Mathematics for Data Science

Lecture 3

Eva FEILLET[1]

[1]LISN
Paris-Saclay University

M1[AI], Fall 2025

Previously covered topics

- (Lecture 1) Vector spaces, subspaces, linear transformations. Rank, image, kernel
- (Lecture 2) Matrices, link with linear transformations, linear systems. Range, rank, kernel of a matrix, inverse of a matrix

Previously covered topics

- (Lecture 1) Vector spaces, subspaces, linear transformations. Rank, image, kernel
- (Lecture 2) Matrices, link with linear transformations, linear systems. Range, rank, kernel of a matrix, inverse of a matrix

In this lecture: determinant and its properties, diagonalization

Table of Contents

Table of Contents

Introduction

Determinants were first introduced in the context of systems of linear equations with as many unknowns as equations, in order to **determine** whether a system admitted a unique solution.
Works by Cardan (2 equations with 2 unknowns, 1545), Leibniz (3 by 3, 1678), Maclaurin (4 by 4, 1748), Cramer (formula for n by n but no proof), Bézout, Lagrange, Gauss, Cauchy...

Figure 1: Gerolamo Cardano, Italian mathematician from the XVI^{th} century

Determinant: definitions

The **determinant** of a square matrix A, denoted $det(A)$, or $|A|$, can be defined in several different ways.

Determinant: definitions

The **determinant** of a square matrix A, denoted $det(A)$, or $|A|$, can be defined in several different ways.

1. Can be defined via the Leibniz formula (explicit, 3 by 3), and generalized to an $n \times n$ matrix involving permutations and their signatures.
2. In the Euclidean space, its absolute value can be interpreted in terms of area and volume and its sign in terms of orientation.
3. Cen be defined using the notion of *signature* of a *permutation* of coefficients
4. Can be characterized as the unique function, defined on the entries of a matrix, satisfying a given set of properties.

Determinant: explicit formulae

The determinant of a 2×2 matrix is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

and the determinant of a 3×3 matrix is

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

Determinant: explicit formulae

The determinant of a 2×2 matrix is

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

and the determinant of a 3×3 matrix is

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

Example

On the board

Geometrical Interpretation in 2D

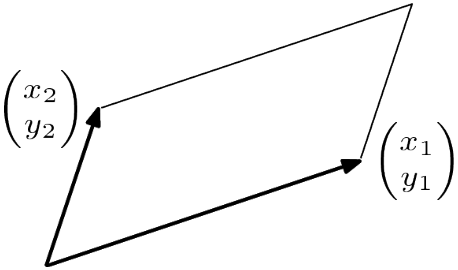


Figure 2: Two vectors in $\mathbb{R}^2$.

Geometrical Interpretation in 2D

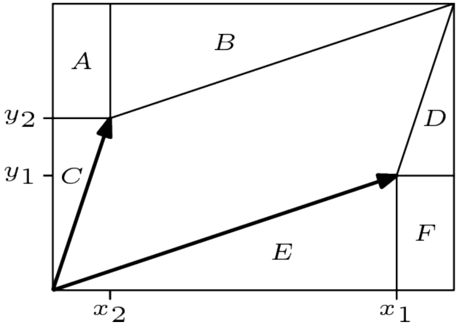
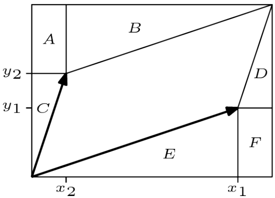


Figure 3: Computing the area of the parallelogram defined by the two vectors in $\mathbb{R}^2$. Image source.

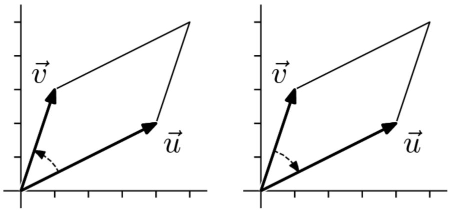
Geometrical Interpretation in 2D



$$
\begin{aligned}
&\text{area of parallelogram} \\
&\quad = \text{area of rectangle} - \text{area of } A - \text{area of } B \\
&\quad\quad - \cdots - \text{area of } F \\
&\quad = (x_1 + x_2)(y_1 + y_2) - x_2 y_1 - x_1 y_1/2 \\
&\quad\quad - x_2 y_2/2 - x_2 y_2/2 - x_1 y_1/2 - x_2 y_1 \\
&\quad = x_1 y_2 - x_2 y_1
\end{aligned}
$$

Figure 4: Computing the area of the parallelogram defined by the two vectors in $\mathbb{R}^2$. Image source.

