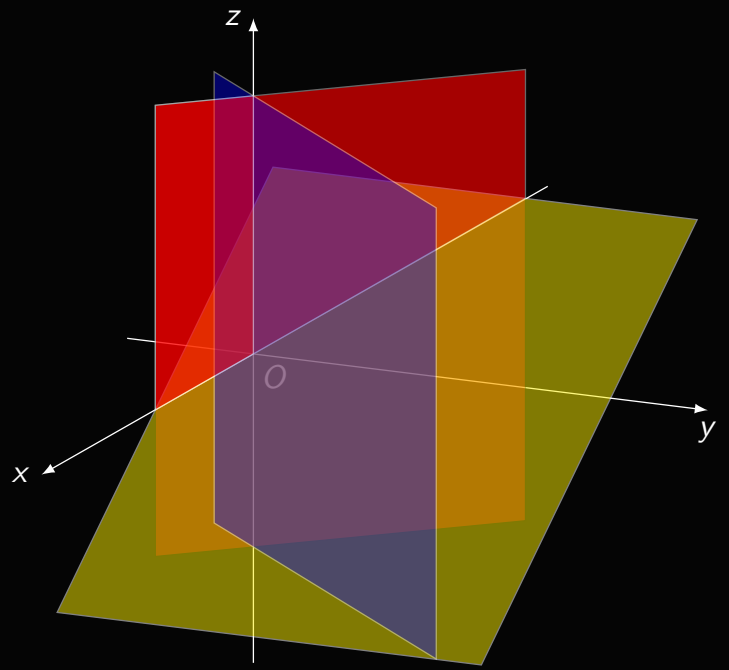


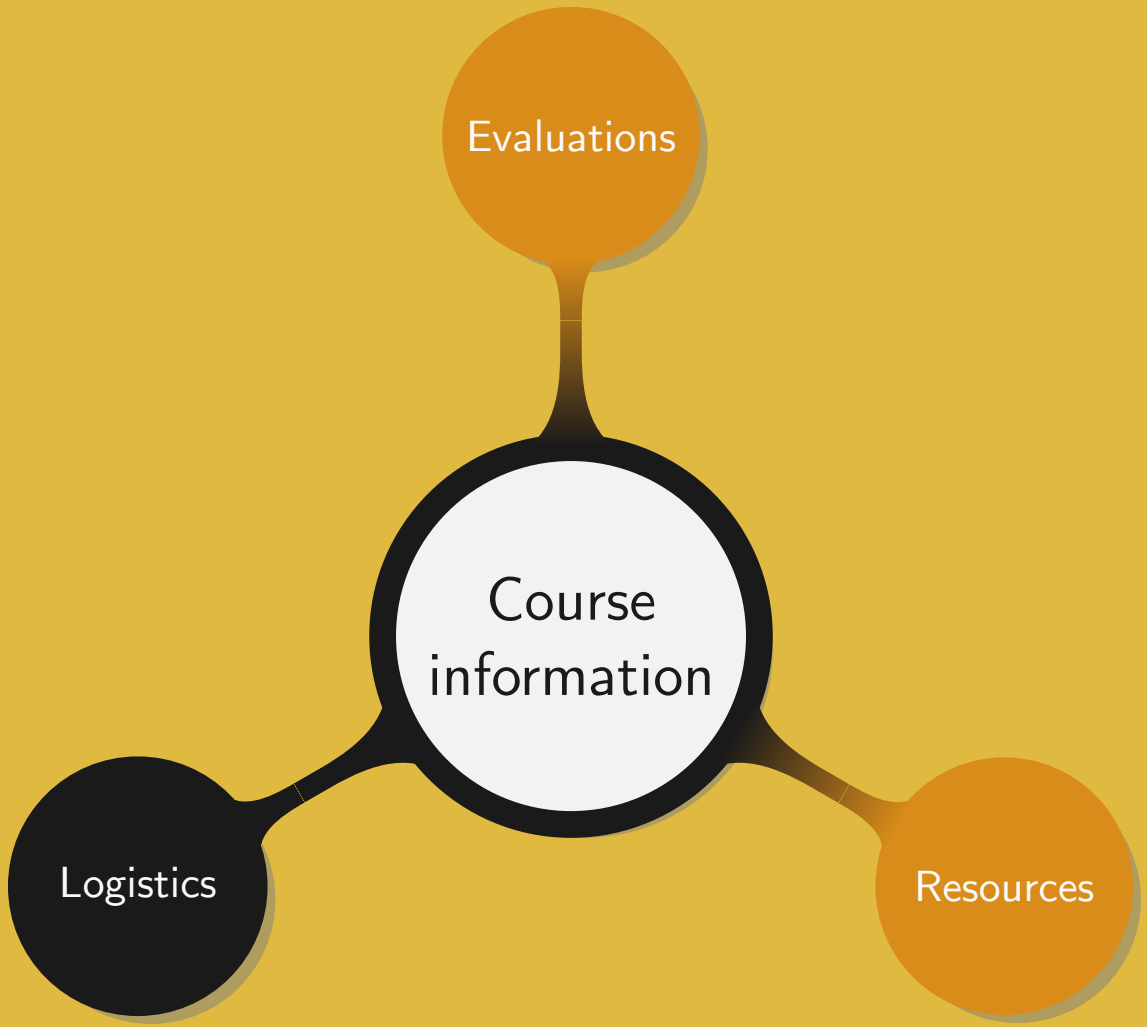
Linear algebra

Manuel – Spring 2023

Table of contents

0. Course information

Chapter organisation



Basic information

Teaching team:

