcal{Q}}]^T [\boldsymbol{y}_{\mathcal{Q}}]$$

Proof.

Suppose $\boldsymbol{x} = \sum_{i=1}^{n} x_i \boldsymbol{q}_i$ and $\boldsymbol{y} = \sum_{j=1}^{n} y_j \boldsymbol{q}_j$.

$$\begin{aligned}(\boldsymbol{x}, \boldsymbol{y}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j (\boldsymbol{q}_i, \boldsymbol{q}_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \delta_{ij} = \sum_{i=1}^{n} x_i y_i \\ &= [\boldsymbol{x}_{\mathcal{Q}}]^T [\boldsymbol{y}_{\mathcal{Q}}]\end{aligned}$$

□

Projection to a unit vector

Let $\boldsymbol{q}$ be a unit vector. The matrix for the projection to $\boldsymbol{q}$ is

$$\boldsymbol{P} = \boldsymbol{q}\boldsymbol{q}^T$$

Proof.

We have $\boldsymbol{q}^T\boldsymbol{q} = \|\boldsymbol{q}\|^2 = 1$, so

$$\boldsymbol{P} = \frac{\boldsymbol{q}\boldsymbol{q}^T}{\boldsymbol{q}^T\boldsymbol{q}} = \boldsymbol{q}\boldsymbol{q}^T$$

□

PROJECTION TO COLUMN SPACE: ORTHOGONAL MATRIX

Let $\boldsymbol{Q}$ be a matrix with orthonormal column vectors. The matrix for the projection to $\mathbb{C}(\boldsymbol{Q})$ is

$$\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^T$$

$$\boldsymbol{P} = \boldsymbol{Q}\left(\boldsymbol{Q}^T\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}^T = \boldsymbol{Q}\boldsymbol{Q}^T$$

Projection to a space with an orthonormal basis

Let $\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}$ be an orthonormal basis of space $\mathbb{V}$. Any $\boldsymbol{x} \in \mathbb{V}$ is the sum of the projections of $\boldsymbol{x}$ to the basis vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$.

Proof.

Let $\boldsymbol{Q}$ be the matrix with columns $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$. The projection matrix to $\mathbb{C}(\boldsymbol{Q}) = \mathbb{V}$ is

$$\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^T = \sum_{j=1}^{n} \boldsymbol{q}_j \boldsymbol{q}_j^T$$

For any $\boldsymbol{x} \in \mathbb{V}$, we have

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x} = \left(\sum_{j=1}^{n} \boldsymbol{q}_j \boldsymbol{q}_j^T \right) \boldsymbol{x} = \sum_{j=1}^{n} \boldsymbol{q}_j \left(\boldsymbol{q}_j^T \boldsymbol{x} \right)$$

□

Gram-Schmidt process

G.-S. process converts a basis to an orthonormal one

$$\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \longrightarrow \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$$

For $j = 1, \ldots, n$, do the following operations

- **projection** of $\boldsymbol{a}_j$ to span$(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{j-1})$

$$\boldsymbol{p}_j = \boldsymbol{q}_1(\boldsymbol{q}_1, \boldsymbol{a}_j) + \cdots + \boldsymbol{q}_{j-1}(\boldsymbol{q}_{j-1}, \boldsymbol{a}_j)$$

- **normalization**

$$\boldsymbol{b}_j = \boldsymbol{a}_j - \boldsymbol{p}_j \neq 0, \quad \boldsymbol{q}_j = \frac{\boldsymbol{b}_j}{\|\boldsymbol{b}_j\|}$$

Note

$$\textbf{span}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_j) = \textbf{span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_j)$$

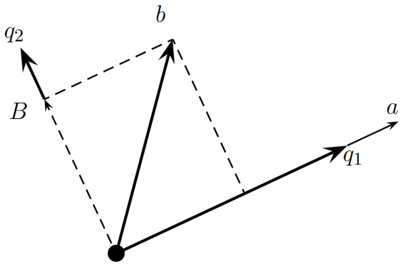


Figure 3.10: The q_i component of b is removed; a and B normalized to q_1 and q_2.

Example (Gram-Schmidt process)

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$\Rightarrow \boldsymbol{b}_1 = \boldsymbol{a}_1, \quad \boldsymbol{q}_1 = \frac{\boldsymbol{b}_1}{\|\boldsymbol{b}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\boldsymbol{b}_2 = \boldsymbol{a}_2 - \boldsymbol{q}_1(\boldsymbol{q}_1, \boldsymbol{a}_2) = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{-1}{2} \end{bmatrix}, \quad \boldsymbol{q}_2 = \frac{\boldsymbol{b}_2}{\|\boldsymbol{b}_2\|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$$

$$\boldsymbol{b}_3 = \boldsymbol{a}_3 - \boldsymbol{q}_1(\boldsymbol{q}_1, \boldsymbol{a}_3) - \boldsymbol{q}_2(\boldsymbol{q}_2, \boldsymbol{a}_3) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{q}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Theorem (QR factorization)

Suppose $\boldsymbol{A}$ has linearly independent columns. Then $\boldsymbol{A} = \boldsymbol{QR}$, where $\boldsymbol{Q}$ has orthonormal columns and $\boldsymbol{R}$ is right-triangular.