Geometrical Interpretation in 2D: Orientation matters !



$$\begin{vmatrix} 4 & 1 \\ 2 & 3 \end{vmatrix} = 10 \qquad \begin{vmatrix} 1 & 4 \\ 3 & 2 \end{vmatrix} = -10$$

Figure 5: Computing the area of the parallelogram defined by the two vectors in $\mathbb{R}^2$. Orientation matters ! Oriented surfaces. Image source.

Proposition

1. *Exchanging two columns flips the sign of the determinant.*
2. *Scaling a vector by a factor λ multiplies the determinant by λ.*

Geometrical Interpretation in 2D

Proposition

1. *Exchanging two columns flips the sign of the determinant.*
2. *Scaling a vector by a factor λ multiplies the determinant by λ.*
3. *The determinant is linear with respect to one row, given that the others are fixed.*

$$\begin{vmatrix} b & a \\ d & c \end{vmatrix} = (bc - ad) = -\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\begin{vmatrix} \lambda a & b \\ \lambda c & d \end{vmatrix} = \lambda(ad - bc) = \lambda\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = (ad - bc) + (a'd - b'c) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

Geometrical Interpretation in 2D - Area

In $\mathbb{R}^2$, **the absolute value of** $\det(A)$ **is the scaling factor of area under the transformation** A.

$$A = \begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix}, \quad v = (1,2)^\top \quad w = (2,1)^\top$$

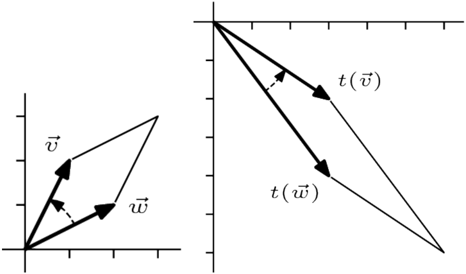


Figure 6: Example with a linear transformation Image source.

Geometrical Interpretation in 3D

Volume of a parallelepiped:

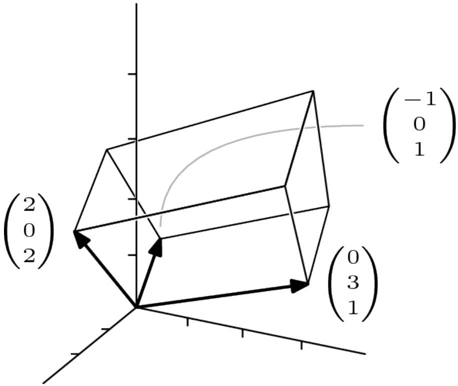


Figure 7: The volume of the box defined by vectors v_1, v_2, v_3 is the absolute value of the determinant of the matrix with v_1, v_2, v_3 as columns. Image source.

In higher dimension

In higher dimensions: A unit ball is transformed into a hyper-ellipsoid via multiplication by A. Its volume is scaled by $|\det(A)|$.

$$\text{New volume} = |\det(A)| \times \text{Original volume}$$

$\diamond$ Practice with exercises / solutions

Permutation and transposition

Definition (Permutation)

Let $n \in \mathbb{N}^*$. We denote by $\mathcal{S}_n$ the set of permutations in $[\![1, n]\!]$.

NB : it corresponds to the bijections from $[\![1, n]\!]$ to $[\![1, n]\!]$. What is the cardinal of $\mathcal{S}_n$?

Permutation and transposition

Definition (Permutation)

Let $n \in \mathbb{N}^*$. We denote by $\mathcal{S}_n$ the set of permutations in $[\![1, n]\!]$.

NB : it corresponds to the bijections from $[\![1, n]\!]$ to $[\![1, n]\!]$. What is the cardinal of $\mathcal{S}_n$? $Card(\mathcal{S}_n)$: $n!$.

Permutation and transposition

Definition (Permutation)

Let $n \in \mathbb{N}^*$. We denote by $\mathcal{S}_n$ the set of permutations in $[\![1, n]\!]$.

NB : it corresponds to the bijections from $[\![1, n]\!]$ to $[\![1, n]\!]$. What is the cardinal of $\mathcal{S}_n$? $Card(\mathcal{S}_n)$: $n!$.

Definition (Transposition)

Let $(i, j) \in [\![1, n]\!], i \leq j$. We call **transposition**, and denote $\tau_{i,j}$ the map in $\mathcal{S}_n$ defined by: $\tau_{i,j} : [\![1, n]\!] \to [\![1, n]\!]$,

$$\tau_{i,j}(k) = k \text{ if } k \notin \{i, j\},$$
$$\tau_{i,j}(k) = i \text{ if } k = j,$$
$$\tau_{i,j}(k) = j \text{ if } k = i$$

NB : What is the inverse map of $\tau_{i,j}$?

