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 transormations. Rank, image, kernel
- (Lecture 2) Matrices, link with linear transformations, linear systems
- (Lecture 3) Determinant, diagonalization, Eigendecomposition (part 1)

In this lecture: Eigendecomposition (part 2), diagonalization, Triangularisability. Inner product spaces,normed spaces.

Table of Contents

Table of Contents

Endomorphisms

Definition

An endomorphism f of a vector space E is a linear map from the space to itself:

$$f : E \to E.$$

NB: we denote by $\mathcal{L}(E)$ the set of endomorphisms on E.

Finite-dimensional case

- When E is finite-dimensional, the study of f reduces to the study of its matrix with respect to a chosen basis.
- The resulting matrix is square; often the same basis of E is used at the source and at the target.

Reduction of endomorphisms

In linear algebra, the reduction of an endomorphism aims to express matrices and endomorphisms in a simpler form, for example to make computations easier.

Method

Reduction essentially consists in decomposing the vector space as a direct sum of invariant subspaces on which the induced endomorphism is simpler:

$$V = U_1 \oplus \cdots \oplus U_k \quad \text{with } f(U_i) \subseteq U_i.$$

NB: Geometrically, this amounts to choosing a basis of the space in which the endomorphism has a simple expression (e.g., block or diagonal form).

Diagonalization

- In finite dimension, diagonalizing an endomorphism means finding a basis in which the matrix of f is diagonal.
- Not every endomorphism is diagonalizable; in some cases one can at best triangularize it, i.e. put it in upper-triangular form.
- Diagonalization is useful for analyzing f, computing powers f^k, searching for square roots of f, etc.

Eigenvalue, eigenvector

Let E be a vector space defined over a field $\mathbb{K}$. For an endomorphism $f \in \mathcal{L}(E)$, there *may* be vectors which, when f is applied to them, are simply scaled by some constant.

Definition

We say that a nonzero vector $\mathbf{x} \in E$ is an **eigenvector** of f corresponding to the **eigenvalue** λ if

$$f(\mathbf{x}) = \lambda \mathbf{x}$$

NB: The zero vector is excluded from this definition because $f(\mathbf{0}) = \mathbf{0} = \lambda \mathbf{0}$ for every $\lambda \in \mathbb{K}$.

Eigenspace and eigenspectrum

Definition (Eigenspectrum)

The **eigenspectrum** (or **spectrum**) of an endomorphism is the list of its eigenvalues, repeated according to their multiplicity. We denote it $sp(f)$.

$$sp(f) = \{\lambda \in \mathbb{K}, \exists v \in E, v \neq 0_E \text{ and } f(v) = \lambda v\}$$

Notations : m_λ the multiplicity of eigenvalue λ.
NB: usually, we write the eigenspectrum as a set and specify the multiplicity of each eigenvalue separately.

Eigenspace

Definition (Eigenspace)

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

Notation: Let $\lambda \in \mathbb{K}$ be an eigenvalue of f. We denote the eigenspace of f associated to the eigenvalue λ by $E_\lambda(f)$ or $Eig(f, \lambda)$.

$$E_\lambda(f) = \{v \in E, f(v) = \lambda v\}$$

Can you express $E_\lambda(f)$ as the kernel of a linear application ?

Eigenspace

Definition (Eigenspace)

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

Notation: Let $\lambda \in \mathbb{K}$ be an eigenvalue of f. We denote the eigenspace of f associated to the eigenvalue λ by $E_\lambda(f)$ or $Eig(f, \lambda)$.

$$E_\lambda(f) = \{v \in E, f(v) = \lambda v\}$$

Can you express $E_\lambda(f)$ as the kernel of a linear application ?

$$E_\lambda(f) = Ker(f - \lambda Id)$$

What can you say about the dimension of this subspace ?

Eigenspace

Definition (Eigenspace)

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

Notation: Let $\lambda \in \mathbb{K}$ be an eigenvalue of f. We denote the eigenspace of f associated to the eigenvalue λ by $E_\lambda(f)$ or $Eig(f, \lambda)$.

$$E_\lambda(f) = \{v \in E, f(v) = \lambda v\}$$

Can you express $E_\lambda(f)$ as the kernel of a linear application ?

$$E_\lambda(f) = Ker(f - \lambda Id)$$

What can you say about the dimension of this subspace ?

$$dim(E_\lambda(f)) \geq 1$$

Eigenspace

Definition (Eigenspace)

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

Notation: Let $\lambda \in \mathbb{K}$ be an eigenvalue of f. We denote the eigenspace of f associated to the eigenvalue λ by $E_\lambda(f)$ or $Eig(f, \lambda)$.

$$E_\lambda(f) = \{v \in E, f(v) = \lambda v\}$$

Can you express $E_\lambda(f)$ as the kernel of a linear application ?

$$E_\lambda(f) = Ker(f - \lambda Id)$$

What can you say about the dimension of this subspace ?

$$dim(E_\lambda(f)) \geq 1$$

We will see that

$$dim(E_\lambda(f)) \leq m_\lambda$$

Manipulating eigenvalues

Proposition

Let $\mathbf{x}$ *be an eigenvector of* $\mathbf{A} \in \mathcal{M}_n(\mathbb{R})$ *with corresponding eigenvalue* λ. *Then*