- Instructor: Manuel (charlem@sjtu.edu.cn)
- Teaching assistant: Yuqi (aressegetes_stery@sjtu.edu.cn)

Important rules:

- When contacting a TA for an important matter, `CC` the instructor
- Prepend [MATH214] to the subject, e.g. `Subject: [MATH214] Grades`
- Use SJTU jBox service to share large files ($>$ 2 MB)

Never send large files by email

Course schedule

Course arrangements:

- Lectures:
  - Monday 8:00 – 9:40
  - Tuesday 8:00 – 9:40
  - Friday 8:00 – 9:40
- Manuel's office hours: Tuesday 10:00 – 11:20 (JI-439C)
- Yuqi's office hours: TBA

Course objectives

Primary goals:

- Become familiar with the concepts of vector space and matrix;
- Understand how to apply common proof methods;
- Be able to relate various topics to linear algebra;

Be ready for further studies in fields where linear algebra is needed

Course workflow

Learning strategy:

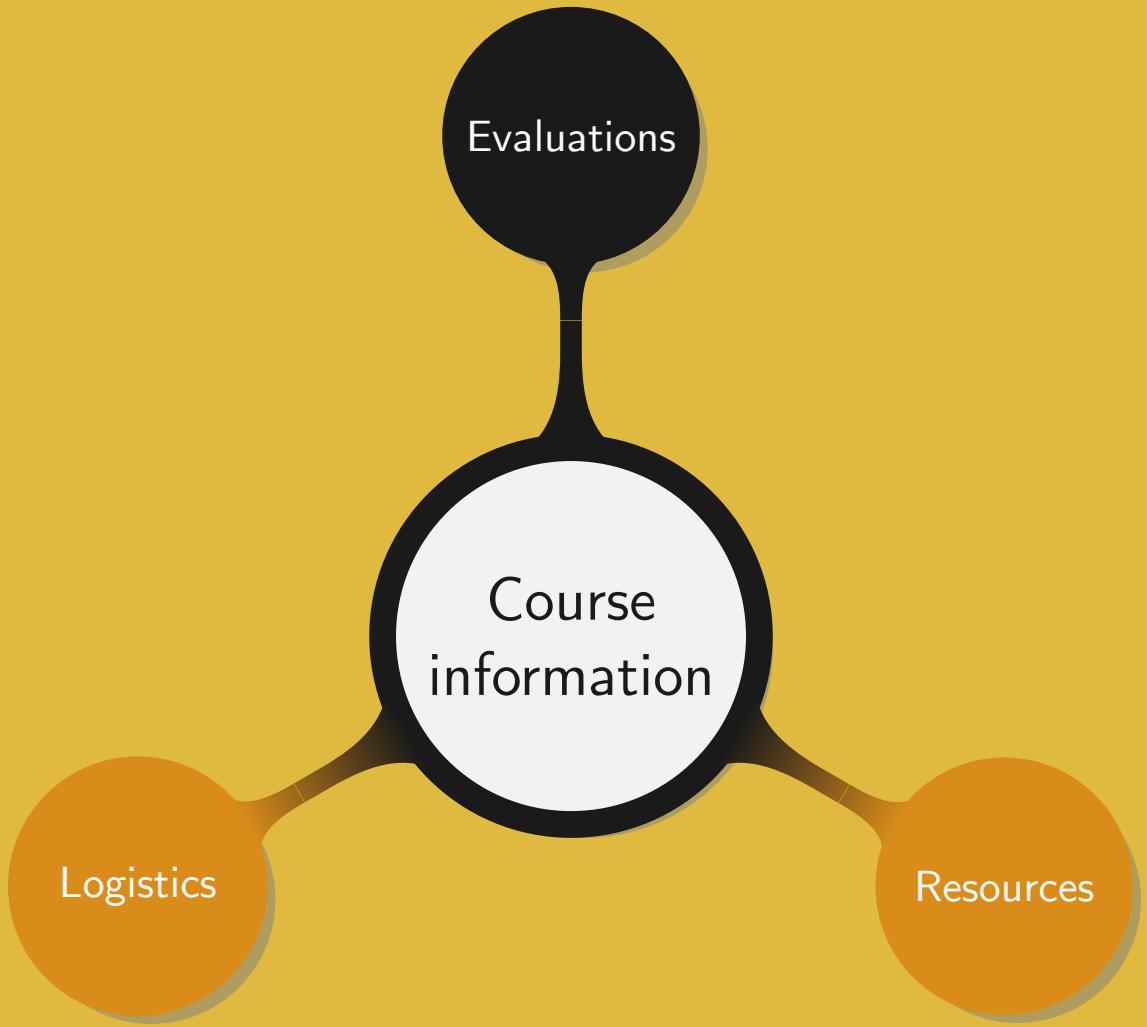
- Lectures and discussions:
  - Learn the basic concepts of linear algebra
  - Understand the relation between theory and application
  - Apply and derive common proof methods
- Personal side:
  - Work on a regular basis
  - Get some training at proving and calculating
  - Research how linear algebra is used in other fields