Permutation and transposition

Definition (Permutation)

Let $n \in \mathbb{N}^*$. We denote by $\mathcal{S}_n$ the set of permutations in $[\![1, n]\!]$.

NB : it corresponds to the bijections from $[\![1, n]\!]$ to $[\![1, n]\!]$. What is the cardinal of $\mathcal{S}_n$? $Card(\mathcal{S}_n)$: $n!$.

Definition (Transposition)

Let $(i, j) \in [\![1, n]\!], i \leq j$. We call **transposition**, and denote $\tau_{i,j}$ the map in $\mathcal{S}_n$ defined by: $\tau_{i,j} : [\![1, n]\!] \to [\![1, n]\!]$,

$$\tau_{i,j}(k) = k \text{ if } k \notin \{i, j\},$$
$$\tau_{i,j}(k) = i \text{ if } k = j,$$
$$\tau_{i,j}(k) = j \text{ if } k = i$$

NB : What is the inverse map of $\tau_{i,j}$? $\tau_{i,j} \circ \tau_{i,j} = Id$. a transposition is its own inverse.

Permutation and transposition

Proposition (Composition of transpositions)

Any application $\sigma \in \mathcal{S}_n$ can be written as a composition of transpositions, i.e. for any permutation $\sigma \in \mathcal{S}_n$, there exists $\tau_1, \tau_2, \dots \tau_T$, T transpositions in $\mathcal{S}_n$ such that

$$\tau_1 \circ \tau_2 \circ \cdots \circ \tau_T = \sigma$$

NB : Proof by induction over n. Case $n = 1 : \mathcal{S}_1 = \{Id\}$. Case $n = 2 : \mathcal{S}_2 = \{Id, \tau_{1,2}\}$. Case $n = 3 \dots$.

Permutation and transposition

Definition (Signature of a permutation)

Let $\sigma \in \mathcal{S}_n$ We call **signature** of a permutation σ the real number, denoted $\epsilon(\sigma)$ defined by:

$$\epsilon(\sigma) = \prod_{1 \leq i < j \leq n} \frac{\sigma(j) - \sigma(i)}{(j - i)}$$

NB : what values can the signature take ?

Permutation and transposition

Definition (Signature of a permutation)

Let $\sigma \in \mathcal{S}_n$ We call **signature** of a permutation σ the real number, denoted $\epsilon(\sigma)$ defined by:

$$\epsilon(\sigma) = \prod_{1 \leq i < j \leq n} \frac{\sigma(j) - \sigma(i)}{(j - i)}$$

NB : what values can the signature take ? $\epsilon(\sigma) \in \{-1, 1\}$
what about compositions of signatures ?

Permutation and transposition

Definition (Signature of a permutation)

Let $\sigma \in \mathcal{S}_n$ We call **signature** of a permutation σ the real number, denoted $\epsilon(\sigma)$ defined by:

$$\epsilon(\sigma) = \prod_{1 \leq i < j \leq n} \frac{\sigma(j) - \sigma(i)}{(j - i)}$$

NB : what values can the signature take ? $\epsilon(\sigma) \in \{-1, 1\}$
what about compositions of signatures ? $\epsilon(\sigma \circ \sigma') = \epsilon(\sigma)\epsilon(\sigma')$
Exercise : prove it !

Proposition

The signature of a transposition is -1.

Signature of a permutation / Recap

- Permutation σ of the elements of a finite set X
- $\mathcal{S}_n$ the set of permutations of integers in $[1, n]$. NB: $Card(S_n) = n!$.
- **Transposition** $\tau_{i,j}$
- the signature of a permutation can be defined from its decomposition into a product of T transpositions: $\epsilon(\sigma) = (-1)^T$
- **Inversion** of pairs of elements $x, y \in S$: if $x < y$ and $\sigma(x) > \sigma(y)$
- alternatively, we can write the signature of a permutation as : $\epsilon(\sigma) = (-1)^{N(\sigma)}$ where $N(\sigma)$ is the number of inversions in σ

Signature of a permutation / Recap

- Permutation σ of the elements of a finite set X
- $\mathcal{S}_n$ the set of permutations of integers in $[1, n]$. NB: $Card(S_n) = n!$.
- **Transposition** $\tau_{i,j}$
- the signature of a permutation can be defined from its decomposition into a product of T transpositions: $\epsilon(\sigma) = (-1)^T$
- **Inversion** of pairs of elements $x, y \in S$: if $x < y$ and $\sigma(x) > \sigma(y)$
- alternatively, we can write the signature of a permutation as : $\epsilon(\sigma) = (-1)^{N(\sigma)}$ where $N(\sigma)$ is the number of inversions in σ

But we wanted to talk about determinants ?!