1. *For any* $\gamma \in \mathbb{R}$, $\mathbf{x}$ *is an eigenvector of* $\mathbf{A} + \gamma\mathbf{I}$ *with eigenvalue* $\lambda + \gamma$.
2. *If* $\mathbf{A}$ *is invertible, then* $\mathbf{x}$ *is an eigenvector of* $\mathbf{A}^{-1}$ *with eigenvalue* λ^{-1}.
3. *If* $\mathbf{A}$ *is invertible,* $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$ *for any* $k \in \mathbb{Z}$ *(where* $\mathbf{A}^0 = \mathbf{I}$ *by definition).*

Manipulating eigenvalues

Proof

(i) By computing

$$(\mathbf{A} + \gamma \mathbf{I})\mathbf{x} = \mathbf{A}\mathbf{x} + \gamma \mathbf{I}\mathbf{x} = \lambda \mathbf{x} + \gamma \mathbf{x} = (\lambda + \gamma)\mathbf{x}$$

(ii) Suppose $\mathbf{A}$ is invertible. Then

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{A}\mathbf{x} = \mathbf{A}^{-1}(\lambda \mathbf{x}) = \lambda \mathbf{A}^{-1}\mathbf{x}$$

Dividing by λ, which is valid because the invertibility of $\mathbf{A}$ implies $\lambda \neq 0$, gives $\lambda^{-1}\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}$.

(iii) Cf previous lecture. The case $k \geq 0$ follows by induction on k. Then the general case $k \in \mathbb{Z}$ follows by combining the $k \geq 0$ case with (ii).

Properties

Proposition

Let $f \in \mathcal{L}(E)$ an endomorphism. Then 0 is an eigenvalue of f if and only if $Ker(f) \neq \{0_E\}$ i.e. f is not injective.

$$0 \in \mathrm{Sp}(f) \iff \ker(f) \neq \{0\} \iff f \text{ is not injective}$$

NB: directly follows from expressing $E_\lambda(f)$ as $Ker(f - \lambda Id)$.

Proposition (Determinant and eigenspace)

Let $f \in \mathcal{L}(E)$ an endomorphism, with E a finite-dimensional vector space. Then

$$\lambda \in sp(f) \text{ iff } det(\lambda Id_E - f) = 0$$

Proof: exercise

Properties

Let's prove the property on the determinant of the linear map $\lambda Id_E - f$.
$(\implies)$ Let $\lambda \in sp(f)$. Assume there exists $u \in E$ with $u \neq 0$ such that $f(u) = \lambda u$. Then

$$\lambda u - f(u) = 0 \iff (\lambda Id_E - f)(u) = 0 \iff u \in \ker(\lambda Id_E - f).$$

Hence $\lambda Id_E - f$ is not bijective (i.e. its kernel contains a nonzero vector), and therefore

$$\det(\lambda Id_E - f) = 0.$$

$(\Longleftarrow)$ If $\det(\lambda Id_E - f) = 0$, then $\lambda Id_E - f$ is not bijective, hence not injective (since E is finite-dimensional).
Thus there exists $u \in \ker(\lambda Id_E - f)$ with $u \neq 0$, which means $f(u) = \lambda u$.
Therefore λ is an eigenvalue of f (i.e., $\lambda \in \mathrm{Sp}(f)$).

Properties

Proposition

Let $f \in \mathcal{L}(E)$ *and let* $\lambda_1, \dots \lambda_p$ *be* p *pairwise-distinct eigenvalues of* f*.*

1. *The sum of subspaces* $E_{\lambda_1}(f) + \cdots + E_{\lambda_p}(f)$ *is a direct sum.*
2. *If* $(u_1, u_2, \dots u_p) \in E^p$ *such that*

$$\forall i \in [\![1, p]\!], u_i \neq 0, \text{ and } f(u_i) = \lambda_i u_i$$

   Then the family $(u_1, u_2, \dots u_p)$ *is linearly independent.*

Proof of (1): by induction. Proof of (2): follows from (1) with the definiton of a direct sum.

Properties

Proposition (Cardinal of the eigenspectrum)

If E is a finite-dimensional vector space, with $dim(E) = n$, and $f \in \mathcal{L}(E)$, then

$$Card(sp(f)) \leq n$$

Proof:

Properties

Proposition (Cardinal of the eigenspectrum)

If E is a finite-dimensional vector space, with $dim(E) = n$, and $f \in \mathcal{L}(E)$, then

$$Card(sp(f)) \leq n$$

Proof: suppose $Card(sp(f)) \geq n+1$. Using the previous property, we would have a linearly independent family of vectors $(u_1, u_2, \dots u_n, u_{n+1})$ associated with eigenvalues $\lambda_1, \dots \lambda_n, \lambda_{n+1}$. But $dim(E) = n$. Absurd.

Properties (bonus exercise)

Proposition

Let $f, g \in \mathcal{L}(E)$ two endomorphisms that commute. Then the eigenspaces of f are stable by g.

NB: Let $\lambda \in sp(f)$. This proposition writes :

$$g(E_\lambda(f)) \subseteq E_\lambda(f)$$

Since f and g are commutative, $\lambda Id_E - f$ and g are commutative too. Writing $E_\lambda(f)$ as $Ker(\lambda Id_E - f)$, the eigenspace of f associated to λ is stable by g.

Characteristic polynomial

We consider a finite-dimensional vector space E, $dim(E) = n$.

Definition

Let $f \in \mathcal{L}(E)$. We call characteristic polynomial of f the polynomial, denoted χ_f, defined as

$$\chi_f = det(\lambda Id_E - f)$$

NB: alternative notation $P_\lambda(f)$. Similarly, for a square matrix $A \in \mathcal{M}_n(\mathbb{K})$, we will denote $\chi_A = det(\lambda I_n - A)$.

