Orthogonality

Bioinformatics Degree
Algebra

Departament de Matemàtiques

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Outline

Outline

Definition

Definition

The dot product (or scalar product) $u \cdot v$ of two vectors $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ is

$$u \cdot v := u^t v = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n.$$

Example:

$$u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, v = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \Rightarrow u \cdot v = (123) \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = 1 \times 1 + 2 \times 0 + 3 \times 2 = 7$$

Definition

Definition

The dot product (or scalar product) $u \cdot v$ of two vectors

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n \text{ is}$$

$$u \cdot v := u^t v = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n.$$

Example:

$$u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, v = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \Rightarrow u \cdot v = (123) \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = 1 \times 1 + 2 \times 0 + 3 \times 2 = 7$$

Properties:

1. $u \cdot u \geq 0 \; \forall u$
2. $u \cdot u = 0 \Leftrightarrow u = 0.$
3. $u \cdot v = v \cdot u.$
4. $(a_1 u_1 + a_2 u_2) \cdot v = a_1 u_1 \cdot v + a_2 u_2 \cdot v;$
5. $u \cdot (a_1 v_1 + a_2 v_2) = a_1 u \cdot v_1 + a_2 u \cdot v_2.$

Properties:

1. $u \cdot u \geq 0 \ \forall u$
2. $u \cdot u = 0 \Leftrightarrow u = 0.$
3. $u \cdot v = v \cdot u.$
4. $(a_1 u_1 + a_2 u_2) \cdot v = a_1 u_1 \cdot v + a_2 u_2 \cdot v;$
5. $u \cdot (a_1 v_1 + a_2 v_2) = a_1 u \cdot v_1 + a_2 u \cdot v_2.$

Properties:

1. $u \cdot u \geq 0 \; \forall u$
2. $u \cdot u = 0 \Leftrightarrow u = 0.$
3. $u \cdot v = v \cdot u.$
4. $(a_1 u_1 + a_2 u_2) \cdot v = a_1 u_1 \cdot v + a_2 u_2 \cdot v;$
5. $u \cdot (a_1 v_1 + a_2 v_2) = a_1 u \cdot v_1 + a_2 u \cdot v_2.$

Properties:

1. $u \cdot u \geq 0 \; \forall u$
2. $u \cdot u = 0 \Leftrightarrow u = 0.$
3. $u \cdot v = v \cdot u.$
4. $(a_1 u_1 + a_2 u_2) \cdot v = a_1 u_1 \cdot v + a_2 u_2 \cdot v;$
5. $u \cdot (a_1 v_1 + a_2 v_2) = a_1 u \cdot v_1 + a_2 u \cdot v_2.$

Properties:

1. $u \cdot u \geq 0 \; \forall u$
2. $u \cdot u = 0 \Leftrightarrow u = 0.$
3. $u \cdot v = v \cdot u.$
4. $(a_1 u_1 + a_2 u_2) \cdot v = a_1 u_1 \cdot v + a_2 u_2 \cdot v;$
5. $u \cdot (a_1 v_1 + a_2 v_2) = a_1 u \cdot v_1 + a_2 u \cdot v_2.$

Norm and distance

The norm of $u \in \mathbb{R}^n$ is $\|u\| = \sqrt{u \cdot u}$.
Example: $\|(1,2,0)\| = \sqrt{1 \times 1 + 2 \times 2 + 0} = \sqrt{5}$.
Properties:

- $\|u\| \geq 0$;
- $\|cu\| = |c|\|u\|$ $c \in \mathbb{R}$;
- $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality);
- $\|u\| = 0 \Leftrightarrow u = 0$

A vector u is called a unit vector if $\|u\| = 1$. Given a vector $v \neq 0$, we can always find a unit vector in its direction: $v/\|v\|$ (we say that we have normalized v).

The distance between two points P, $Q \in \mathbb{R}^n$, is $d(P, Q) = \|P - Q\|$.

Norm and distance

The norm of $u \in \mathbb{R}^n$ is $\|u\| = \sqrt{u \cdot u}$.
Example: $\|(1,2,0)\| = \sqrt{1 \times 1 + 2 \times 2 + 0} = \sqrt{5}$.
Properties:

- $\|u\| \geq 0$;
- $\|cu\| = |c|\|u\|$ $c \in \mathbb{R}$;
- $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality);
- $\|u\| = 0 \Leftrightarrow u = 0$

A vector u is called a unit vector if $\|u\| = 1$. Given a vector $v \neq 0$, we can always find a unit vector in its direction: $v/\|v\|$ (we say that we have normalized v).