Signature of a permutation / Recap

- Permutation σ of the elements of a finite set X
- $\mathcal{S}_n$ the set of permutations of integers in $[1, n]$. NB: $Card(S_n) = n!$.
- **Transposition** $\tau_{i,j}$
- the signature of a permutation can be defined from its decomposition into a product of T transpositions: $\epsilon(\sigma) = (-1)^T$
- **Inversion** of pairs of elements $x, y \in S$: if $x < y$ and $\sigma(x) > \sigma(y)$
- alternatively, we can write the signature of a permutation as : $\epsilon(\sigma) = (-1)^{N(\sigma)}$ where $N(\sigma)$ is the number of inversions in σ

But we wanted to talk about determinants ?!

Definition (Determinant / version with signature of permutations)

Let $A = (a_{ij})$ be a $n \times n$ matrix. We define its **determinant** as:

$$det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) a_{\sigma(1),1}, \ldots a_{\sigma(n),n} = \sum_{\sigma \in \mathcal{S}_n} \epsilon(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}$$

Multilinear, alternating map

Definition (Multilinear form)

A **multilinear form** on a vector space E over a field $\mathbb{K}$ is a map $f\colon E^n \to \mathbb{K}$ that is separately $\mathbb{K}$-linear in each of its n arguments, i.e. for each $i \in [\![1, n]\!]$ and $(u_1, u_2, ...u_n) \in E^n$, the map $f_i : E \to \mathbb{K}$ defined as follows is linear:

$$x \mapsto f(u_1, \ldots, u_{i-1}, x, u_{i+1}, \ldots, u_n)$$

Definition (Alternated form)

Let $f : E^n \to \mathbb{K}$. It is **alternating** if :

$$\forall (u_1, \ldots u_n) \in E^n,$$

$$[\exists (i, j) \in [\![1, n]\!]^2, (i \neq j) \text{ and } u_i = u_j] \implies f(u_1, \ldots u_n) = 0$$

Determinant: definition via properties

Proposition

Let E be a $\mathbb{K}$-vector space of dimension n and $\mathcal{B}$ a basis of E. There exist a unique map, denoted $det_{\mathcal{B}}$ such that

1. *$det_{\mathcal{B}}$ is multilinear*
2. *$det_{\mathcal{B}}$ alternating, and*
3. *$det_{\mathcal{B}} = 1$.*

Determinant: definition via properties

Proposition

Let E be a $\mathbb{K}$-vector space of dimension n and $\mathcal{B}$ a basis of E. There exist a unique map, denoted $det_{\mathcal{B}}$ such that

1. *$det_{\mathcal{B}}$ is multilinear*
2. *$det_{\mathcal{B}}$ alternating, and*
3. *$det_{\mathcal{B}} = 1$.*

Link with the determinant of a matrix A: if we denote the columns of A $C_1, C_2, ...C_n$ then

$$det(A) = det_{\mathcal{B}}(C_1, C_2, ...C_n)$$

Determinant: definition via properties / Recap

Proposition

1. *The determinant of the identity matrix is 1.*
2. *The exchange of two rows multiplies the determinant by* -1.
3. *Multiplying a row (or column) by a scalar multiplies the determinant by this scalar.*
4. *Adding a multiple of one row (resp. column) to another row (resp. column) does not change the determinant.*

Laplace expansion (cofactor expansion)

Notation. For $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{K})$, let M_{ij} be the $(n-1) \times (n-1)$ matrix obtained by deleting row i and column j from A. The cofactor is $C_{ij} = (-1)^{i+j} \det\big(M_{ij}\big)$.

Proposition (Developing along column j (fixed j))

$$\det(A) = \sum_{i=1}^{n} a_{ij}\, C_{ij} \;=\; \sum_{i=1}^{n} (-1)^{i+j}\, a_{ij}\, \det\big(M_{ij}\big).$$

Proposition (Developing along row i (fixed i))

$$\det(A) = \sum_{j=1}^{n} a_{ij}\, C_{ij} \;=\; \sum_{j=1}^{n} (-1)^{i+j}\, a_{ij}\, \det\big(M_{ij}\big).$$

Properties of the determinant

Proposition

Let A, B be square matrices of size $n \times n$.