Course outcomes

Detailed goals:

- Be familiar with the concepts of vector, vector space, and linear map
- Know how to apply advanced matrix manipulations
- Understand the concepts of determinant and trace
- Be familiar with with endomorphism reduction
- Know how to work with isometries in the Euclidean plane
- Become familiar with common proof techniques and methods

Chapter organisation



Assignments

Homework:

- Total: 6 to 8
- Content: mathematical proofs and some calculations

Projects:

- Total: 1
- Content: select and study a topic related to linear algebra

Grading policy

Grade weighting:

- Homework: 15%
- Discussions: 15%
- Project: 20%
- Midterm: 25%
- Final: 25%

Assignment submissions:

- Bonus: $+10\%$ for a work fully written in LaTeX, bounded to 100%
- Penalty: -10% for a work not written in a neat and legible fashion
- Late policy: -10% per day, not accepted after three days

Grades will be curved with the median in the range $[\![B, B+]\!]$

Honor Code

Not allowed:

- Reuse the code or work from others
- Share too many details on how to complete a task

Allowed:

- Reuse short parts of the course or textbooks, quoting the source
- Share ideas and understandings on the course
- Provide hints on where or how to find information

Group works:

- Every student in a group is responsible for his group's submission
- If a student breaks the Honor Code, the whole group is guilty

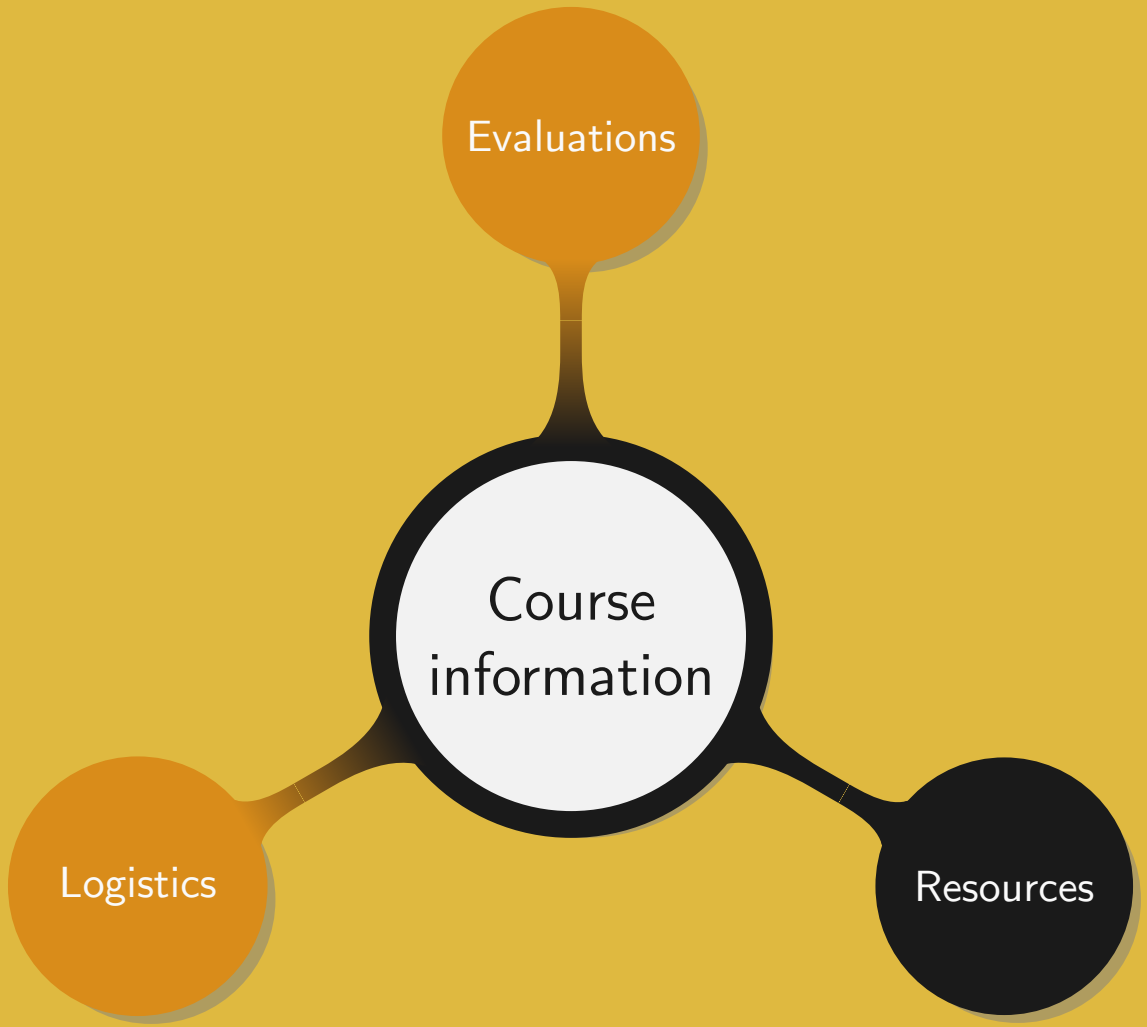
Special circumstances

Contact us as early as possible when:

