# Orthogonality

*Bioinformatics Degree*
*Algebra*

Departament de Matemàtiques

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

# Outline

Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# Outline


Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# 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 $\{v_1, \ldots, v_n\}$ 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:

Example: the standard basis is an orthonormal basis.

# 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 $\{v_1, \dots, v_n\}$ 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

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.
- If $A$ is orthogonal, then $A^{-1} = A^t$.

# 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.
- If $A$ is orthogonal, then $A^{-1} = A^t$.

# 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.
- If $A$ is orthogonal, then $A^{-1} = A^t$.

# 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.
- If $A$ is orthogonal, then $A^{-1} = A^t$.

# 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.
- If $A$ is orthogonal, then $A^{-1} = A^t$.

# 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.

# 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.

# Outline


Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# 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 $F = [v_1, \dots, v_d]$, 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: $(F^\perp)^\perp = F$.

# 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 $F = [v_1, \dots, v_d]$, 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: $(F^\perp)^\perp = F$.

- 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.

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

# Outline


Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# 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 $u_1, \dots, u_d$, then $proj_F(v)$ is the unique vector $w$ such that*

$$w = c_1 u_1 + \dots c_d u_d \in F \quad and \begin{cases} u_1 \cdot w = u_1 \cdot v \\ \vdots \\ u_d \cdot w = u_d \cdot v \end{cases}$$

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

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

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


Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# 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 $\{v_1, \ldots, v_n\}$ 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).

# Outline

Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# Orthogonal matrices

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

- If we call the columns $v_1, \dots, v_n$, $A = (v_1 \dots v_n)$, then,

  $$A^t A = Id \quad \text{if and ony if } \{v_1, \dots, v_n\} \text{ is an orthonormal basis.}$$

- $A$ is orthogonal if and only if $A^{-1} = A^t$.
- If $A$ is orthogonal, then the corresponding endomorphism preserves norms (preserves the measures of vectors):

  $$||Ax|| = ||x|| \text{ for all } x,$$

  and preserves dot products and angles (and hence preserves orthogonality).

# Orthogonal matrices

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

- If we call the columns $v_1, \ldots, v_n$, $A = (v_1 \ldots v_n)$, then,

$$A^t A = Id \quad \text{if and ony if } \{v_1, \ldots, v_n\} \text{ is an orthonormal basis.}$$

- $A$ is orthogonal if and only if $A^{-1} = A^t$.
- If $A$ is orthogonal, then the corresponding endomorphism preserves norms (preserves the measures of vectors):

$$\|Ax\| = \|x\| \text{ for all } x,$$

and preserves dot products and angles (and hence preserves orthogonality).

# Orthogonal matrices

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

- If we call the columns $v_1, \dots, v_n$, $A = (v_1 \dots v_n)$, then,

$$A^t A = Id \quad \text{if and ony if } \{v_1, \dots, v_n\} \text{ is an orthonormal basis.}$$

- $A$ is orthogonal if and only if $A^{-1} = A^t$.
- If $A$ is orthogonal, then the corresponding endomorphism preserves norms (preserves the measures of vectors):

$$\|Ax\| = \|x\| \text{ for all } x,$$

  and preserves dot products and angles (and hence preserves orthogonality).

# Singular value decomposition (SVD)

## Theorem (Singular value decomposition)

*Let $A$ be an $m \times n$ matrix. There there exists a decomposition $A = U \cdot D \cdot V^t$, where $U$ is $m \times m$, $V$ is $n \times n$, $U, V$ are orthogonal and*

$$D = \begin{pmatrix} \sigma_1 & & 0 & \cdots & 0 \\ & \ddots & & & \vdots \\ 0 & & \sigma_r & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

*with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ and $r = \text{rank } A$.*
*$\sigma_1, \ldots, \sigma_r$ are called singular values of $A$.*

# How to get the SVD?

The singular values are determined by the matrix $A$:

$$A = UDV^t \Rightarrow A^tA = VD^tU^tUDV^t = VD^tDV^t$$

but $U$ and $V$ are not (although are almost determined in most cases). Ho do we compute the SVD?