- $\det(\mathbf{I}) = 1$
- $\det\left(\mathbf{A}^\top\right) = \det(\mathbf{A})$
- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ *i.e. The determinant of a product of matrices is the product of their determinants.*
- $\det(\alpha\mathbf{A}) = \alpha^n \det(\mathbf{A})$
- *The determinant is linear with respect to one row, given the other rows are fixed.*

Example

Exercise session: use these properties to compute determinants.

Determinant and invertibility

Proposition (Basis change and determinant)

Let $\mathcal{B}, \mathcal{B}'$ be two bases of a vector space E, and $(v_1, \dots v_n) \in E^n$. Then:

$$det_{\mathcal{B}'}(v_1, \dots v_n) = det_{\mathcal{B}'}(\mathcal{B}) det_{\mathcal{B}}(v_1, \dots v_n)$$

i.e. with P is the change-of-basis matrix whose j-th column is $[v_j]_{\mathcal{B}}$,

$$\det_{\mathcal{B}'}(v_1, \dots, v_n) = \underbrace{\det(P)}_{=\det_{\mathcal{B}'}(\mathcal{B})} \det_{\mathcal{B}}(v_1, \dots, v_n)$$

Determinant and invertibility

Proposition (Basis change and determinant)

Let $\mathcal{B}, \mathcal{B}'$ be two bases of a vector space E, and $(v_1, \dots v_n) \in E^n$. Then:

$$det_{\mathcal{B}'}(v_1, \dots v_n) = det_{\mathcal{B}'}(\mathcal{B}) det_{\mathcal{B}}(v_1, \dots v_n)$$

i.e. with P is the change-of-basis matrix whose j-th column is $[v_j]_{\mathcal{B}}$,

$$\det_{\mathcal{B}'}(v_1, \dots, v_n) = \underbrace{\det(P)}_{=\det_{\mathcal{B}'}(\mathcal{B})} \det_{\mathcal{B}}(v_1, \dots, v_n)$$

Proposition

Let $\mathcal{B}$ be a basis of E, and $(v_1, v_2, \dots v_n) \in E^n$. Then $(v_1, v_2, \dots v_n)$ is a basis of E if and only if $det_B(v_1, v_2, \dots v_n) \neq 0$.

Proof : $\implies$ then $\impliedby$ by contradiction

Determinant and invertibility

Reminder: if A is invertible, it corresponds to an isomorphism.

Proposition

$$(A \textit{ is } \textbf{invertible}) \;\Leftrightarrow\; \det(A) \neq 0$$

$$(A \textit{ is } \textbf{singular}) \;\Leftrightarrow\; \det(A) = 0$$

Proposition

If A is invertible:

$$\det\left(A^{-1}\right) = \frac{1}{\det(A)}$$

NB: We have

$$det(AA^{-1}) = det(A)det(A^{-1}) \text{ and } det(AA^{-1}) = det(I_n) = 1$$

Reminder : Theorem of basis change

Theorem (Theorem of basis change)

Let $\mathcal{B} = (u_1, u_2, ...u_n)$ *and* $\mathcal{B}' = (v_1, v_2, ...v_n)$ *be two bases of a vector space* E, $L : E \mapsto E$ *a linear map,* $A = [L]_{\mathcal{B}}$ *the matrix of* L *in* $\mathcal{B}$ *and* $B = [L]_{\mathcal{B}'}$ *the matrix of* L *in* $\mathcal{B}'$.

Reminder : Theorem of basis change

Theorem (Theorem of basis change)

Let $\mathcal{B} = (u_1, u_2, ...u_n)$ and $\mathcal{B}' = (v_1, v_2, ...v_n)$ be two bases of a vector space E, $L : E \mapsto E$ a linear map, $A = [L]_{\mathcal{B}}$ the matrix of L in $\mathcal{B}$ and $B = [L]_{\mathcal{B}'}$ the matrix of L in $\mathcal{B}'$.
Let P be the matrix such that the j^{th} column is $[v_j]_{\mathcal{B}}$, the coordinates of basis vector v_j of $\mathcal{B}'$ in the basis $\mathcal{B}$.

$$P = [[v_1]_{\mathcal{B}} \ldots [v_n]_{\mathcal{B}}], \text{ so that } [x]_{\mathcal{B}} = P[x]_{\mathcal{B}'}$$

Then P is invertible and we have

$$B = P^{-1}AP.$$

Determinant of an endomorphism

Proposition

The determinant is invariant under matrix similarity.