Characteristic polynomial

We consider a finite-dimensional vector space E, $dim(E) = n$.

Definition

Let $f \in \mathcal{L}(E)$. We call characteristic polynomial of f the polynomial, denoted χ_f, defined as

$$\chi_f = det(\lambda Id_E - f)$$

NB: alternative notation $P_\lambda(f)$. Similarly, for a square matrix $A \in \mathcal{M}_n(\mathbb{K})$,

we will denote $\chi_A = det(\lambda I_n - A)$.

What is the degree of χ_f ? What is the value of its leading coefficient ?

Characteristic polynomial

We consider a finite-dimensional vector space E, $dim(E) = n$.

Definition

Let $f \in \mathcal{L}(E)$. We call characteristic polynomial of f the polynomial, denoted χ_f, defined as

$$\chi_f = det(\lambda Id_E - f)$$

NB: alternative notation $P_\lambda(f)$. Similarly, for a square matrix $A \in \mathcal{M}_n(\mathbb{K})$,

we will denote $\chi_A = det(\lambda I_n - A)$.
What is the degree of χ_f ? What is the value of its leading coefficient ?
The leading coefficient is 1, we say the polynomial is **monic**, of degree n.

Characteristic polynomial

- Following the property on the determinant of the linear map $\lambda Id_E - f$, a value λ is an eigenvalue of f if and only if it is a root of its characteristic polynomial.

$$\lambda \in \mathrm{Sp}(f) \iff \chi_f(\lambda) = 0, \qquad \chi_f(\lambda) = \det\big(\lambda \, \mathrm{Id}_E - [f]_{\mathcal{B}}\big).$$

- $\chi_A(\lambda)$ is a polynomial of degree n.
- Remark that the leading coefficient of $\chi_A(\lambda)$ is 1 (unitary/monic polynomial).
- If A is the matrix of f in a given basis of E, then we have the equivalence

$$\chi_A(\lambda) = \chi_f(\lambda)$$

This means that the characteristic polynomial of a linear map does not depend on the choice of basis. Like the determinant, it does not vary under matrix similarity (NB: FR "invariant de similitude")

Multiplicity of an eigenvalue

Definition

Let $f \in \mathcal{L}(E)$, $A \in \mathcal{M}_n(\mathbb{K})$, and $\lambda \in sp(f)$. We call **multiplicity** of an eigenvalue λ its multiplicity as a root of the caracteristic polynomial of f (or of A, respectively). It is denoted m_λ.

$$m_\lambda = max(k \in \mathbb{N}^*, (X - \lambda)^k \mid \chi_A(X))$$

NB: For any $\lambda \in sp(f)$,

$$1 \leq m_\lambda \leq n$$

Proposition

Let $\lambda \in sp(f)$. We have $dim(E_\lambda(f)) \leq m_\lambda$. Similarly, for a square matrix $A \in \mathcal{M}_n(\mathbb{K})$, $dim(E_\lambda(A)) \leq m_\lambda$.

Multiplicity of an eigenvalue

Not always equal

The dimension of the eigenspace associated to a given eigenvalue λ is not necessaril egal to the multiplicity of this eigenvalue in the characteristic polynomial.

Example

Example : take the matrix A with zeros everywhere, except on its diagonal and upper diagonal, filled with ones. $sp(A) = \{1\}$, $\chi_A(\lambda) = (\lambda - 1)^n$. $E_1(A) = span((1, 0, \ldots 0)^\top)$ is of dimension 1, whereas the multiplicity of 1 as root of the caracteristic polynomial of A is n.

NB: Equality holds if $m_\lambda = 1$.

Recap - Characteristic polynomial

- The **characteristic polynomial** of an endomorphism of a finite-dimensional vector space is the characteristic polynomial of the matrix of that endomorphism over any basis.
- This means that the characteristic polynomial does not depend on the choice of a basis.
- The characteristic polynomial of a square matrix is a polynomial which is invariant under matrix similarity
- It has the eigenvalues of this matrix as roots.
- The **characteristic equation** is the equation obtained by equating the characteristic polynomial to zero.

Table of Contents

Diagonalizable matrix

In this section we consider a finite-dimensional vector space E, $dim(E) = n$.

Definition

- We say that $f \in \mathcal{L}(E)$ is **diagonalizable** if there exists a basis $\mathcal{B}$ of E formed by eigenvectors of f, i.e. $[f]_{\mathcal{B}}$ the matrix of f in $\mathcal{B}$ is diagonal.
- We say that $A \in \mathcal{M}_n(\mathbb{K})$ is **diagonalizable** if A is similar to a diagonal matrix, i.e. there exists an invertible matrix $P \in \mathcal{M}_n(\mathbb{K})$ and a diagonal matrix D such that

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

Diagonalizable matrix

In this section we consider a finite-dimensional vector space E, $dim(E) = n$.