Proof.

Let the columns of $\boldsymbol{A}$ be $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. Apply G.-S. to $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ to get an orthonormal basis $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n\}$.

$$\boldsymbol{a}_j = \sum_{i=1}^{n} \boldsymbol{q}_i (\boldsymbol{q}_i, \boldsymbol{a}_j), \; j = 1, \ldots, n$$

Construct $\boldsymbol{Q}$ with columns $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ and $\boldsymbol{R}$ with elements $r_{ij} = (\boldsymbol{q}_i, \boldsymbol{a}_j)$ so $\boldsymbol{A} = \boldsymbol{QR}$. $\boldsymbol{R}$ is right-triangular since for $i > j$

$$\begin{aligned} \boldsymbol{q}_i \perp \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_j\} \;&\Rightarrow\; \boldsymbol{q}_i \perp \boldsymbol{a}_j \in \textbf{span}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_j) \\ &\Rightarrow\; (\boldsymbol{q}_i, \boldsymbol{a}_j) = 0 \end{aligned}$$

□

Example (QR factorization)

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix} \xrightarrow{\text{G.-S.}} \boldsymbol{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix}$$

$$\boldsymbol{R} = \{(\boldsymbol{q}_i, \boldsymbol{a}_j)\} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$$

Function Approximation*

SPACE OF FUNCTIONS AND AN INNER PRODUCT

- The set of real-valued functions is a space
- We denote $f(t)$ by $\boldsymbol{f}$ since it is a vector in a space
- An inner product in this space is defined by

$$(\boldsymbol{f}, \boldsymbol{g}) = \int_I f(t)g(t)dt$$

- Two functions are orthogonal if

$$(\boldsymbol{f}, \boldsymbol{g}) = \int_I f(t)g(t)dt = 0$$

- The length of a function is defined by

$$\|\boldsymbol{f}\|^2 = (\boldsymbol{f}, \boldsymbol{f}) = \int_I f^2(t)dt$$

FUNCTION APPROXIMATION BY POLYNOMIAL

Consider the function approximation problem

$$f(t) \doteq c_0 + c_1 t + c_2 t^2$$

- Denote $\boldsymbol{f} = f(t),\ \boldsymbol{f}_1 = 1,\ \boldsymbol{f}_2 = t,\ \boldsymbol{f}_3 = t^2$. We have

$$\boldsymbol{f} \doteq c_0 \boldsymbol{f}_1 + c_1 \boldsymbol{f}_2 + c_2 \boldsymbol{f}_3$$

- In matrix and vectors

$$\overbrace{\begin{bmatrix} \\ \boldsymbol{f}_1 & \boldsymbol{f}_2 & \boldsymbol{f}_3 \\ \\ \end{bmatrix}}^{\boldsymbol{F}} \overbrace{\begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}}^{\boldsymbol{x}} \doteq \boldsymbol{f}$$

- This is an over-determined system of linear equations

- The over-determined system is

$$\boldsymbol{F}\boldsymbol{x} = \boldsymbol{f}$$

- The normal equation is

$$\boldsymbol{F}^T\boldsymbol{F}\hat{\boldsymbol{x}} = \boldsymbol{F}^T\boldsymbol{f}$$

- That is

$$\begin{bmatrix} (\boldsymbol{f}_1, \boldsymbol{f}_1) & (\boldsymbol{f}_1, \boldsymbol{f}_2) & (\boldsymbol{f}_1, \boldsymbol{f}_3) \\ (\boldsymbol{f}_2, \boldsymbol{f}_1) & (\boldsymbol{f}_2, \boldsymbol{f}_2) & (\boldsymbol{f}_2, \boldsymbol{f}_3) \\ (\boldsymbol{f}_3, \boldsymbol{f}_1) & (\boldsymbol{f}_3, \boldsymbol{f}_2) & (\boldsymbol{f}_3, \boldsymbol{f}_3) \end{bmatrix} \begin{bmatrix} \hat{c}_0 \\ \hat{c}_1 \\ \hat{c}_2 \end{bmatrix} = \begin{bmatrix} (\boldsymbol{f}_1, \boldsymbol{f}) \\ (\boldsymbol{f}_2, \boldsymbol{f}) \\ (\boldsymbol{f}_3, \boldsymbol{f}) \end{bmatrix}$$

- The least-squares solution is

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{F}^T\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\boldsymbol{f}$$

Example (function approximation)

Approximate $\boldsymbol{f} = t^5$ by $\boldsymbol{p} = c + dt$ in the interval $I = (0, 1)$.