The distance between two points P, $Q \in \mathbb{R}^n$, is $d(P,Q) = \|P - Q\|$.

Norm and distance

The norm of $u \in \mathbb{R}^n$ is $\|u\| = \sqrt{u \cdot u}$.
Example: $\|(1,2,0)\| = \sqrt{1 \times 1 + 2 \times 2 + 0} = \sqrt{5}$.
Properties:

- $\|u\| \geq 0$;
- $\|cu\| = |c|\|u\|$ $c \in \mathbb{R}$;
- $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality);
- $\|u\| = 0 \Leftrightarrow u = 0$

A vector u is called a unit vector if $\|u\| = 1$. Given a vector $v \neq 0$, we can always find a unit vector in its direction: $v/\|v\|$ (we say that we have normalized v).

The distance between two points P, $Q \in \mathbb{R}^n$, is $d(P,Q) = \|P - Q\|$.

Norm and distance

The norm of $u \in \mathbb{R}^n$ is $\|u\| = \sqrt{u \cdot u}$.
Example: $\|(1,2,0)\| = \sqrt{1 \times 1 + 2 \times 2 + 0} = \sqrt{5}$.
Properties:

- $\|u\| \geq 0$;
- $\|cu\| = |c|\|u\|$ $c \in \mathbb{R}$;
- $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality);
- $\|u\| = 0 \Leftrightarrow u = 0$

A vector u is called a unit vector if $\|u\| = 1$. Given a vector $v \neq 0$, we can always find a unit vector in its direction: $v/\|v\|$ (we say that we have normalized v).

The distance between two points P, $Q \in \mathbb{R}^n$, is $d(P, Q) = \|P - Q\|$.

Norm and distance

The norm of $u \in \mathbb{R}^n$ is $\|u\| = \sqrt{u \cdot u}$.
Example: $\|(1,2,0)\| = \sqrt{1 \times 1 + 2 \times 2 + 0} = \sqrt{5}$.
Properties:

- $\|u\| \geq 0$;
- $\|cu\| = |c|\|u\|$ $c \in \mathbb{R}$;
- $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality);
- $\|u\| = 0 \Leftrightarrow u = 0$

A vector u is called a unit vector if $\|u\| = 1$. Given a vector $v \neq 0$, we can always find a unit vector in its direction: $v/\|v\|$ (we say that we have normalized v).

The distance between two points $P, Q \in \mathbb{R}^n$, is $d(P, Q) = \|P - Q\|$.

Norm and distance

The norm of $u \in \mathbb{R}^n$ is $\|u\| = \sqrt{u \cdot u}$.
Example: $\|(1, 2, 0)\| = \sqrt{1 \times 1 + 2 \times 2 + 0} = \sqrt{5}$.
Properties:

- $\|u\| \geq 0$;
- $\|cu\| = |c|\|u\|$ $c \in \mathbb{R}$;
- $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality);
- $\|u\| = 0 \Leftrightarrow u = 0$

A vector u is called a unit vector if $\|u\| = 1$. Given a vector $v \neq 0$, we can always find a unit vector in its direction: $v/\|v\|$ (we say that we have normalized v).

The distance between two points P, $Q \in \mathbb{R}^n$, is $d(P, Q) = \|P - Q\|$.

Angle

- Two vectors u, v are orthogonal (also denoted $u \perp v$) if $u \cdot v = 0$.
- The angle between two vectors $u, v \in \mathbb{R}^n$ is the angle that has $\cos(\widehat{uv}) = \frac{u \cdot v}{||u|| \cdot ||v||}$ (the sign of $\widehat{uv}$ depends on the orientation we choose).
- Two orthogonal vectors have $\widehat{uv} = \pm\frac{\pi}{2}$.

Angle

- Two vectors u, v are orthogonal (also denoted $u \perp v$) if $u \cdot v = 0$.
- The angle between two vectors $u, v \in \mathbb{R}^n$ is the angle that has $\cos(\widehat{uv}) = \frac{u \cdot v}{||u|| \cdot ||v||}$ (the sign of $\widehat{uv}$ depends on the orientation we choose).
- Two orthogonal vectors have $\widehat{uv} = \pm\frac{\pi}{2}$.