(1) Diagonalize the symmetric matrix $S = A^t \cdot A$

(2) If $\lambda_1 \geq \cdots \geq \lambda_r$ are the non-zero eigenvalues of $S \Rightarrow$ the singular values are $\sigma_1 = \sqrt{\lambda_1}, \ldots \sigma_r = \sqrt{\lambda_r}$ (fact: $A^tA$ always has non-negative eigenvalues).

(3) The columns of $V$ are an orthonormal basis $v_1, \ldots, v_n$ of eigenvectors of $S$.

(4) $u_1 = \frac{1}{\sigma_1}Av_1, \ldots, u_r = \frac{1}{\sigma_r}Av_r$ are orthonormal vectors in $\mathbb{R}^m$ (which can be completed to an orthonormal basis of $\mathbb{R}^m$ if necessary) and they form the columns of $U$.

# How to get the SVD?

The singular values are determined by the matrix $A$:

$$A = UDV^t \Rightarrow A^tA = VD^tU^tUDV^t = VD^tDV^t$$

but $U$ and $V$ are not (although are almost determined in most cases). Ho do we compute the SVD?

(1) Diagonalize the symmetric matrix $S = A^t \cdot A$

(2) If $\lambda_1 \geq \cdots \geq \lambda_r$ are the non-zero eigenvalues of $S \Rightarrow$ the singular values are $\sigma_1 = \sqrt{\lambda_1}, \ldots \sigma_r = \sqrt{\lambda_r}$ (fact: $A^tA$ always has non-negative eigenvalues).

(3) The columns of $V$ are an orthonormal basis $v_1, \ldots, v_n$ of eigenvectors of $S$.

(4) $u_1 = \frac{1}{\sigma_1}Av_1, \ldots, u_r = \frac{1}{\sigma_r}Av_r$ are orthonormal vectors in $\mathbb{R}^m$ (which can be completed to an orthonormal basis of $\mathbb{R}^m$ if necessary) and they form the columns of $U$.

# How to get the SVD?

The singular values are determined by the matrix $A$:

$$A = UDV^t \Rightarrow A^tA = VD^tU^tUDV^t = VD^tDV^t$$

but $U$ and $V$ are not (although are almost determined in most cases). Ho do we compute the SVD?

(1) Diagonalize the symmetric matrix $S = A^t \cdot A$

(2) If $\lambda_1 \geq \cdots \geq \lambda_r$ are the non-zero eigenvalues of $S \Rightarrow$ the singular values are $\sigma_1 = \sqrt{\lambda_1}, \ldots \sigma_r = \sqrt{\lambda_r}$ (fact: $A^tA$ always has non-negative eigenvalues).

(3) The columns of $V$ are an orthonormal basis $v_1, \ldots, v_n$ of eigenvectors of $S$.

(4) $u_1 = \frac{1}{\sigma_1}Av_1, \ldots, u_r = \frac{1}{\sigma_r}Av_r$ are orthonormal vectors in $\mathbb{R}^m$ (which can be completed to an orthonormal basis of $\mathbb{R}^m$ if necessary) and they form the columns of $U$.

## How to get the SVD?

The singular values are determined by the matrix $A$:

$$A = UDV^t \Rightarrow A^tA = VD^tU^tUDV^t = VD^tDV^t$$

but $U$ and $V$ are not (although are almost determined in most cases). Ho do we compute the SVD?

(1) Diagonalize the symmetric matrix $S = A^t \cdot A$

(2) If $\lambda_1 \geq \cdots \geq \lambda_r$ are the non-zero eigenvalues of $S \Rightarrow$ the singular values are $\sigma_1 = \sqrt{\lambda_1}, \ldots \sigma_r = \sqrt{\lambda_r}$ (fact: $A^tA$ always has non-negative eigenvalues).

(3) The columns of $V$ are an orthonormal basis $v_1, \ldots, v_n$ of eigenvectors of $S$.

(4) $u_1 = \frac{1}{\sigma_1}Av_1, \ldots, u_r = \frac{1}{\sigma_r}Av_r$ are orthonormal vectors in $\mathbb{R}^m$ (which can be completed to an orthonormal basis of $\mathbb{R}^m$ if necessary) and they form the columns of $U$.