From the theorem of basis change, we write:

$$det(B) = det(P^{-1})det(A)det(P) = det(A)det(P)det(P^{-1})$$

$$= det(A)det(PP^{-1}) = det(A)det(I_n) = det(A)$$

In other words, given a linear endomorphism of a finite-dimensional vector space, the determinant of the matrix that represents this linear endomorphism on a given basis does not depend on this chosen basis. **This allows defining the determinant of a linear endomorphism, which does not depend on the choice of a coordinate system**.

Determinant of an endomorphism

Proposition

Let f and g two linear applications from E to E. Then:

- *f is bijective iff $det(f) \neq 0$.*
- $\det(f \circ g) = det(f)det(g)$
- *If $\mathcal{B} = (v_1, \ldots v_n)$ is a basis of E, then*

$$det_{\mathcal{B}}(f(v_1), \ldots f(v_n)) = det_{\mathcal{B}}(f)det_{\mathcal{B}}(v_1, \ldots v_n)$$

- *If f is bijective, then $det(f^{-1}) = \frac{1}{det(f)}$.*

Table of Contents

Determinant of a diagonal matrix

Proposition

The determinant of a diagonal matrix is the product of its diagonal entries.

$$det(\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)) = \prod_{i=1}^{n} \lambda_i$$

Corollary

A diagonal matrix $D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *is invertible if and only if all its diagonal coefficients are not null.*

Determinant of a triangular matrix

Proposition

The determinant of a triangular matrix is the product of its diagonal entries.

Exercise : show this property. Hint : proof by induction over n the size of the matrix.

Determinant of an orthogonal matrix

Reminder: Definition of an orthogonal matrix

Determinant of an orthogonal matrix

Reminder: Definition of an orthogonal matrix
If U is orthogonal, then $U^\top U = I_n$. So $U^{-1} = U^\top$. Taking determinants:

$$\det\left(U^\top\right)\det(U) = 1$$

Since $\det\left(U^\top\right) = \det(U)$, we have:

$$(\det(U))^2 = 1 \quad \Rightarrow \quad \det(U) = \pm 1$$

Proposition

The determinant of an orthogonal matrix $U \in \mathcal{M}_n(\mathbb{R})$ is either 1 or -1.

NB: $\det(U) \in \{+1, -1\}$ means that orientation is preserved if $+1$, flipped if -1.

Why the name "orthogonal" matrix

Reminder: Scalar product of two vectors of $\mathbb{R}^n$.

Proposition

Let us consider $U \in \mathbb{R}^{n \times n}$ an orthogonal matrix. Columns (and rows) of Q form an orthonormal set.

Example

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

What is the inverse of U ? Verify orthogonality of the column vectors of U.

Classic example: rotation in $\mathbb{R}^2$

Example

For an angle $\theta \in \mathbb{R}$, we define the rotation matrix as

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

- $R(\theta)$ rotates every vector in $\mathbb{R}^2$ counterclockwise by θ.
- $\det R(\theta) = \cos^2\theta + \sin^2\theta = 1$ (so orientation is preserved).
- $R(0) = I_2, \quad R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$.

Classic example: rotation in $\mathbb{R}^2$

Example

For an angle $\theta \in \mathbb{R}$, we define the rotation matrix as

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

- $R(\theta)$ rotates every vector in $\mathbb{R}^2$ counterclockwise by θ.
- $\det R(\theta) = \cos^2\theta + \sin^2\theta = 1$ (so orientation is preserved).
- $R(0) = I_2, \quad R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$.

Actually useful in ML ! e.g. see RoPE paper.

Verifying $R(\theta)$ is orthogonal

Compute $R(\theta)^\top R(\theta)$:

$$R(\theta)^\top R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Verifying $R(\theta)$ is orthogonal

Compute $R(\theta)^\top R(\theta)$:

$$R(\theta)^\top R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} \cos^2\theta + \sin^2\theta & -\cos\theta\sin\theta + \sin\theta\cos\theta \\ -\sin\theta\cos\theta + \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{pmatrix}$$

Verifying $R(\theta)$ is orthogonal

Compute $R(\theta)^\top R(\theta)$:

$$R(\theta)^\top R(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} \cos^2\theta + \sin^2\theta & -\cos\theta\sin\theta + \sin\theta\cos\theta \\ -\sin\theta\cos\theta + \cos\theta\sin\theta & \sin^2\theta + \cos^2\theta \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2.$$

Hence $R(\theta)^{-1} = R(\theta)^\top = R(-\theta)$.