Angle

- Two vectors u, v are orthogonal (also denoted $u \perp v$) if $u \cdot v = 0$.
- The angle between two vectors $u, v \in \mathbb{R}^n$ is the angle that has $\cos(\widehat{uv}) = \frac{u \cdot v}{||u|| \cdot ||v||}$ (the sign of $\widehat{uv}$ depends on the orientation we choose).
- Two orthogonal vectors have $\widehat{uv} = \pm\frac{\pi}{2}$.

Orthonormal basis

Definition

An orthogonal basis is a basis $\{v_1, \dots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$

Definition

An orthonormal basis is a basis $\{v_1, \dots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$
- and normalized: $\|v_i\| = 1$ for $i = 1, 2, \dots, n$.

Example: the standard basis is an orthonormal basis.

Orthonormal basis

Definition

An orthogonal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$

Definition

An orthonormal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$
- and normalized: $\|v_i\| = 1$ for $i = 1, 2, \ldots, n$.

Example: the standard basis is an orthonormal basis.

Orthonormal basis

Definition

An orthogonal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$

Definition

An orthonormal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$
- and normalized: $\|v_i\| = 1$ for $i = 1, 2, \ldots, n.$

Example: the standard basis is an orthonormal basis.

Orthonormal basis

Definition

An orthogonal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$

Definition

An orthonormal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$
- and normalized: $\|v_i\| = 1$ for $i = 1, 2, \ldots, n$.

Example: the standard basis is an orthonormal basis.

Orthonormal basis

Definition

An orthogonal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$

Definition

An orthonormal basis is a basis $\{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ such that its vectors are

- pairwise orthogonal: $v_i \cdot v_j = 0$ if $i \neq j, 1,$
- and normalized: $\|v_i\| = 1$ for $i = 1, 2, \ldots, n$.

Example: the standard basis is an orthonormal basis.

Orthonormal basis

Given a basis $B = \{v_1, \dots, v_n\}$ of $\mathbb{R}^n$ consider the matrix $A = (v_1 \dots v_n)$ then,

- $A^t A$ is a diagonal matrix if and only if B is orthogonal.
- $A^t A = Id$ if and only if B is orthonormal.

Definition

An $n \times n$ matrix that satisfies $A^t A = Id$ is called an orthogonal matrix.

Note: The columns of an orthogonal matrix form an orthonormal basis.

Orthonormal basis

Given a basis $B = \{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ consider the matrix $A = (v_1 \ldots v_n)$ then,

- $A^t A$ is a diagonal matrix if and only if B is orthogonal.
- $A^t A = Id$ if and only if B is orthonormal.

Definition
An $n \times n$ matrix that satisfies $A^t A = Id$ is called an orthogonal matrix.

Note: The columns of an orthogonal matrix form an orthonormal basis.

Orthonormal basis

Given a basis $B = \{v_1, \ldots, v_n\}$ of $\mathbb{R}^n$ consider the matrix $A = (v_1 \ldots v_n)$ then,

- $A^t A$ is a diagonal matrix if and only if B is orthogonal.
- $A^t A = Id$ if and only if B is orthonormal.

Definition

An $n \times n$ matrix that satisfies $A^t A = Id$ is called an orthogonal matrix.

Note: The columns of an orthogonal matrix form an orthonormal basis.

Orthonormal basis

Given a basis $B = \{v_1, \dots, v_n\}$ of $\mathbb{R}^n$ consider the matrix $A = (v_1 \dots v_n)$ then,

- $A^t A$ is a diagonal matrix if and only if B is orthogonal.
- $A^t A = Id$ if and only if B is orthonormal.

Definition

An $n \times n$ matrix that satisfies $A^t A = Id$ is called an orthogonal matrix.

Note: The columns of an orthogonal matrix form an orthonormal basis.

Cross-product in $\mathbb{R}^3$

The cross-product between two vectors $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ is the following vector (in standard basis)

$$\begin{aligned}(u_1, u_2, u_3) \times (v_1, v_2, v_3) &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \\ &= (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).\end{aligned}$$

Main properties:

- $v \times u = -u \times v$ (anti-commutative)
- $u \times v$ is orthogonal to both u and v
- If u, v are orthogonal and normalized, then $u, v, u \times v$ is an orthonormal basis.

Cross-product in $\mathbb{R}^3$

The cross-product between two vectors $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ is the following vector (in standard basis)

$$\begin{aligned}(u_1, u_2, u_3) \times (v_1, v_2, v_3) &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \\ &= (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).\end{aligned}$$

Main properties:

- $v \times u = -u \times v$ (anti-commutative)
- $u \times v$ is orthogonal to both u and v
- If u, v are orthogonal and normalized, then $u, v, u \times v$ is an orthonormal basis.