- Facing special circumstances, e.g. full time work, illness, etc.
- Feeling late in the course
- Feeling to work hard without any result

Any late request will be rejected

Chapter organisation



Canvas

Information and documents available on the Canvas platform:

- Course materials:
  - Syllabus
  - Lecture slides
  - Homework
  - Projects
- Course information:
  - Announcements
  - Notifications
  - Grades
  - Surveys

References

Useful places where to find information:

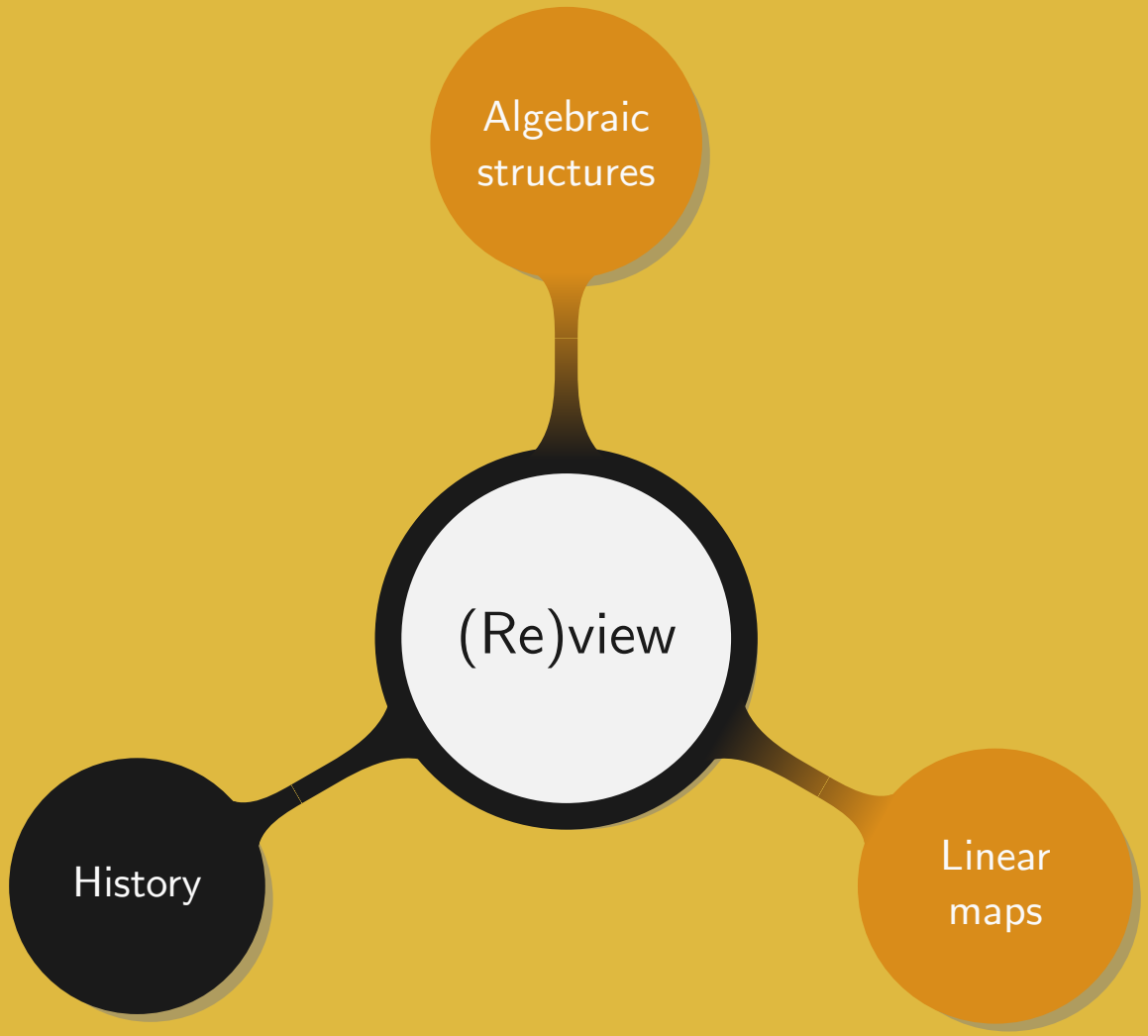
- *Introduction to Linear Algebra*, by Gilbert Strang
- *Linear Algebra and Geometry*, by Jean Dieudonne
- Search information online, i.e. $\{\textit{internet} \setminus \{\textit{non-English websites}\}\}$
- Piazza
- Mattermost