Definition

- We say that $f \in \mathcal{L}(E)$ is **diagonalizable** if there exists a basis $\mathcal{B}$ of E formed by eigenvectors of f, i.e. $[f]_{\mathcal{B}}$ the matrix of f in $\mathcal{B}$ is diagonal.
- We say that $A \in \mathcal{M}_n(\mathbb{K})$ is **diagonalizable** if A is similar to a diagonal matrix, i.e. there exists an invertible matrix $P \in \mathcal{M}_n(\mathbb{K})$ and a diagonal matrix D such that

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

NB : "Diagonalizing" an endomorphism means finding a basis of E in which the matrix representaiton of f is diagonal.
"Diagonalizing" a matrix $A \in \mathcal{M}_n(\mathbb{K})$ means finding an invertible matrix $P \in \mathcal{M}_n(\mathbb{K})$ and a diagonal matrix $D \in \mathcal{M}_n(\mathbb{K})$ such that $A = PDP^{-1}$.

Properties

Let $f \in \mathcal{L}(E)$.

Proposition

- *Let $\mathcal{B}$ be a basis such that $[f]_{\mathcal{B}} = \mathrm{diag}(\lambda_1, ...\lambda_n)$ Then $sp(f) = \{\lambda_1, \dots \lambda_n\}$.*
- *The following statements are equivalent.*
  1. *f is diagonalizable*
  2. *There exists a basis $\mathcal{B}$ of E such that $[f]_{\mathcal{B}}$ is diagonalizable.*
  3. *For any basis $\mathcal{B}$ of E, $[f]_{\mathcal{B}}$ is diagonalizable.*

NB : (1) $\chi_f(\lambda) = \chi_D(\lambda)$ with $D = \mathrm{diag}(\lambda_1, \dots \lambda_n)$. Since $\chi_D(\lambda) = \prod_{i=1}^{n}(\lambda - \lambda_i)$ we get $sp(f) = \{\lambda_1, \dots \lambda_n\}$.
(2) use property on similar matrices.

Not all matrices are diagonalizable

Example

$$A = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & 1 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & & & \dots & & \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

Not all matrices are diagonalizable

Example

$$A = \begin{bmatrix} 1 & 1 & 0 & ... & 0 & 0 \\ 0 & 1 & 1 & ... & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 \\ ... & & & ... & & \\ 0 & 0 & 0 & ... & 1 & 1 \\ 0 & 0 & 0 & ... & 0 & 1 \end{bmatrix}$$

$sp(A) = \{1\}$. If A was diagonalizable, then A would be similar to the identity matrix. i.e. $D = \text{diag}(1, \dots 1) = I_n$.
But $PI_nP^{-1} = I_n$, so I_n is the only matrix similar to I_n.
This leaves us with $A = I_n$, absurd. So A is not diagonalizable.

Not all matrices are diagonalizable

Example

$$A = \begin{bmatrix} 1 & 1 & 0 & ... & 0 & 0 \\ 0 & 1 & 1 & ... & 0 & 0 \\ 0 & 0 & 1 & ... & 0 & 0 \\ ... & & & ... & & \\ 0 & 0 & 0 & ... & 1 & 1 \\ 0 & 0 & 0 & ... & 0 & 1 \end{bmatrix}$$

$sp(A) = \{1\}$. If A was diagonalizable, then A would be similar to the identity matrix. i.e. $D = \text{diag}(1, \dots 1) = I_n$.
But $PI_nP^{-1} = I_n$, so I_n is the only matrix similar to I_n.
This leaves us with $A = I_n$, absurd. So A is not diagonalizable.

We can generalize on this example : if a square matrix M is diagonalizable and has a unique eigenvalue λ, then this matrix must be similar to λI_n. In particular, if $\lambda = 1$, it must be similar to the identity matrix, so in fact it must *be* the identity matrix.

Equivalent statements on diagonalizability

Proposition

Let $f \in \mathcal{L}(E)$. Then the following statements are equivalent.

1. *f is diagonalizable*
2. $\sum_{\lambda \in sp(f)} dim(E_\lambda(f)) = dim(E)$
3. $\sum_{\lambda \in sp(f)} E_\lambda(f) = E$

Equivalent statements on diagonalizability

Recall that the eigenspaces are in direct sum. Hence

$$dim(\sum_{\lambda \in sp(f)} E_\lambda(f)) = \sum_{\lambda \in sp(f)} dim(E_\lambda(f))$$

and the equivalence between (2) and (3).

If we suppose that (1) f is diagonalizable, then there exists a basis $\mathcal{B}$ of E formed by eigenvectors of f. So $\sum_{\lambda \in sp(f)} E_\lambda(f)$ contains a basis of E. It follows that $\sum_{\lambda \in sp(f)} E_\lambda(f) = E$ (3).

Now suppose (3): we have $E = \bigoplus_{\lambda \in sp(f)} E_\lambda(f)$. So there exists a basis of E adapted to this decomposition of E. This basis is formed by eigenvectors of f... So f is diagonalizable (1).

Split Polynomial

Definition

Let $\mathbb{K}$ be a field. A non-constant polynomial $P \in \mathbb{K}[X]$ *splits over* $\mathbb{K}$ if

$$P(X) = C \prod_{j=1}^{n} (X - a_j),$$

where each $a_j \in \mathbb{K}$ and $C \in \mathbb{K}$. Equivalently, P can be written as a product of degree-1 polynomials with coefficients in $\mathbb{K}$. In this case, C is called the **leading coefficient** of P.

Split Polynomial

Definition

Let $\mathbb{K}$ be a field. A non-constant polynomial $P \in \mathbb{K}[X]$ *splits over* $\mathbb{K}$ if

$$P(X) = C \prod_{j=1}^{n} (X - a_j),$$

where each $a_j \in \mathbb{K}$ and $C \in \mathbb{K}$. Equivalently, P can be written as a product of degree-1 polynomials with coefficients in $\mathbb{K}$. In this case, C is called the **leading coefficient** of P.