Cross-product in $\mathbb{R}^3$

The cross-product between two vectors $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ is the following vector (in standard basis)

$$(u_1, u_2, u_3) \times (v_1, v_2, v_3) = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$
$$= (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).$$

Main properties:

- $v \times u = -u \times v$ (anti-commutative)
- $u \times v$ is orthogonal to both u and v
- If u, v are orthogonal and normalized, then $u, v, u \times v$ is an orthonormal basis.

Outline

Orthogonal complement

The orthogonal complement to a given subspace $F \subset \mathbb{R}^n$ is

$$F^\perp = \{u \in \mathbb{R}^n \,|\, u \perp v \text{ for all } v \in F\}.$$

If $F = [v_1, \ldots, v_d]$, then

$$F^\perp = \left\{ u \in \mathbb{R}^n \;\middle|\; \begin{array}{c} u \cdot v_1 = 0 \\ \vdots \\ u \cdot v_d = 0 \end{array} \right\}$$

- If $F \subseteq \mathbb{R}^n$ has dimension d, then $F^\perp$ has dimension $n - d$.
- The orthogonal of the orthogonal is the subspace itself: $(F^\perp)^\perp = F$.
- $F \cap F^\perp = 0$, $F + F^\perp = \mathbb{R}^n$.

Orthogonal complement

The orthogonal complement to a given subspace $F \subset \mathbb{R}^n$ is

$$F^\perp = \{u \in \mathbb{R}^n \,|\, u \perp v \text{ for all } v \in F\}.$$

If $F = [v_1, \ldots, v_d]$, then

$$F^\perp = \left\{ u \in \mathbb{R}^n \;\middle|\; \begin{array}{c} u \cdot v_1 = 0 \\ \vdots \\ u \cdot v_d = 0 \end{array} \right\}$$

- If $F \subseteq \mathbb{R}^n$ has dimension d, then $F^\perp$ has dimension $n - d$.
- The orthogonal of the orthogonal is the subspace itself: $(F^\perp)^\perp = F$.
- $F \cap F^\perp = 0$, $F + F^\perp = \mathbb{R}^n$.

Orthogonal complement

The orthogonal complement to a given subspace $F \subset \mathbb{R}^n$ is

$$F^\perp = \{u \in \mathbb{R}^n \,|\, u \perp v \text{ for all } v \in F\}.$$

If $F = [v_1, \ldots, v_d]$, then

$$F^\perp = \left\{ u \in \mathbb{R}^n \;\middle|\; \begin{array}{c} u \cdot v_1 = 0 \\ \vdots \\ u \cdot v_d = 0 \end{array} \right\}$$

- If $F \subseteq \mathbb{R}^n$ has dimension d, then $F^\perp$ has dimension $n - d$.
- The orthogonal of the orthogonal is the subspace itself: $(F^\perp)^\perp = F$.
- $F \cap F^\perp = 0$, $F + F^\perp = \mathbb{R}^n$.

- If F is defined by generators $\Rightarrow$ the equations of $F^{\perp}$ are easy to get: their coefficients are the generators coordinates.
- If F is given by equations $\Rightarrow$ the generators of $F^{\perp}$ are easy to get: their coordinates are the coefficients of the equations.

F	$F^{\perp}$
$[(1,3,2),(-2,1,8)]$	$\begin{cases} x+3y+2z=0 \\ -2x+y+8z=0 \end{cases}$
$\{3x-5y+\frac{11}{2}z\}=0$	$[(3,-5,\frac{11}{2})]$

- If F is defined by generators $\Rightarrow$ the equations of $F^\perp$ are easy to get: their coefficients are the generators coordinates.
- If F is given by equations $\Rightarrow$ the generators of $F^\perp$ are easy to get: their coordinates are the coefficients of the equations.

F	$F^\perp$
$[(1,3,2),(-2,1,8)]$	$\begin{cases} x+3y+2z=0 \\ -2x+y+8z=0 \end{cases}$
$\{3x-5y+\frac{11}{2}z\}=0$	$[(3,-5,\frac{11}{2})]$

Outline

Orthogonal projection

Theorem (Orthogonal Decomposition)

Any $v \in \mathbb{R}^n$ can be written in a unique way as $v = w + w'$ where $w \in F$ and $w' \in F^\perp$.