Key points – Key points

- Work regularly, do not wait the last minute
- Respect the Honor Code
- Go beyond what is taught
- Do not learn, understand
- Keep in touch with us
- Advice and suggestions are always much appreciated

1. (Re)view

Chapter organisation



Mathematics

Pure and applied mathematics:

- Pure: study concepts independently of their applications
- Applied: study concepts with the aim of solving real-world problems

The birth of pure mathematics:

- Plato distinguished:
  - Arithmetic: for philosophers
  - Logistic: for business men and men of war
- 19th century: progression of axiomatic systems and rigorous proofs
- 20th century: development of systematic abstraction, generalization, axiomatization, and demonstrations

The origins of algebra

Early mathematics was motivated by various needs:

- Mesopotamia: definition of numbers
- Egypt: empirical evaluation of fractions
- India: table of sinus, approximations

Equations or the origin or algebra:

Solutions to problems expressed using degree 2 equations

Diophantus formalizes the notion of equation

Al Khwarizmi writes the first treatise for systematically solving degree 2 equations

Al Khwarizmi's contribution

Al Khwarizmi's treaty had a lasting impact on mathematics:

- The word الجبر appearing in the title means "by reduction" and gave birth to the word *algebra*
- Al Khwarizmi became Algorithmi in Latin and later gave the word *algorithm*
- While not containing any new insight it gathered and synthesized the mathematical knowledge from Greece, India, and Persia
- Arabic numerals, derived from the Indian numerical system, were also introduced in Occident by Al Khwarizmi

Resolution of equations

Solving equations:

- Degree 2: Babylonians and Egyptians
- Degree 3: Cardano, 1545 (method lead to complex numbers)
- Degree 4: Ferrari and Bombelli, 1572 (reduction to degree 3)
- Degree 5: Abel-Ruffini, 1824 (no solution)
- Degree 5 or larger: Galois, 1831
  - Equation *solvable in radicals*: solution can be expressed using only the coefficients and basic operations
  - Provide criteria to decide whether an equation is solvable in radicals
  - Set the foundations for group and extension field theories

Resolution of systems of linear equations

Most important treatise in the history of Chinese mathematics:

- *Nine Chapters of the Mathematical Art*, Liu Hui, 263
- Compiles 246 problems with their solutions
- Describes a procedure for solving linear systems
- Refers to lost works of Chang Ts'ang, 3 or 4 century earlier

The western re-discovery:

- Found by Gauss and Jordan, and initially named *eliminatio vulgaris*
- Frobenius fully resolves the problem of linear equations with real and complex coefficients

Towards algebraic structures

Basic concept:

- Problems looking unrelated at first can share common properties
- The algebraic structure of the sets on which problems are defined have a bearing on the set of the solutions

The *algebraic structure* of a set essentially relies on the set itself and the operations it is endowed with. The various properties verified by these operations will define how "rich" this algebraic structure is.

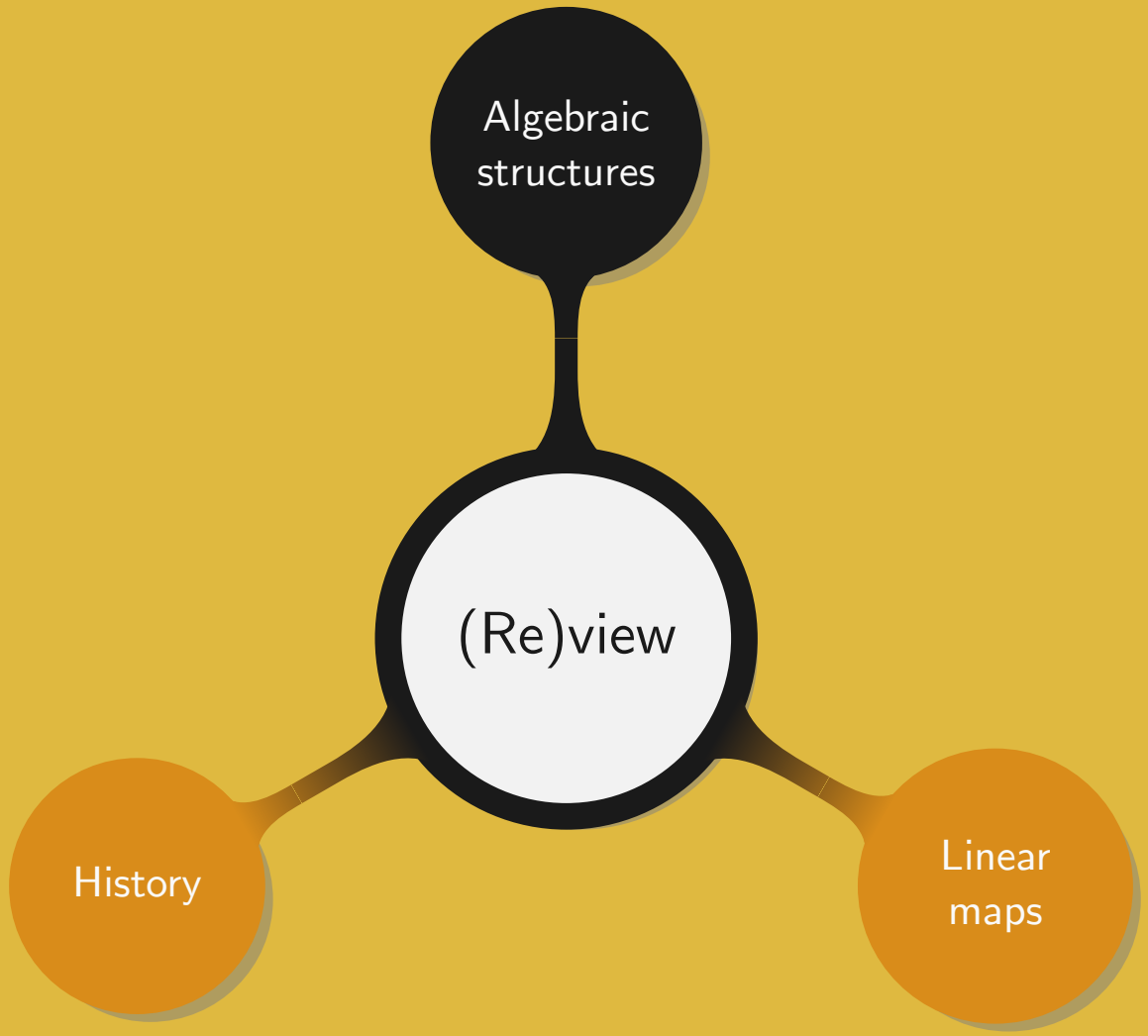
From an informal perspective, *algebra* started with the study of applied problems and turned into a more abstract discipline during the 19th century. It aim at classifying sets with respect to their behaviour under various types of operations.