Dependence on the Field

The splitting property depends on $\mathbb{K}$. For example,

$$X^2 + 1 = (X - i)(X + i) \quad \text{splits over } \mathbb{C},$$

but it does not split over $\mathbb{R}$.

Simple Roots

Definition

A split polynomial P *has simple roots* if, in the previous factorization, the numbers $a_1, \ldots, a_n$ are pairwise distinct (i.e., each root has multiplicity 1).

Example

Over $\mathbb{K} = \mathbb{Q}$ Consider

$$P(X) = 5(X-1)(X+2)(X-3) \in \mathbb{Q}[X].$$

All roots $1, -2, 3$ lie in $\mathbb{Q}$, so P splits over $\mathbb{K}$ with leading coefficient 5.
The roots are pairwise distinct, hence P is split with simple roots.
NB: Expanding,

$$P(X) = 5\left(X^3 - 2X^2 - 5X + 6\right) = 5X^3 - 10X^2 - 25X + 30.$$

Characteristic polynomial and Diagonalizability

Theorem (for endomorphisms)

Let E be a vector space defined over a field $\mathbb{K}$, $dim(E) = n$ and $f \in \mathcal{L}(E)$. The endomorphism f is diagonalizable if and only if

1. *Its characteristic polynomial χ_f is split.*
2. *For each $\lambda \in sp(f)$, $dim(E_\lambda(f)) = m_\lambda$.*

Theorem (for matrices)

Let $A \in \mathcal{M}_n(\mathbb{K})$. It is diagonalizable if and only if

1. *Its characteristic polynomial χ_A is split.*
2. *For each $\lambda \in sp(A)$, $dim(E_\lambda(A)) = m_\lambda$, or equivalently,*

$$rank(\lambda I_n - A) = n - m_\lambda$$

Annihilating Polynomials

Definition

Let E be a finite-dimensional $\mathbb{K}$-vector space with $\dim E = n$. Let $f \in \mathcal{L}(E)$ be an endomorphism and let $P \in \mathbb{K}[X]$ be a polynomial. P is called an *annihilating polynomial* for f if

$$P(f) = 0_{\mathcal{L}(E)} \quad \text{(i.e., the zero endomorphism).}$$

Cayley-Hamilton Theorem

This theorem tells us that the characteristic polynomial of a given endomorphism is an annihilating polynomial for this endomorphism.

Theorem (Cayley-Hamilton)

For any $n \times n$ matrix A over a field K, let

$$\chi_A(\lambda) = \det(\lambda I_n - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$$

be its characteristic polynomial. Then

$$\chi_A(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I_n = 0.$$

NB: notice that the polynomial $det(\lambda I_n - A)$ is null when substituting λ with A.

Cayley-Hamilton Theorem

Existence of an Annihilating Polynomial

Since $\dim \mathcal{L}(E) = n^2$, the family

$$\left(\mathrm{Id}_E,\ u,\ u^2,\ \ldots,\ u^{n^2}\right)$$

of $n^2 + 1$ endomorphisms is linearly dependent. Hence there exist $a_0, \ldots, a_{n^2} \in K$, not all zero, such that

$$a_0 \, \mathrm{Id}_E + a_1 u + \cdots + a_{n^2} u^{n^2} \ = \ 0.$$

Setting $P(X) = a_0 + a_1 X + \cdots + a_{n^2} X^{n^2}$ yields a nonzero polynomial with $P(u) = 0$.

Example: Characteristic polynomial in 2×2

Computation (using $\chi_A(\lambda) = \det(\lambda I_2 - A)$) Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$\lambda I_2 - A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda - a & -b \\ -c & \lambda - d \end{pmatrix}.$$

Hence

$$\begin{aligned} \chi_A(\lambda) = \det \begin{pmatrix} \lambda - a & -b \\ -c & \lambda - d \end{pmatrix} &= (\lambda - a)(\lambda - d) - (-b)(-c) \\ &= \lambda^2 - (a + d)\lambda + (ad - bc). \end{aligned}$$

Example: Characteristic polynomial in 2×2

Computation (using $\chi_A(\lambda) = \det(\lambda I_2 - A)$) Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$\lambda I_2 - A = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda - a & -b \\ -c & \lambda - d \end{pmatrix}.$$

Hence

$$\begin{aligned} \chi_A(\lambda) = \det \begin{pmatrix} \lambda - a & -b \\ -c & \lambda - d \end{pmatrix} &= (\lambda - a)(\lambda - d) - (-b)(-c) \\ &= \lambda^2 - (a+d)\lambda + (ad - bc). \end{aligned}$$

Identification
Since $\mathrm{tr}(A) = a + d$ and $\det(A) = ad - bc$, we obtain

$$\chi_A(\lambda) = \lambda^2 - (\mathrm{tr}\, A)\,\lambda + \det(A).$$

Example: Characteristic polynomial in 2×2

Now let us suppose A of size 2×2 is invertible. Using the previous formula, find an explicit formula for the inverse of A.