Orthonormal columns (Lecture 5)

The columns of $R(\theta)$ are

$$u_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \qquad u_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

We have

$$\|u_1\|_2^2 = \cos^2\theta + \sin^2\theta = 1, \quad \|u_2\|_2^2 = \sin^2\theta + \cos^2\theta = 1,$$

$$u_1^\top u_2 = -\cos\theta\sin\theta + \sin\theta\cos\theta = 0.$$

Thus the columns of $R(\theta)$ form an orthonormal basis of $\mathbb{R}^2$.

Orthogonal matrix and norms (Lecture 5)

Proposition

Inner products and norms are preserved when applying a multiplication by an orthogonal matrix.

$$(Qx)\cdot(Qy) = x\cdot y, \qquad \|Qx\|_2 = \|x\|_2.$$

Exercise : proof (Lecture 5)

Example (Numerical example ($\theta = \pi/6$))

$$R\left(\tfrac{\pi}{6}\right) = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}, \qquad R\left(\tfrac{\pi}{6}\right)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}.$$

$$\left\| R\left(\tfrac{\pi}{6}\right)\begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{\left(\frac{\sqrt{3}}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = 1 \quad \text{(norm preserved)}.$$

Orthogonal matrices

Example

More examples of orthogonal matrices

- **Reflections** (determinant -1), e.g. $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ (reflection across the x-axis).
- **Permutation matrices**, e.g. $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (swap coordinates).

Summary

- Q orthogonal $\iff Q^\top Q = I \iff Q^{-1} = Q^\top$.
- Orthogonal matrices preserve lengths, angles, and dot products.
- In $\mathbb{R}^2$, $R(\theta)$ is orthogonal and models rotation by θ.
- Orthogonal matrices have determinant ± 1 (rotations vs. reflections).

Table of Contents

Eigenvalues and eigenvectors

Definition

Let A be an $n \times n$ matrix. A nonzero vector $v \in \mathbb{R}^n$ is called an *eigenvector* of A if there exists $\lambda \in \mathbb{R}$ such that:

$$Av = \lambda v.$$

The scalar λ is called an *eigenvalue* of A associated with v.

Example

Find eigenvalues of I_n and $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots \lambda_n)$

Eigenvalues and eigenvectors

Definition

Let A be an $n \times n$ matrix. A nonzero vector $v \in \mathbb{R}^n$ is called an *eigenvector* of A if there exists $\lambda \in \mathbb{R}$ such that:

$$Av = \lambda v.$$

The scalar λ is called an *eigenvalue* of A associated with v.

Example

Find eigenvalues of I_n and $D = \text{diag}(\lambda_1, \lambda_2, \ldots \lambda_n)$

- Any vector $x \in \mathbb{R}^n$ satisfies $I_n x = 1 \cdot x$, i.e. is associated with the eigenvalue 1 of I_n.
- Let e_i be the $i - th$ vector in the standard basis of $\mathbb{R}^n$. We remark that e_i is an eigenvector of D associated with the eigenvalue λ_i.

Geometrical interpretation

Example

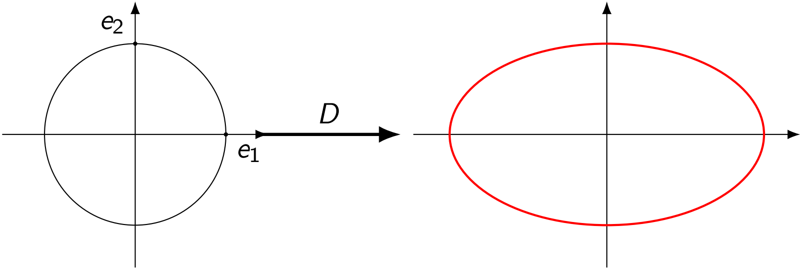
In 2D : represent v an eigenvector of matrix A associated with eigenvalue λ. Discuss according to the value of λ.

Geometrical interpretation

Example

In $\mathbb{R}^2$, how is the unit circle transformed via D?

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{Unit circle } \left\{ x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;\middle|\; x_1^2 + x_2^2 = 1 \right\}.$$



Geometrical interpretation

The circle is transformed into the red ellipse.