# Significance of the SVD

If $A$ is the standard matrix of a linear map $f : \mathbb{R}^n \to \mathbb{R}^m$, and we call $u_1, \ldots, u_m$, $v_1, \ldots, v_n$, the columns of $U$ and $V$ respectively, then $D$ the matrix associated to $f$ in orthonormal basis $v_1, \ldots, v_n$ and $u_1, \ldots, u_m$:

$$A = M_e(f) = \underbrace{U}_{A_{u\to e}} * \underbrace{D}_{M_{v,u}(f)} * \underbrace{V^t}_{A_{e\to v}}$$

(note that $V^t = V^{-1} = A_{e\to v}$).

# 2-norm of a matrix

To "measure" a linear map we measure how big the image of the unit sphere is under this map:

## Definition

The 2-norm of a matrix $A$ is

$$\|A\|_2 = \max_{||x||=1} ||Ax||.$$

Geometric consequence of the *SVD*:

## Proposition

- $\|A\|_2 = \sigma_1$,
- $\max_{||x||=1} ||Ax|| = ||Av_1||$ *(the maximum is attained at $v_1$)*.
- *If $A$ is invertible,* $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$.

# 2-norm of a matrix

To "measure" a linear map we measure how big the image of the unit sphere is under this map:

## Definition

The 2-norm of a matrix $A$ is

$$\|A\|_2 = \max_{||x||=1} ||Ax||.$$

Geometric consequence of the $SVD$:

## Proposition

- $\|A\|_2 = \sigma_1$,
- $\max_{||x||=1} \|Ax\| = \|Av_1\|$ *(the maximum is attained at* $\pm v_1$*).*
- *If* $A$ *is invertible,* $\|A^{-1}\|_2 = \frac{1}{\sigma_r}$.

# 2-norm of a matrix

To "measure" a linear map we measure how big the image of the unit sphere is under this map:

## Definition

The 2-norm of a matrix $A$ is

$$\|A\|_2 = \max_{||x||=1} ||Ax||.$$

Geometric consequence of the $SVD$:

## Proposition

- $\|A\|_2 = \sigma_1$,
- $\max_{||x||=1} \|Ax\| = \|Av_1\|$ *(the maximum is attained at $\pm v_1$).*
- *If $A$ is invertible, $\|A^{-1}\|_2 = \frac{1}{\sigma_r}$.*

## 2-norm of a matrix

To "measure" a linear map we measure how big the image of the unit sphere is under this map:

### Definition

The 2-norm of a matrix $A$ is

$$\|A\|_2 = \max_{||x||=1} ||Ax||.$$

Geometric consequence of the *SVD*:

### Proposition

- $\|A\|_2 = \sigma_1$,
- $\max_{||x||=1} \|Ax\| = \|Av_1\|$ *(the maximum is attained at* $\pm v_1$*).*
- *If* $A$ *is invertible,* $\|A^{-1}\|_2 = \frac{1}{\sigma_r}$.

# 2-norm of a matrix

To "measure" a linear map we measure how big the image of the unit sphere is under this map:

## Definition

The 2-norm of a matrix $A$ is

$$\|A\|_2 = \max_{||x||=1} ||Ax||.$$

Geometric consequence of the *SVD*:

## Proposition

- $\|A\|_2 = \sigma_1$,
- $\max_{||x||=1} \|Ax\| = \|Av_1\|$ *(the maximum is attained at* $\pm v_1$*).*
- *If $A$ is invertible,* $\|A^{-1}\|_2 = \frac{1}{\sigma_r}$.

# SVD and rank approximation

## Theorem

*Let $A$ be any matrix. If $A = UDV^t$ and the singular values of $A$ are $\sigma_1, \ldots, \sigma_r$ then for any $k \leq r$,*

$$M = U \begin{pmatrix} \sigma_1 & & 0 & \cdots & 0 \\ & \ddots & & & \vdots \\ 0 & & \sigma_k & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} V^t$$

*is the matrix of rank $k$ closest to $A$ (in the sense that $\|A - M\|_2$ is minimal among matrices $M$ of rank $k$).*

# Outline


Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# Outline

Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications

- Linear least squares
- Principal component analysis

Python

# 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:

$$\begin{array}{c}\text{Incompatible}\\ Ax = b\\ \text{system}\end{array} \quad \Leftrightarrow \quad b \notin \mathrm{Im}(A)$$

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 to 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 to $Ax = b$.
- $\tilde{x}$ 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 to 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 to $Ax = b$.
- $\tilde{x}$ 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 to 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 to $Ax = b$.
- $\tilde{x}$ 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 to 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 to $Ax = b$.
- $\tilde{x}$ 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 to 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 to $Ax = b$.
- $\tilde{x}$ 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 to 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 to $Ax = b$.
- $\tilde{x}$ 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).