Physics and linear algebra

Phenomenon that can be linearly characterized are common in physics:

- Many observable initially appeared as linear, e.g. gravitation
- *Linearization* is a common technique applied by physicists as preliminary approximation
- *Linear algebra* was initiated by Grassman in 1844 in his thesis
- Around 1860 mathematicians discovered Grassman's work and extended it
- In the 1920s formal definitions of linear spaces were proposed

Chapter organisation



Definitions

Let S and S' be two sets.

1. An *internal composition law* $(\circ)$ is an map from $S \times S$ to S such that
$$S \times S \longrightarrow S$$
$$(x, y) \longmapsto x \circ y.$$
2. An *external composition law* $(*)$ is an map from $S' \times S$ to S such that
$$S' \times S \longrightarrow S$$
$$(\alpha, x) \longmapsto \alpha * x.$$

Example. For a set S, the intersection $(\bigcap)$ and union $(\bigcup)$ define two internal composition laws for the class of subsets of S.

Definition (Group)

A *group* is a pair $(G, \circ)$ consisting of a set G and an internal composition law that verifies the following properties:

i. *Associativity:* $a \circ (b \circ c) = (a \circ b) \circ c$ for all $a, b, c \in G$

ii. *Existence of a unit element:* there exists an element $e \in G$ such that $a \circ e = e \circ a = a$ for all $a \in G$

iii. *Existence of inverse:* for every $a \in G$ there exists an element $a^{-1} \in G$ such that $a \circ a^{-1} = a^{-1} \circ a = e$

A group is called *abelian* if in addition to the above properties

iv. *Commutativity:* $a \circ b = b \circ a$ for all $a, b \in G$.

Definition (Ring)

A *ring* is a triple $(R, +, \cdot)$ consisting of a set R and two internal composition laws $(+)$ and $(\cdot)$, such that

i. $(R, +)$ is an abelian group

ii. *Multiplicative unit:* there exists an element $1 \in R$ such that

$$a \cdot 1 = 1 \cdot a = a \qquad \text{for all } a \in R$$

iii. *Associativity:* for any $a, b, c \in R$,

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c$$

iv. *Distributivity:* for any $a, b, c \in R$,

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c), \quad (b + c) \cdot a = (b \cdot a) + (c \cdot a)$$

A ring is called *commutative* if in addition to the above properties

v. *Commutativity:* $a \cdot b = b \cdot a$ for all $a, b \in R$

Definition (Field)

Let $(F, +, \cdot)$ be a commutative ring with unit element of addition 0 and unit element of multiplication 1. Then F is a *field* if

i. $0 \neq 1$

ii. For every $a \in F \setminus \{0\}$ there exists an element a^{-1} such that

$$a \cdot a^{-1} = 1.$$

Remark. Another way of writing this definition is to say that $(F, +, \cdot)$ is a field if $(F, +)$ and $(F \setminus \{0\}, \cdot)$ are abelian groups, $0 \neq 1$, and $\cdot$ distributes over $+$.