Let

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = Ax \quad \text{with } x \in \text{unit circle } \big(\text{i.e. } x_1^2 + x_2^2 = 1\big), \qquad y = \begin{pmatrix} 2x_1 \\ x_2 \end{pmatrix}.$$

Hence $x_1 = \frac{y_1}{2}$, $x_2 = y_2$, and substituting into $x_1^2 + x_2^2 = 1$ gives

$$\boxed{\frac{y_1^2}{4} + y_2^2 = 1} \qquad \text{(Equation of an ellipsoid).}$$

Eigenspace and eigenspectrum

Definition (Eigenspace)

The set of all eigenvectors of A corresponding to the same eigenvalue λ, together with the zero vector, is called an **eigenspace**

Exercise : prove that it is a subspace.
NB : If a set of eigenvectors of A forms a basis of the domain of A, then this basis is called an **eigenbasis**.

Definition (Eigenspectrum)

The **eigenspectrum** (or **spectrum**) of a matrix is the list of its eigenvalues, repeated according to their multiplicity.

NB: We will see that an important quantity associated with the spectrum is the maximum absolute value of any eigenvalue. This is known as the **spectral radius** of the matrix.

Diagonalization

Proposition

If A has n linearly independent eigenvectors $\{v_1, \ldots, v_n\}$ with associated eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$, then these eigenvectors form a basis of $\mathbb{R}^n$ and A is **diagonalizable***:*

$$A = VDV^{-1}$$

where $V = (v_1 \ \ldots \ v_n)$ and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

NB: with the above notations, we can also write

$$AV = VD$$

Diagonalization

In other words, if $V \in \mathcal{M}_{n\times n}(\mathbb{R})$ is composed of n linearly independent vectors:

$$\mathrm{span}(v_1, \dots, v_n) = \mathbb{R}^n \;\Rightarrow\; \mathrm{rank}(V) = n \;\Rightarrow\; V \text{ is an invertible matrix.}$$

From $AV = VD$ we get

$$AV = VD \iff AVV^{-1} = VDV^{-1} \iff A = V \underset{\text{Diagonal}}{D} V^{-1}.$$

We say that the matrix A is **diagonalizable**.

Computing AV in detail

$$AV = A\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} Av_1 & \cdots & Av_n \end{pmatrix} = \begin{pmatrix} \lambda_1 v_1 & \cdots & \lambda_n v_n \end{pmatrix} = VD.$$

$$V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}, \qquad D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix},$$

$$VD = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1 v_1 & \cdots & \lambda_n v_n \\ | & & | \end{bmatrix}.$$

$$(VD)_{ij} = \sum_{k=1}^{n} V_{ik} D_{kj} = V_{ij}\lambda_j, \qquad (AV)_{ij} = (Av_j)_i = (\lambda_j v_j)_i = V_{ij}\lambda_j,$$

hence $AV = VD$.

Notes around $AV = VD$

Warning

$$\boxed{VD \neq DV}$$

- $V\boldsymbol{D}$ *multiplies the columns* of V by the diagonal entries d_i.
- $\boldsymbol{D}V$ *multiplies the rows* of V by the diagonal entries d_i.

Exercise: Powers of a matrix with a basis of eigenvectors

Example

Let $\{(\lambda_i, v_i)\}_{i=1}^{n}$ be eigenvalue–eigenvector pairs of a matrix $A \in \mathbb{R}^{n\times n}$ such that $\{v_1, \ldots, v_n\}$ are linearly independent. Identify the eigenvalues and eigenvectors of $A^2, \ldots, A^k$ for $k \in \mathbb{N}$. Are these matrices diagonalizable?

Exercise: Powers of a matrix with a basis of eigenvectors

Since $Av_i = \lambda_i v_i$ and the v_i's form a basis, we have, by induction on k,

$$A^k v_i \;=\; \lambda_i^k v_i \qquad \text{for } i = 1, \ldots, n,\; k \in \mathbb{N}.$$

Hence for every $k \in \mathbb{N}$ the vectors $v_1, \ldots, v_n$ are eigenvectors of A^k and the corresponding eigenvalues are $\lambda_1^k, \ldots, \lambda_n^k$.
Because the same n eigenvectors remain linearly independent, A^k is diagonalizable: if $V = (v_1 \;\cdots\; v_n)$ and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then

$$A = VDV^{-1} \quad\Longrightarrow\quad A^k = VD^kV^{-1} = V \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)\, V^{-1}.$$

Remarks.

- If some $\lambda_i = 0$, then 0 is an eigenvalue of A^k;
- for $k = 0$ we get $A^0 = I$ with eigenvalues all equal to 1 and the same eigenvectors v_i.

Next class

- Positive definite matrices, positive semi-definite matrices, Gram matrix
- Spectral theorem
- Polar decomposition
- Singular value decomposition

Prepare by reading Chapter 4 of the MML book (4.1 to 4.4 to review determinants and eigendecomposition)