Cayley-Hamilton theorem - Examples

Example

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}, \qquad \chi_A(\lambda) = (\lambda - 2)(\lambda - 3) = \lambda^2 - 5\lambda + 6.$$

Compute $A^2 = \begin{pmatrix} 4 & 5 \\ 0 & 9 \end{pmatrix}$. Then

$$A^2 - 5A + 6I_2 = \begin{pmatrix} 4 & 5 \\ 0 & 9 \end{pmatrix} - \begin{pmatrix} 10 & 5 \\ 0 & 15 \end{pmatrix} + \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix} = 0.$$

Bonus: Compute the inverse from Cayley-Hamilton. $\det(A) = 6$, $\operatorname{tr}(A) = 5$, hence

$$A^{-1} = \frac{1}{6}(5I_2 - A) = \begin{pmatrix} \frac{1}{2} & -\frac{1}{6} \\ 0 & \frac{1}{3} \end{pmatrix}.$$

Table of Contents

Trigonalizability

Definition

- Let $f \in \mathcal{L}(E)$. We say that f is **trigonalizable** if there exists a basis $\mathcal{B}$ of E such that $[f]_{\mathcal{B}}$ the matrix of f in $\mathcal{B}$, is triangular.
- Let $A \in \mathcal{M}_n(\mathbb{K})$. We say that A is **trigonalizable** if A is <u>similar</u> to a triangular matrix.

NB: we often consider triangular superior matrices.
If $f \in \mathcal{L}(E)$, and $\mathcal{B} = (e_1, e_2, \ldots e_n)$ is a basis of E, then $\mathcal{M}_{\mathcal{B}}(f)$ is triangular superior if and only if

$$\forall i \in [[1, n]], f(e_i) \in span(e_1, e_2, ...e_i)$$

Warning

A diagonalizable matrix is trigonalizable, but not reciprocally.

Trigonalizability

Trigonalizing a linear map f means finding a basis $\mathcal{B}$ such that $[f]_{\mathcal{B}}$ is triangular.
Trigonalizing a matrix $A \in \mathcal{M}_n(\mathbb{K})$ consists in finding P invertible and T triangular such that $A = PTP^{-1}$.

Theorem (Criterion of Triangularisability)

Let $f \in \mathcal{L}(E)$. Then f is trigonalizable if and only if its characteristic polynomial χ_f is split. (FR: scindé) Similarly, a square matrix A is trigonalizable if and only if its characteristic polynomial χ_A is split.

Proof : by induction. NB : all matrices of $\mathcal{M}_n(\mathbb{C})$ are trigonalizable because all the polynomials from $\mathbb{C}[X]$ are split, but all real valued matrices are not trigonalizable.

Trigonalizability, trace and determinant

Proposition

Let $f \in \mathcal{L}(E)$ $(A \in \mathcal{M}_n(\mathbb{K})$, *respectively). Then*

1. $\sum_{\lambda \in sp(f)} m_\lambda = n$
2. $det(f) = \prod_{\lambda \in sp(f)} \lambda^{m_\lambda}$
3. $tr(f) = \sum_{\lambda \in sp(f)} m_\lambda \lambda$

Method - Triangularisability

Let $A \in \mathcal{M}_n(\mathbb{K})$ such that χ_A is split. Let us denote $sp(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ with multiplicity $m_1, m_2, \dots m_p$. NB : we know that $\sum_{i=1}^{p} \lambda_i = n$.
For each $i \in [[1, p]]$ we look for a basis $\mathcal{B}_i$ of $ker(\lambda_i I_n - A)$.
If $Card(\mathcal{B}_i) = m_i$ (i.e. $dim(ker(\lambda_i I_n - A)) = m_i$) then we can diagonalize.
Otheriwse, we look for vectors to add to $\mathcal{B}_i$ to make it a basis of $ker((\lambda_i I_n - A)^2)$, etc.
Finally, the family of vectors $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \mathcal{B}_p$ is a basis where A can be trigonalized.

Nilpotent matrices

Definition

A **nilpotent** matrix N is a square matrix such that there exist a positive integer k such that

$$N^k = 0$$

NB: the smallest power k for which N^k is null is sometimes called the **index** of N.

Example

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

N is nilpotent with index 2, since $N^2 = 0$.

We can prove that any triangular matrix $T \in \mathcal{M}_n(\mathbb{R})$
with zeros along the main diagonal is nilpotent, with index $\leq n$.

Nilpotent matrices

Example

$$B = \begin{bmatrix} 0 & 2 & 1 & 6 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B^2 = \begin{bmatrix} 0 & 0 & 2 & 7 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad B^3 = \begin{bmatrix} 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad B^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore B is nilpotent, with index 4.

Nilpotent matrices

What to use nilpotent matrices for ?

Example

Triangularisability then decomposition of an upper triangular matrix as the sum of a diagonal matrix and of a nilpotent matrix.

NB: The determinant and trace of a nilpotent matrix are always zero. So a nilpotent matrix cannot be invertible.

Cayley-Hamilton theorem - Examples

Example (Jordan block)

$$J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \chi_J(\lambda) = (\lambda - 1)^3.$$

$$\chi_J(J) = 0 \iff (J - I_3)^3 = 0 \iff J^3 - 3J^2 + 3J - I_3 = 0.$$

Writing $N = J - I_3$ (nilpotent with $N^3 = 0$), any power J^m reduces to a polynomial of degree ≤ 2 in N:

$$J^m = (I_3 + N)^m = I_3 + mN + \binom{m}{2} N^2 \quad (N^3 = 0).$$

This illustrates how we can use C–H to compute powers of a matrix even when the matrix is not diagonalizable.