Definition (Vector space)

A *vector space* $(V, +, \cdot)$ over a field F is a set with an internal and external composition laws $(+)$ and $(\cdot)$, respectively. Moreover, for any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $\alpha, \beta \in F$ the following conditions hold:

i. $(V, +)$ is an abelian group;

ii. *Commutativity of scalar multiplication:* $\alpha \cdot (\beta \mathbf{v}) = (\alpha\beta) \cdot \mathbf{v}$;

iii. *Distributivity of scalar multiplication for vector addition:* $\alpha \cdot (\mathbf{u} + \mathbf{v}) = \alpha \cdot \mathbf{u} + \alpha \cdot \mathbf{v}$;

iv. *Distributivity of scalar multiplication for field addition:* $(\alpha+\beta)\cdot\mathbf{v} = \alpha \cdot \mathbf{v} + \beta \cdot \mathbf{v}$;

v. *Scalar multiplication identity:* for all $\mathbf{v} \in V, 1_F \cdot \mathbf{v} = \mathbf{v}$;

Elements of V are called *vectors* and the ones of F *scalar*.

Vector spaces

For the sake of simplicity, in the rest of the course, we will not denote vectors using bold fonts, for instance we now write v instead of $\mathbf{v}$. Besides, unless specified otherwise, the field $\mathbb{K}$ will refer to $\mathbb{R}$ or $\mathbb{C}$. Most results will however apply to any field. Unless specified otherwise, $\alpha, \beta, \lambda, \mu$ will refer to elements of the base field; V and W will be $\mathbb{K}$-vector spaces and their elements will most often be denoted u, v, w, x, y, or z, and n will denote a positive integer.

Remark. Vector spaces are viewed as "rigid" structures in the sense that they are not only closed under the *inner operation* but also for its *outer operation*. In fact as we will see it is even stronger, since a linear combination of elements remains in the vector space.

Vector spaces can be viewed as a simple generalisation of euclidean geometry. However they cannot take into account more complicated setups such as curved spaces. The study of those more advanced structures require the introduction of more complex objects called *varieties*.

Examples.

- $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{R}^2$ are $\mathbb{R}$-vector spaces.
- More generally $\mathbb{K}^n$ is a $\mathbb{K}$-vector space for the operations.

$$(x_0, x_1, \cdots, x_n) + (x_0', x_1' \cdots, x_n') = (x_0 + x_0', x_1 + x_1', \cdots, x_n + x_n')$$
$$\alpha \cdot (x_0, x_1, \cdots, x_n) = (\alpha \cdot x_0, \alpha \cdot x_1, \cdots, \alpha \cdot x_n).$$

- If V is a vector space over $\mathbb{K}$ and X is a non-empty set, then $\mathcal{F}(X, V)$ is a $\mathbb{K}$-vector space for the operations

$$f + g : X \longrightarrow V \qquad\qquad \alpha \cdot f : X \longrightarrow V$$
$$t \longmapsto f(t) + g(t) \qquad\qquad t \longmapsto \alpha \cdot f(t).$$

- The set of the polynomials over $\mathbb{K}$, denoted $\mathbb{K}[X]$ is a $\mathbb{K}$-vector space.

Proposition

Let V be a $\mathbb{K}$-vector space, then for any $\alpha \in \mathbb{K}$ and $x \in V$, $0_{\mathbb{K}} x = 0_V$, $\alpha \cdot 0_V = 0_V$, and $-(\alpha x) = \alpha(-x) = (-\alpha)x$.

Proof. For $x \in V$, we have $0_V + 0_{\mathbb{K}} x = 0_{\mathbb{K}} x = (0_{\mathbb{K}} + 0_{\mathbb{K}})x = 0_{\mathbb{K}} x + 0_{\mathbb{K}} x$. Hence $0_V = 0_{\mathbb{K}} x$. Similarly we obtain that $\alpha \cdot 0_V = 0_V$.

For the last equality observe that

$$\alpha x + (-\alpha)x = (\alpha - \alpha)x = 0_{\mathbb{K}} x = 0_V.$$

It means that $(-\alpha)x$ is the inverse of αx, i.e. $(-\alpha)x = -(\alpha x)$. Similarly $\alpha(-x) = -(\alpha x)$. □

Proposition

Let V be a $\mathbb{K}$-vector space, $\alpha \in \mathbb{K}$ and $x \in V$. If $\alpha \cdot x = 0$, then either $\alpha = 0_{\mathbb{K}}$ or $x = 0_V$.

Proof. Assume $\alpha \cdot x = 0$ and α is not $0_{\mathbb{K}}$. Since $\mathbb{K}$ is a field, α is invertible and $x = 1_{\mathbb{K}} \cdot x = \alpha^{-1}(\alpha \cdot x)$, and we have $x = \alpha^{-1} \cdot 0_V$, which by proposition 1.36 is 0_V. □

Definition (Linear combination)

Let V be a $\mathbb{K}$-vector space and $x_0, x_1, \cdots, x_n$ be vectors of V. Any vector in the form $\sum_{i \in \{0,1,\cdots,n\}} \lambda_i x_i$ is called a *linear combination* of vectors of V.

Cartesian product of vector spaces

Example. In the $\mathbb{R}$-vector space $\mathbb{C}$, any complex number is a linear combination of 1 and i.