- w is called the *orthogonal projection* of v on F and is denoted as $proj_F(v)$,
- w' is called the *orthogonal projection* of v on $F^\perp$ and is denoted as $proj_{F^\perp}(v)$.
- Thus, $v = proj_F(v) + proj_{F^\perp}(v)$ and $proj_F(v)$ is the unique vector of F such that $v - proj_F(v)$ belongs to $F^\perp$.

Geometric property: $proj_F(v)$ is the vector of F that is closest to v; this is, $\min\{\|v - w\| \mid w \in F\}$ is achieved at $\|v - proj_F(v)\|$ (and equals $\|proj_{F^\perp}(v)\|$). The orthogonal projection $proj_F(v)$ is the best approximation to v in F.

Orthogonal projection

Theorem (Orthogonal Decomposition)

Any $v \in \mathbb{R}^n$ can be written in a unique way as $v = w + w'$ where $w \in F$ and $w' \in F^\perp$.

- w is called the *orthogonal projection* of v on F and is denoted as $proj_F(v)$,
- w' is called the *orthogonal projection* of v on $F^\perp$ and is denoted as $proj_{F^\perp}(v)$.
- Thus, $v = proj_F(v) + proj_{F^\perp}(v)$ and $proj_F(v)$ is the unique vector of F such that $v - proj_F(v)$ belongs to $F^\perp$.

Geometric property: $proj_F(v)$ is the vector of F that is closest to v; this is, $\min\{\|v - w\| \mid w \in F\}$ is achieved at $\|v - proj_F(v)\|$ (and equals $\|proj_{F^\perp}(v)\|$). The orthogonal projection $proj_F(v)$ is the best approximation to v in F.

Orthogonal projection

Theorem (Orthogonal Decomposition)

Any $v \in \mathbb{R}^n$ can be written in a unique way as $v = w + w'$ where $w \in F$ and $w' \in F^\perp$.

- w is called the *orthogonal projection* of v on F and is denoted as $proj_F(v)$,
- w' is called the *orthogonal projection* of v on $F^\perp$ and is denoted as $proj_{F^\perp}(v)$.
- Thus, $v = proj_F(v) + proj_{F^\perp}(v)$ and $proj_F(v)$ is the unique vector of F such that $v - proj_F(v)$ belongs to $F^\perp$.

Geometric property: $proj_F(v)$ is the vector of F that is closest to v; this is, $\min\{\|v - w\| \mid w \in F\}$ is achieved at $\|v - proj_F(v)\|$ (and equals $\|proj_{F^\perp}(v)\|$). The orthogonal projection $proj_F(v)$ is the best approximation to v in F.

Orthogonal projection

Theorem (Orthogonal Decomposition)

Any $v \in \mathbb{R}^n$ can be written in a unique way as $v = w + w'$ where $w \in F$ and $w' \in F^\perp$.

- w is called the *orthogonal projection* of v on F and is denoted as $proj_F(v)$,
- w' is called the *orthogonal projection* of v on $F^\perp$ and is denoted as $proj_{F^\perp}(v)$.
- Thus, $v = proj_F(v) + proj_{F^\perp}(v)$ and $proj_F(v)$ is the unique vector of F such that $v - proj_F(v)$ belongs to $F^\perp$.

Geometric property: $proj_F(v)$ is the vector of F that is closest to v; this is, $\min\{\|v - w\| \mid w \in F\}$ is achieved at $\|v - proj_F(v)\|$ (and equals $\|proj_{F^\perp}(v)\|$). The orthogonal projection $proj_F(v)$ is the best approximation to v in F.

Computation of the orthogonal projection

Proposition

$proj_F(v)$ is the unique vector w that satisfies $w \in F$ and $v - w \in F^\perp$. If F has basis $u_1, \ldots, u_d$, then $proj_F(v)$ is the unique vector w such that

$$w = c_1 u_1 + \ldots c_d u_d \in F \quad and \left\{ \begin{array}{c} u_1 \cdot w = u_1 \cdot v \\ \vdots \\ u_d \cdot w = u_d \cdot v \end{array} \right.$$

Thus, $proj_F(v)$ is the vector $c_1 u_1 + \cdots + c_d u_d$ such that $c_1, \ldots, c_d$ are solution to the system

$$\begin{pmatrix} u_1 \cdot u_1 & \ldots & u_1 \cdot u_d \\ \vdots & \vdots & \vdots \\ u_d \cdot u_1 & \ldots & u_d \cdot u_d \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} = \begin{pmatrix} v \cdot u_1 \\ \vdots \\ v \cdot u_d \end{pmatrix}$$

If A is the matrix $\begin{pmatrix} u_1 & \cdots & u_d \end{pmatrix}$, then $c_1, \ldots, c_d$ are the solution to the system

$$A^t A \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} = A^t v.$$

(If $u_1, \ldots, u_d$ is a basis, then $A^t A$ is invertible).