## Theorem

- $\tilde{x}$ *is a least squares solution to* $Ax = b$ *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 to $Ax = b$ 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 to $Ax = b$ 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 to $Ax = b$ 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.*

# 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.

# Outline

Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

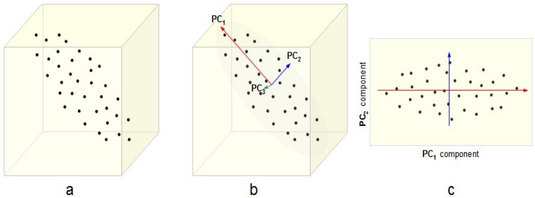
Applications

- Linear least squares
- Principal component analysis

Python

# Principal component analysis

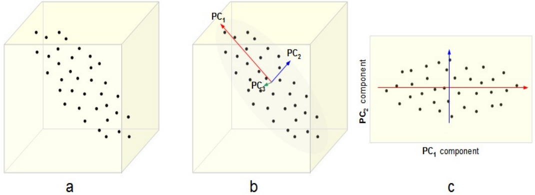
**Goal:** Given $N$ data points in $\mathbb{R}^3$, $p_i = (x_i, y_i, z_i)$, $i = 1, \ldots, N$ highly correlated, one wants to find $v_1 = (a, b, c)$ of norm 1 such that the set $\{t_i = ax_i + by_i + cz_i\}_i$ has maximum variance:




- Note that $proj_{[v_1]}(p_i) = t_i v_1$
- $v_1 = (a, b, c)$ is called the first principal component.
- Then one can look for $v_2 \in [v_1]^\perp$ (2nd principal component) maximizing variance of $proj_{[v_1]^\perp}(p_i)$.
- Keep going or project down to the first components in order to reduce the dimension of the problem.

# Principal component analysis

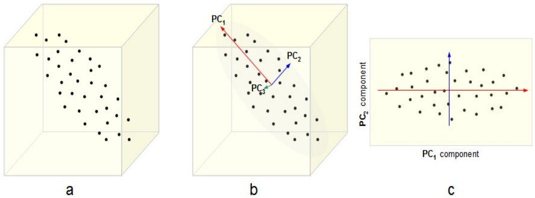
**Goal:** Given $N$ data points in $\mathbb{R}^3$, $p_i = (x_i, y_i, z_i)$, $i = 1, \dots, N$ highly correlated, one wants to find $v_1 = (a, b, c)$ of norm 1 such that the set $\{t_i = ax_i + by_i + cz_i\}_i$ has maximum variance:




- Note that $proj_{[v_1]}(p_i) = t_i v_1$
- $v_1 = (a, b, c)$ is called the first principal component.
- Then one can look for $v_2 \in [v_1]^\perp$ (2nd principal component) maximizing variance of $proj_{[v_1]^\perp}(p_i)$.
- Keep going or project down to the first components in order to reduce the dimension of the problem.

# Principal component analysis

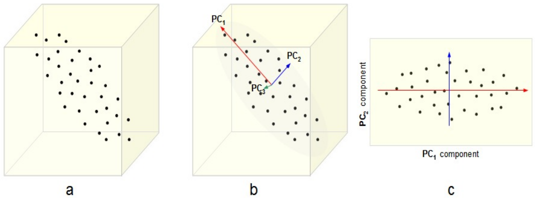
**Goal:** Given $N$ data points in $\mathbb{R}^3$, $p_i = (x_i, y_i, z_i)$, $i = 1, \ldots, N$ highly correlated, one wants to find $v_1 = (a, b, c)$ of norm 1 such that the set $\{t_i = ax_i + by_i + cz_i\}_i$ has maximum variance:




- Note that $proj_{[v_1]}(p_i) = t_i v_1$
- $v_1 = (a, b, c)$ is called the first principal component.
- Then one can look for $v_2 \in [v_1]^\perp$ (2nd principal component) maximizing variance of $proj_{[v_1]^\perp}(p_i)$.
- Keep going or project down to the first components in order to reduce the dimension of the problem.