Practical implications of Cayley-Hamilon theorem

- **Reduce powers.** Any A^m ($m \geq n$) can be rewritten as a linear combination of $I, A, \ldots, A^{n-1}$ using $\chi_A(A) = 0$.
- **Compute inverses.** If $\det(A) \neq 0$, C–H yields a polynomial expression for A^{-1} in terms of $I, A, \ldots, A^{n-1}$ (e.g., explicit in the 2×2 example we saw previously).
- **Minimal polynomial.** The minimal polynomial μ_A divides χ_A; in particular, every eigenvalue of A is a root of χ_A.
- **Linear recurrences.** For vectors $x_k = A^k x_0$, the sequence satisfies the recurrence with coefficients from χ_A (useful in control theory... or in deep learning!).

Example exercise

Example

Consider the following matrix, with $a > 0$:

$$A = \begin{bmatrix} -1 & a & -a \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

1. Compute χ_A and show that A is not diagonalizable.
2. Find v_1, v_2, v_3 in $\mathcal{M}_{3,1}(\mathbb{R})$ such that $Av_1 = -v_1$, $Av_2 = v_1 - v_2$, $Av_3 = v_1 + v_2 - v_3$.
3. Show that A is similar to $T = \begin{bmatrix} -1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$
4. Use a nilpotent matrix to compute A^n for any $n \in \mathbb{N}$.

Table of Contents

Inner product

Definition

An **inner product** on a real vector space V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying, for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all $\alpha, \beta \in \mathbb{R}$:

1. (symmetric) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
2. (bilinear) $\langle \alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle$
3. (positive-definite) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{0_V}$

Vocabulary: A vector space endowed with an inner product is called an **inner product space**, or a **pre-Hilbert space** (FR: *espaces préhilbertiens*).
NB: in the above definition we have stated linearity in the first slot; with symmetry this implies linearity in the second.

Scalar product (dot product)

The usual **scalar product** $(.|.)$ defined on $\mathbb{R}^n$ is an inner product.

$$(.|.) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, (x_1, x_2, \ldots x_n), (y_1, y_2, \ldots y_n) \mapsto \sum_{i=1}^{n} x_i y_i$$

The inner product on $\mathbb{R}^n$ is also often written $\mathbf{x} \cdot \mathbf{y}$ (hence the alternate name **dot product**).

NB: If $x, y \in \mathbb{R}^n$, the dot product can be expressed as:

$$\langle x, y \rangle = x^\top y.$$

i.e. this inner product is a special case of matrix multiplication where we regard the resulting 1×1 matrix as a scalar.

NB: for $\alpha_1, \ldots \alpha_n \in \mathbb{R}^+$, the map defined on $\mathbb{R}^n$ defined as follows is also an inner product.

$$(x_1, x_2, \ldots x_n), (y_1, y_2, \ldots y_n) \mapsto \sum_{i=1}^{n} \alpha_i x_i y_i$$

Properties of the Dot Product ($\mathbb{R}^n$)

Using the previous definition of the scalar/dot product in $\mathbb{R}^n$, prove the following properties.

Proposition (Properties of the Dot Product)

For all $x, y, z \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$:

1. *Symmetry:* $\langle x, y\rangle = \langle y, x\rangle$.
2. *Homogeneity:* $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$.
3. *Linearity in the first argument:* $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
4. *Linearity in the second argument:* $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.

Proof: exercise.

Another example of inner product

Example with functions

Another example of inner product

Example with functions

Example

Take the vector space of continuous real-valued functions defined over a segment $[a, b]$ of $\mathbb{R}$, $E = \mathcal{C}^0([a, b], \mathbb{R})$. Prove that the map $(.|.)$ defined on E^2 as follows is an inner product .

$$(f|g) = \int_a^b f(x)g(x)\,dx$$

Another example of inner product

Example with functions

Example

Take the vector space of continuous real-valued functions defined over a segment $[a, b]$ of $\mathbb{R}$, $E = \mathcal{C}^0([a, b], \mathbb{R})$. Prove that the map $(.|.)$ defined on E^2 as follows is an inner product .

$$(f|g) = \int_a^b f(x)g(x)\,dx$$

NB: Symmetry, bilinearity, positive-definite.
f^2 is continuous too on $[a, b]$. If $f \in \mathcal{C}^0([a, b])$ and $\int_a^b f(x)^2\,dx = 0$, then by continuity f is the null function on $[a, b]$.

Orthogonal vectors

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be **orthogonal** if their inner product is zero

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0$$

Notation : we can write $\mathbf{x} \perp \mathbf{y}$ for short.

NB: **Orthogonality** generalizes the notion of **perpendicularity** we use in 2D.

Table of Contents

Metric spaces

A **metric space** is a set together with a notion of **distance** between its elements, usually called **points**. The distance is measured by a function called a **metric** or **distance** function.

Definition (Metric)

A **metric** on a set S is a function $d : S \times S \to \mathbb{R}^+$ that satisfies, for all $x, y, z \in S$:

1. (Positivity) $d(x, y) \geq 0$, with equality if and only if $x = y$
2. (Symmetry) $d(x, y) = d(y, x)$
3. (Triangle ineguality) $d(x, z) \leq d(x, y) + d(y, z)$

Metric spaces

A **metric space** is a set together with a notion of **distance** between its elements, usually called **points**. The distance is measured by a function called a **metric** or **distance** function.