Corollary

If $\dim F = 1$, $F = [u]$, *then* $proj_F(v) = \frac{v \cdot u}{u \cdot u}\ u$.

Orthogonal projection with orthonormal basis

Proposition

If $u_1, \ldots, u_d$ is an orthonormal basis of F and $v \in \mathbb{R}^n$, then

$$proj_F(v) = (v \cdot u_1)u_1 + \cdots + (v \cdot u_d)u_d.$$

Outline

Spectral theorem

Theorem

Let A be a symmetric $n \times n$ matrix. Then A has real eigenvalues, diagonalizes, and there exists an orthonormal basis $\{v_1, \ldots, v_n\}$ of eigenvectors. If V has columns $v_1, \ldots, v_n$ and D is the diagonal matrix of eigenvalues, then V is an orthogonal matrix and

$$A = VDV^t.$$

The orthonormal basis of eigenvectors is not difficult to find:

- If u, v are eigenvectors of A of eigenvalues $\lambda \neq \mu$, then $u \perp v$.
- If the eigenvalues are all distinct, then normalizing the eigenvectors we obtain an orthonormal basis of eigenvectors.
- If the eigenvalues are not all distinct, use Gram-Schmidt algorithm (not studied in this course).

Spectral theorem

Theorem

Let A be a symmetric $n \times n$ matrix. Then A has real eigenvalues, diagonalizes, and there exists an orthonormal basis $\{v_1, \ldots, v_n\}$ of eigenvectors. If V has columns $v_1, \ldots, v_n$ and D is the diagonal matrix of eigenvalues, then V is an orthogonal matrix and

$$A = VDV^t.$$

The orthonormal basis of eigenvectors is not difficult to find:

- If u, v are eigenvectors of A of eigenvalues $\lambda \neq \mu$, then $u \perp v$.
- If the eigenvalues are all distinct, then normalizing the eigenvectors we obtain an orthonormal basis of eigenvectors.
- If the eigenvalues are not all distinct, use Gram-Schmidt algorithm (not studied in this course).

Spectral theorem

Theorem

Let A be a symmetric $n \times n$ matrix. Then A has real eigenvalues, diagonalizes, and there exists an orthonormal basis $\{v_1, \dots, v_n\}$ of eigenvectors. If V has columns $v_1, \dots, v_n$ and D is the diagonal matrix of eigenvalues, then V is an orthogonal matrix and

$$A = VDV^t.$$

The orthonormal basis of eigenvectors is not difficult to find:

- If u, v are eigenvectors of A of eigenvalues $\lambda \neq \mu$, then $u \perp v$.
- If the eigenvalues are all distinct, then normalizing the eigenvectors we obtain an orthonormal basis of eigenvectors.
- If the eigenvalues are not all distinct, use Gram-Schmidt algorithm (not studied in this course).

Spectral theorem

Theorem

Let A be a symmetric $n \times n$ matrix. Then A has real eigenvalues, diagonalizes, and there exists an orthonormal basis $\{v_1, \ldots, v_n\}$ of eigenvectors. If V has columns $v_1, \ldots, v_n$ and D is the diagonal matrix of eigenvalues, then V is an orthogonal matrix and

$$A = VDV^t.$$

The orthonormal basis of eigenvectors is not difficult to find:

- If u, v are eigenvectors of A of eigenvalues $\lambda \neq \mu$, then $u \perp v$.
- If the eigenvalues are all distinct, then normalizing the eigenvectors we obtain an orthonormal basis of eigenvectors.
- If the eigenvalues are not all distinct, use Gram-Schmidt algorithm (not studied in this course).

Outline

Linear least squares approximation

Problem: $Ax = b$ might be incompatible due to measure errors in b, but we would still like to have an approximated solution:

$$\underset{\text{system}}{\overset{\text{Incompatible}}{Ax = b}} \quad \Leftrightarrow \quad b \notin \mathrm{Im}(A)$$

Want: $\tilde{x}$ such that $A\tilde{x}$ is as close to b as possible.