# Principal component analysis

**Goal:** Given $N$ data points in $\mathbb{R}^3$, $p_i = (x_i, y_i, z_i)$, $i = 1, \dots, N$ highly correlated, one wants to find $v_1 = (a, b, c)$ of norm 1 such that the set $\{t_i = ax_i + by_i + cz_i\}_i$ has maximum variance:




- Note that $proj_{[v_1]}(p_i) = t_i v_1$
- $v_1 = (a, b, c)$ is called the first principal component.
- Then one can look for $v_2 \in [v_1]^\perp$ (2nd principal component) maximizing variance of $proj_{[v_1]^\perp}(p_i)$.
- Keep going or project down to the first components in order to reduce the dimension of the problem.

## Procedure

Assume that set $\{p_i\}$ is centered at the origin. Let

$$M = \begin{pmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix} \text{ so that } \sum_i x_i = \sum_i y_i = \sum_i z_i = 0.$$

- Want $v_1 = (a, b, c)$ of norm 1 such that $\sum_i t_i^2 = \sum_i (ax_i + by_i + cz_i)^2 = ||Mv_1||$ is maximum.
- $v_1$ is the first column vector of $V$ in the SVD: $M = UDV^t$.
- Then the matrix $M_2 = M - Mv_1v_1^t$ has $proj_{[v_1]^\perp}(p_i)$ in its rows.
- $M_2 = \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.
- The direction which maximizes the variance is $v_2$ (2nd vector in $V$).
- Keep going.

# Procedure

Assume that set $\{p_i\}$ is centered at the origin. Let

$$M = \begin{pmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix} \text{ so that } \sum_i x_i = \sum_i y_i = \sum_i z_i = 0.$$

- Want $v_1 = (a, b, c)$ of norm 1 such that $\sum_i t_i^2 = \sum_i (ax_i + by_i + cz_i)^2 = ||Mv_1||$ is maximum.
- $v_1$ is the first column vector of $V$ in the SVD: $M = UDV^t$.
- Then the matrix $M_2 = M - Mv_1v_1^t$ has $proj_{[v_1]^\perp}(p_i)$ in its rows.
- $M_2 = \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.
- The direction which maximizes the variance is $v_2$ (2nd vector in $V$).
- Keep going.

# Procedure

Assume that set $\{p_i\}$ is centered at the origin. Let

$$M = \begin{pmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix} \text{ so that } \sum_i x_i = \sum_i y_i = \sum_i z_i = 0.$$

- Want $v_1 = (a, b, c)$ of norm 1 such that $\sum_i t_i^2 = \sum_i (ax_i + by_i + cz_i)^2 = ||Mv_1||$ is maximum.
- $v_1$ is the first column vector of $V$ in the SVD: $M = UDV^t$.
- Then the matrix $M_2 = M - Mv_1 v_1^t$ has $proj_{[v_1]^\perp}(p_i)$ in its rows.
- $M_2 = \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.
- The direction which maximizes the variance is $v_2$ (2nd vector in $V$).
- Keep going.

# Procedure

Assume that set $\{p_i\}$ is centered at the origin. Let

$$M = \begin{pmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix} \text{ so that } \sum_i x_i = \sum_i y_i = \sum_i z_i = 0.$$

- Want $v_1 = (a, b, c)$ of norm 1 such that $\sum_i t_i^2 = \sum_i (ax_i + by_i + cz_i)^2 = ||Mv_1||$ is maximum.
- $v_1$ is the first column vector of $V$ in the SVD: $M = UDV^t$.
- Then the matrix $M_2 = M - Mv_1v_1^t$ has $proj_{[v_1]^\perp}(p_i)$ in its rows.
- $M_2 = \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.
- The direction which maximizes the variance is $v_2$ (2nd vector in $V$).
- Keep going.

# Procedure

Assume that set $\{p_i\}$ is centered at the origin. Let

$$M = \begin{pmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix}$$ so that $\sum_i x_i = \sum_i y_i = \sum_i z_i = 0$.