- Shortest distance

$$(\hat{c}, \hat{d}) = \underset{(c,d)}{\arg\min} \|\boldsymbol{f} - \boldsymbol{p}\|^2$$

- Over-determined system and least-squares solution

$$\boldsymbol{F}\boldsymbol{x} = \boldsymbol{f}, \quad \text{i.e.} \begin{bmatrix} (\boldsymbol{f}_1 = 1) & (\boldsymbol{f}_2 = t) \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = (\boldsymbol{f} = t^5)$$

- Projection

SHORTEST DISTANCE

- The distance between $\boldsymbol{f}$ and $\boldsymbol{p}$ is $\|\boldsymbol{f}-\boldsymbol{p}\|$

$$\begin{aligned}\|\boldsymbol{f}-\boldsymbol{p}\|^2 &= \int_0^1 (t^5 - c - dt)^2 dt \\ &= \int_0^1 (t^{10} + c^2 + d^2t^2 - 2ct^5 - 2dt^6 + 2cdt)dt \\ &= \frac{1}{11} + c^2 + \frac{1}{3}d^2 - \frac{1}{3}c - \frac{2}{7}d + cd\end{aligned}$$

- At the shortest distance, the partial derivatives are zero

$$\begin{cases} 2\hat{c} + \hat{d} = \dfrac{1}{3} \\ \hat{c} + \dfrac{2}{3}\hat{d} = \dfrac{2}{7} \end{cases} \quad \Rightarrow \quad \hat{c} = -\frac{4}{21}, \ \hat{d} = \frac{5}{7}$$

Over-determined system and solution

- The over-determined system is

$$\boldsymbol{F}\boldsymbol{x} = \boldsymbol{f}$$

- The least-squares solution is

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{F}^T\boldsymbol{F}\right)^{-1}\boldsymbol{F}^T\boldsymbol{f}$$

- That is

$$\begin{aligned}
\begin{bmatrix}\hat{c}\\ \hat{d}\end{bmatrix} &= \begin{bmatrix}(\boldsymbol{f}_1,\boldsymbol{f}_1) & (\boldsymbol{f}_1,\boldsymbol{f}_2)\\ (\boldsymbol{f}_2,\boldsymbol{f}_1) & (\boldsymbol{f}_2,\boldsymbol{f}_2)\end{bmatrix}^{-1}\begin{bmatrix}(\boldsymbol{f}_1,\boldsymbol{f})\\ (\boldsymbol{f}_2,\boldsymbol{f})\end{bmatrix}\\
&= \begin{bmatrix}1 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{3}\end{bmatrix}^{-1}\begin{bmatrix}\frac{1}{6}\\ \frac{1}{7}\end{bmatrix} = \begin{bmatrix}4 & -6\\ -6 & 12\end{bmatrix}\begin{bmatrix}\frac{1}{6}\\ \frac{1}{7}\end{bmatrix} = \begin{bmatrix}\frac{-4}{21}\\ \frac{5}{7}\end{bmatrix}
\end{aligned}$$

PROJECTION METHOD

- Find the projection directly
- An orthogonal basis makes projection easy
- Find an orthogonal basis from (non-orthogonal) $\{\boldsymbol{f}_1, \boldsymbol{f}_2\}$

$$\boldsymbol{q}_1 = \boldsymbol{f}_1,\ \boldsymbol{b}_2 = \boldsymbol{f}_2 - \boldsymbol{q}_1(\boldsymbol{q}_1, \boldsymbol{f}_2) = t - \frac{1}{2}$$

- The projection is

$$\begin{aligned}
\boldsymbol{p} &= \boldsymbol{q}_1(\boldsymbol{q}_1, \boldsymbol{f}) + \boldsymbol{b}_2 \frac{(\boldsymbol{b}_2, \boldsymbol{f})}{(\boldsymbol{b}_2, \boldsymbol{b}_2)} \\
&= 1 \int_0^1 (t^5)(1) dt + \left(t - \frac{1}{2}\right) \frac{\int_0^1 (t^5)\left(t - \frac{1}{2}\right) dt}{\int_0^1 \left(t - \frac{1}{2}\right)\left(t - \frac{1}{2}\right) dt} \\
&= \frac{1}{6} + \frac{\frac{1}{7} - \frac{1}{12}}{\frac{1}{3} - \frac{1}{2} + \frac{1}{4}} \left(t - \frac{1}{2}\right) = -\frac{4}{21} + \frac{5}{7} t
\end{aligned}$$