Definition

A least squares solution of $Ax = b$ is a vector $\tilde{x}$ that minimizes $\|Ax - b\|$, that is

$$\|A\tilde{x} - b\| \leq \|Ax - b\| \text{ for all } x$$

Solution of the least squares problem

Solution given by Gauss (1801)

- Change b by the vector of $\text{Im}(A)$ that is closest to b: the *orthogonal projection* of b in $\text{Im}(A)$, $proj_{\text{Im}(A)}(b)$.
- Find a solution $\tilde{x}$ to the compatible system $Ax = proj_{\text{Im}(A)}(b)$
- Then $\tilde{x}$ is a least square solution of $Ax = b$.
- $\tilde{x}$ does not satisfy $Ax - b = \vec{0}$, but minimizes the norm $\|Ax - b\|$ among all x.
- The residual measures how far $\tilde{x}$ is from a solution to the system:

$$\text{residual} = A\tilde{x} - b \text{ (which is } = proj_{\text{Im}(A)}(b) - b).$$

  norm of the residual: $\|A\tilde{x} - b\|$
- Important point: we do not need to compute $proj_{\text{Im}(A)}(b)$ (see next slide).

Solution of the least squares problem

Solution given by Gauss (1801)

- Change b by the vector of $\mathrm{Im}(A)$ that is closest to b: the *orthogonal projection* of b in $\mathrm{Im}(A)$, $proj_{\mathrm{Im}(A)}(b)$.
- Find a solution $\tilde{x}$ to the compatible system $Ax = proj_{\mathrm{Im}(A)}(b)$
- Then $\tilde{x}$ is a least square solution of $Ax = b$.
- $\tilde{x}$ does not satisfy $Ax - b = \vec{0}$, but minimizes the norm $\|Ax - b\|$ among all x.
- The residual measures how far $\tilde{x}$ is from a solution to the system:

  $$\text{residual} = A\tilde{x} - b \text{ (which is } = proj_{\mathrm{Im}(A)}(b) - b).$$

  norm of the residual: $\|A\tilde{x} - b\|$
- Important point: we do not need to compute $proj_{\mathrm{Im}(A)}(b)$ (see next slide).

Solution of the least squares problem

Solution given by Gauss (1801)

- Change b by the vector of $\mathrm{Im}(A)$ that is closest to b: the *orthogonal projection* of b in $\mathrm{Im}(A)$, $proj_{\mathrm{Im}(A)}(b)$.
- Find a solution $\tilde{x}$ to the compatible system $Ax = proj_{\mathrm{Im}(A)}(b)$
- Then $\tilde{x}$ is a least square solution of $Ax = b$.
- $\tilde{x}$ does not satisfy $Ax - b = \vec{0}$, but minimizes the norm $\|Ax - b\|$ among all x.
- The residual measures how far $\tilde{x}$ is from a solution to the system:

  $$\text{residual} = A\tilde{x} - b \text{ (which is } = proj_{\mathrm{Im}(A)}(b) - b).$$

  norm of the residual: $\|A\tilde{x} - b\|$
- Important point: we do not need to compute $proj_{\mathrm{Im}(A)}(b)$ (see next slide).

Solution of the least squares problem

Solution given by Gauss (1801)

- Change b by the vector of $\mathrm{Im}(A)$ that is closest to b: the *orthogonal projection* of b in $\mathrm{Im}(A)$, $proj_{\mathrm{Im}(A)}(b)$.
- Find a solution $\tilde{x}$ to the compatible system $Ax = proj_{\mathrm{Im}(A)}(b)$
- Then $\tilde{x}$ is a least square solution of $Ax = b$.
- $\tilde{x}$ does not satisfy $Ax - b = \vec{0}$, but minimizes the norm $\|Ax - b\|$ among all x.
- The residual measures how far $\tilde{x}$ is from a solution to the system:

  $$\text{residual} = A\tilde{x} - b \text{ (which is } = proj_{\mathrm{Im}(A)}(b) - b).$$

  norm of the residual: $\|A\tilde{x} - b\|$
- Important point: we do not need to compute $proj_{\mathrm{Im}(A)}(b)$ (see next slide).