Definition (Metric)

A **metric** on a set S is a function $d : S \times S \to \mathbb{R}^+$ that satisfies, for all $x, y, z \in S$:

1. (Positivity) $d(x, y) \geq 0$, with equality if and only if $x = y$
2. (Symmetry) $d(x, y) = d(y, x)$
3. (Triangle inegualilty) $d(x, z) \leq d(x, y) + d(y, z)$

Examples of metric spaces ?

Metric spaces - Examples

Example

- The real numbers with the distance function $d(x, y) = |y - x|$ given by the absolute difference between two real numbers
- For a given $n > 0$, the **Hamming distance** is a metric on the set of the words of length n, e.g. the set of 100-character Unicode strings can be equipped with the Hamming distance, which measures the number of characters that need to be changed to get from one string to another.

NB: The Hamming distance actually comes from information theory, where it is used to count the minimum number of errors that could have transformed one string into the other.

Metric spaces - Examples

Example

The Euclidean plane $\mathbb{R}^2$ can be equipped with many different metrics:

- The usual Euclidean distance from high school :
  $d_2((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
- The Manhattan (or "taxi-cab") distance :
  $d_1((x_1, y_1), (x_2, y_2)) = |x_2 - x_1| + |y_2 - y_1|$
- The maximum distance :
  $d_\infty((x_1, y_1), (x_2, y_2)) = \max\{|x_2 - x_1|, |y_2 - y_1|\}$
- A discrete metric like the following : $d(p, q) = \begin{cases} 0, & \text{if } p = q, \\ 1, & \text{otherwise.} \end{cases}$

NB: More generally, we call **Euclidean space** a vector space defined over $\mathbb{R}$, that has a finite dimension and is endowed with an inner product.

Table of Contents

Normed spaces

Some (but not all !) metric spaces are **normed spaces**. They are defined on vector spaces and endowed with a **norm** function.

Definition (Norm)

A **norm** on a real vector space V is a function $\|\cdot\| : V \to \mathbb{R}^+$ that satisfies

1. $\|\mathbf{x}\| \geq 0$, with equality if and only if $\mathbf{x} = \mathbf{0}_V$
2. $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (the **triangle inequality** again)

for all $\mathbf{x}, \mathbf{y} \in V$ and all $\alpha \in \mathbb{R}$.

A vector space endowed with a norm is called a **normed vector space**, or simply a **normed space**.

Norm induced by an inner product

Remark

Any inner product on V induces a norm on V:

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

NB: the axioms for norms are satisfied under this definition and follow from the axioms for inner products. Therefore any inner product space is also a normed space (and hence also a metric space).

Norm induced by the dot product on $\mathbb{R}^n$

Example

Verify that the two-norm $\|\cdot\|_2$ (Euclidean norm) on $\mathbb{R}^2$ is induced by the dot product.

$$\langle x, x\rangle = \|x\|_2^2.$$

NB: Moreover, for the angle θ between x and y (vectors different from the null vector):

$$\langle x, y\rangle = \|x\|_2 \, \|y\|_2 \cos\theta.$$

Reminder : If $\langle x, y\rangle = 0$, the vectors x and y are said to be *orthogonal*. We see the link with angles here.

Examples of norms on $\mathbb{R}^n$

$$L_1 : \|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

$$L_2 : \|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \quad \text{(Euclidean norm)}$$

$$L_p : \|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \quad (p \geq 1)$$

$$L_\infty : \|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

NB: The 1- and 2-norms are special cases of the p-norm, and the ∞-norm is the limit of the p-norm as p tends to infinity.
We require $p \geq 1$ because the triangle inequality fails to hold for $0 < p < 1$.

Pythagorean Theorem

The Pythagorean theorem generalizes to arbitrary inner product spaces.

Theorem (Pythagorean Theorem)

Let V a finite-dimensional vector space, and $x, y \in V$. If $\mathbf{x} \perp \mathbf{y}$, then

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Proof:

Pythagorean Theorem

The Pythagorean theorem generalizes to arbitrary inner product spaces.

Theorem (Pythagorean Theorem)

Let V a finite-dimensional vector space, and $x, y \in V$. If $\mathbf{x} \perp \mathbf{y}$, then

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Proof: Suppose $\mathbf{x} \perp \mathbf{y}$, i.e. $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. It follows:

$$\|\mathbf{x}+\mathbf{y}\|^2 = \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Cauchy-Schwarz inequality

Proposition

Let V be an inner product space. For all $\mathbf{x}, \mathbf{y} \in V$,

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

NB: this equality holds exactly iff $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent (i.e. are scalar multiples of each other, including the case when at leat one of them is null).

Remarks

- Two different norms will give us two different "measures" of distances.
- Not all metric spaces are vector spaces !
- A key motivation for metrics is that they allow limits to be defined for mathematical objects other than real numbers.
  e.g. we say that a sequence $\{x_n\} \subseteq S$ converges to the limit x if for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for all $n \geq N$.

Equivalence of norms in a finite-dimensional vector space

For any given finite-dimensional vector space V, all norms on V are equivalent in the sense that for all $\mathbf{x} \in V$, for two norms $\|\cdot\|_A$ and $\|\cdot\|_B$, there exist constants $\alpha, \beta > 0$ such that

$$\alpha\|\mathbf{x}\|_A \leq \|\mathbf{x}\|_B \leq \beta\|\mathbf{x}\|_A$$

NB: the constants depend on the norms, not on the vector. **Therefore convergence in one norm implies convergence in any other norm.**

This is not a general property e.g. may not apply in infinite-dimensional vector spaces such as function spaces!

Next class

Putting everything together, spectral theorem