- Want $v_1 = (a, b, c)$ of norm 1 such that $\sum_i t_i^2 = \sum_i (ax_i + by_i + cz_i)^2 = ||Mv_1||$ is maximum.
- $v_1$ is the first column vector of $V$ in the SVD: $M = UDV^t$.
- Then the matrix $M_2 = M - Mv_1 v_1^t$ has $proj_{[v_1]^\perp}(p_i)$ in its rows.
- $M_2 = \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.
- The direction which maximizes the variance is $v_2$ (2nd vector in $V$).
- Keep going.

# Procedure

Assume that set $\{p_i\}$ is centered at the origin. Let

$$M = \begin{pmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix} \text{ so that } \sum_i x_i = \sum_i y_i = \sum_i z_i = 0.$$

- Want $v_1 = (a, b, c)$ of norm 1 such that $\sum_i t_i^2 = \sum_i (ax_i + by_i + cz_i)^2 = ||Mv_1||$ is maximum.
- $v_1$ is the first column vector of $V$ in the SVD: $M = UDV^t$.
- Then the matrix $M_2 = M - Mv_1 v_1^t$ has $proj_{[v_1]^\perp}(p_i)$ in its rows.
- $M_2 = \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.
- The direction which maximizes the variance is $v_2$ (2nd vector in $V$).
- Keep going.

Remarks:

- If the set $\{p_i\}$ is not centered at the origin we center it: let $(\bar{x}, \bar{y}, \bar{z}) = \sum_i (x_i, y_i, z_i)/N$, and consider

$$M = \begin{pmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_N - \bar{x} & y_N - \bar{y} & z_N - \bar{z} \end{pmatrix}.$$

  Proceed as before with this $M$ and then sum $(\bar{x}, \bar{y}, \bar{z})$ to the final result.
- The matrix $M^t M$ is the *empirical covariance* matrix and the principal component $v_1$ is the dominant eigenvector of this matrix.
- The same can be done for clouds of points in $\mathbb{R}^n$.

Remarks:

- If the set $\{p_i\}$ is not centered at the origin we center it: let $(\bar{x}, \bar{y}, \bar{z}) = \sum_i (x_i, y_i, z_i)/N$, and consider

$$M = \begin{pmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_N - \bar{x} & y_N - \bar{y} & z_N - \bar{z} \end{pmatrix}.$$

  Proceed as before with this $M$ and then sum $(\bar{x}, \bar{y}, \bar{z})$ to the final result.
- The matrix $M^t M$ is the *empirical covariance* matrix and the principal component $v_1$ is the dominant eigenvector of this matrix.
- The same can be done for clouds of points in $\mathbb{R}^n$.

Remarks:

- If the set $\{p_i\}$ is not centered at the origin we center it: let $(\bar{x}, \bar{y}, \bar{z}) = \sum_i (x_i, y_i, z_i)/N$, and consider

$$M = \begin{pmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_N - \bar{x} & y_N - \bar{y} & z_N - \bar{z} \end{pmatrix}.$$

  Proceed as before with this $M$ and then sum $(\bar{x}, \bar{y}, \bar{z})$ to the final result.
- The matrix $M^t M$ is the *empirical covariance* matrix and the principal component $v_1$ is the dominant eigenvector of this matrix.
- The same can be done for clouds of points in $\mathbb{R}^n$.

# Outline

Distance and angle

Orthogonal complement

Orthogonal projection

Spectral Theorem

Singular value decomposition

Applications
- Linear least squares
- Principal component analysis

Python

# Python

```
>>> import numpy as np
>>> from numpy.linalg import *
```
>>> $A = np.array([[a_{11},\ldots,a_{1n}],\ldots,[a_{n1},\ldots,a_{nn}]])$

To get $U$, $V^t$ and the singular values of $A$ we do:

```
>>> U,d,Vt = svd(A)
```

d is not a matrix, it is an array that contains the signular values.
To convert it to a matrix we can do:

```
>>> D=np.diag(d)
```

if $A$ is a square matrix; if not, we can do:

```
>>> D = np.zeros((n,n),dtype='complex128')
>>> for i in range(n):
      D[i,i] = eigenval[i]
```