Solution of the least squares problem

Solution given by Gauss (1801)

- Change b by the vector of $\mathrm{Im}(A)$ that is closest to b: the *orthogonal projection* of b in $\mathrm{Im}(A)$, $proj_{\mathrm{Im}(A)}(b)$.
- Find a solution $\tilde{x}$ to the compatible system $Ax = proj_{\mathrm{Im}(A)}(b)$
- Then $\tilde{x}$ is a least square solution of $Ax = b$.
- $\tilde{x}$ does not satisfy $Ax - b = \vec{0}$, but minimizes the norm $\|Ax - b\|$ among all x.
- The residual measures how far $\tilde{x}$ is from a solution to the system:

  $$\text{residual} = A\tilde{x} - b \text{ (which is } = proj_{\mathrm{Im}(A)}(b) - b\text{)}.$$

  norm of the residual: $\|A\tilde{x} - b\|$
- Important point: we do not need to compute $proj_{\mathrm{Im}(A)}(b)$ (see next slide).

Solution of the least squares problem

Solution given by Gauss (1801)

- Change b by the vector of $\mathrm{Im}(A)$ that is closest to b: the *orthogonal projection* of b in $\mathrm{Im}(A)$, $proj_{\mathrm{Im}(A)}(b)$.
- Find a solution $\tilde{x}$ to the compatible system $Ax = proj_{\mathrm{Im}(A)}(b)$
- Then $\tilde{x}$ is a least square solution of $Ax = b$.
- $\tilde{x}$ does not satisfy $Ax - b = \vec{0}$, but minimizes the norm $\|Ax - b\|$ among all x.
- The residual measures how far $\tilde{x}$ is from a solution to the system:

  $$\text{residual} = A\tilde{x} - b \text{ (which is } = proj_{\mathrm{Im}(A)}(b) - b\text{)}.$$

  norm of the residual: $\|A\tilde{x} - b\|$
- Important point: we do not need to compute $proj_{\mathrm{Im}(A)}(b)$ (see next slide).

Theorem

- $\tilde{x}$ is a least squares solution of $Ax = b$ if and only if it is a solution to the normal equations:

$$A^t A x = A^t b.$$

- If the rank of A equals the number of columns, then the least squares solution is unique and given by

$$\tilde{x} = (A^t A)^{-1} A^t b$$

(although computing the inverse is not efficient)

- If the original system is compatible, $\tilde{x}$ is a solution to the original system as well.

Theorem

- *$\tilde{x}$ is a least squares solution of $Ax = b$ if and only if it is a solution to the normal equations:*

$$A^t Ax = A^t b.$$

- *If the rank of A equals the number of columns, then the least squares solution is unique and given by*

$$\tilde{x} = (A^t A)^{-1} A^t b$$

(although computing the inverse is not efficient)

- *If the original system is compatible, $\tilde{x}$ is a solution to the original system as well.*

Theorem

- *$\tilde{x}$ is a least squares solution of $Ax = b$ if and only if it is a solution to the normal equations:*

$$A^t A x = A^t b.$$

- *If the rank of A equals the number of columns, then the least squares solution is unique and given by*

$$\tilde{x} = (A^t A)^{-1} A^t b$$

(although computing the inverse is not efficient)

- *If the original system is compatible, $\tilde{x}$ is a solution to the original system as well.*

Theorem

- *$\tilde{x}$ is a least squares solution of $Ax = b$ if and only if it is a solution to the normal equations:*

$$A^t A x = A^t b.$$

- *If the rank of A equals the number of columns, then the least squares solution is unique and given by*

$$\tilde{x} = (A^t A)^{-1} A^t b$$

  (although computing the inverse is not efficient)
- *If the original system is compatible, $\tilde{x}$ is a solution to the original system as well.*

Linear regression

Problem: Given n data points $P_i = (x_i, y_i) \in \mathbb{R}^2$, find a line $y = a_1 x + a_0$ such that $a_1 x_i + a_0 = y_i \ \forall i$:

$$\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

If the system is incompatible, use linear least squares to find a_1, a_0 $\rightarrow$ the line is called the regression line.

Statistically and numerically speaking, it is better to center the data x and y first.

Quadratic regression

Problem: given n data points $P_i = (x_i, y_i) \in \mathbb{R}^2$, find the parabola $y = a_2x^2 + a_1x + a_0$ such that $a_2x_i^2 + a_1x_i + a_0 = y_i \; \forall i$.

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{pmatrix} \begin{pmatrix} a_2 \\ a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

If the system is incompatible, use linear least squares approximation.

- This approach can be followed for polynomials of higher degree (polynomial regression)
- The same approach can be followed to fit other types of functions.

- This approach can be followed for polynomials of higher degree (polynomial regression)
- The same approach can be followed to fit other types of functions.