Proposition

Given two $\mathbb{K}$-vector spaces V and W, $(V \times W, +, \cdot)$ where

- $(+)$ is defined by $(v, w) + (v', w') = (v + v', w + w')$, and
- $(\cdot)$ is defined by $\alpha \cdot (v, w) = (\alpha \cdot v, \alpha \cdot w)$, for $\alpha \in \mathbb{K}$,

is a vector space over $\mathbb{K}$.

Proof. According to definition 1.33 we need to prove that $(V \times W, +)$ is an abelian group and then verify the four last properties.

Cartesian product of vector spaces

It is easy to see that $(0_V, 0_W)$ is a unit element for $(+)$, and that any (v, w) has an inverse, namely $(-v, -w)$. We now prove the associativity using the associativity on both V and W. For three elements (v, w), (v', w'), and $(v'', w'') \in V \times W$,

$$\begin{aligned}((v, w) + (v', w')) + (v'', w'') &= (v + v', w + w') + (v'', w'') \\ &= ((v + v') + v'', (w + w') + w'') \\ &= (v + (v' + v''), w + (w' + w'')) \\ &= (v, w) + ((v', w') + (v'', w'')).\end{aligned}$$

Commutativity is proven following a similar pattern.

All four last properties can be proven using a same strategy, so we will only address one.

Let $(v, w) \in V \times W$ and $(\alpha_1, \alpha_2) \in \mathbb{K}^2$. Then

$$\begin{aligned}
\alpha_1 \cdot (\alpha_2 \cdot (v, w)) &= \alpha_1 \cdot (\alpha_2 \cdot v, \alpha_2 \cdot w) \\
&= (\alpha_1 \cdot (\alpha_2 \cdot v), \alpha_1 \cdot (\alpha_2 \cdot w)) \\
&= ((\alpha_1 \alpha_2) \cdot v, (\alpha_1 \alpha_2) \cdot w) \\
&= (\alpha_1 \alpha_2) \cdot (v, w).
\end{aligned}$$

□

Definition (Vector subspace)

A *vector subspace* of a vector space V is a subgroup of $(V, +)$ which is closed under scalar multiplication. Sometimes the short version *subspace* is used instead of vector subspace.

Examples.

- $\mathbb{R}$ and $i\mathbb{R}$ are vector subspaces of the $\mathbb{R}$-vector space $\mathbb{C}$.
- V and $\{0\}$ are called *trivial vector subspaces* of V.
- For a set X, if W is a vector subspace of V, then $\mathcal{F}(X, W)$ is a vector subspace of $\mathcal{F}(X, V)$.

Theorem

Let W be a non-empty subset of V. The following properties are equivalent.

i. W is a vector subspace of V;

ii. For any $\alpha, \beta \in \mathbb{K}$ and $x, y \in W$ we have $\alpha x + \beta y \in W$;

Proof. (i) $\Rightarrow$ (ii) First note that $0 \in W$. Indeed if W is a vector subspace of V, then it is a subgroup of $(V, +)$, and as such contains 0.
By definition of a vector subspace, and using the notations of the theorem, W is closed under scalar multiplication (definition 1.40), meaning that $\alpha \cdot x$ and $\beta \cdot y$ are both in W. It follows from the group structure that $\alpha x + \beta y \in W$.

(ii) $\Rightarrow$ (i) Since W is not empty then we can take $x \in W$, and by proposition 1.36, $0 \cdot x + 0 \cdot x = 0 \in W$.
We also note that W is closed under

- $(+)$: $\forall x, y \in W,\ x + y = 1 \cdot x + 1 \cdot y$;
- $(\cdot)$: $\forall x \in W,\ \forall \alpha \in \mathbb{K},\ \alpha \cdot x = \alpha \cdot x + 0 \cdot x$;

In particular the stability of $(\cdot)$ yields $-x = -1 \cdot x \in W$, i.e. if $x \in W$, then so is its inverse.

Hence W is a subgroup of V and closed under scalar multiplication. $\square$

Remark. Theorem 1.41 is a very important result as it is the most common way to prove that a set is endowed with a vector space structure. Showing that it is vector subspace of a known vector space is much faster and easier than proving it from the definition of a vector space.

Exercise. Assuming the "common operations" on the following sets, which are vector spaces?

- $C([0,1],\mathbb{R})$;
- $\left\{x \in \mathbb{R}^n, \sum_{i=1}^n x_i = 0\right\}$;
- $\left\{x \in \mathbb{R}^n, \sum_{i=1}^n x_i = 1\right\}$;
- $\left\{P \in \mathbb{K}[X], \deg P \leq n \in \mathbb{N}\right\}$;
- The course picture;
- $\emptyset$;

Proposition

Let $(V_i)_{i\in I}$, $I \neq \emptyset$, be a family of vector subspaces of V. Then $\bigcap_{i\in I} V_i$ is a vector subspace of V.

Proof. All the vector subspaces contain 0, therefore, so does their intersection.
If x and y are in the intersection of the vector spaces, then $\alpha x + \beta y$, $\alpha, \beta \in \mathbb{K}$, will belong to each of them. In other words it is in their intersection. Hence by theorem 1.41 the intersection of vector spaces is a vector space. □

Remarks.

- Proposition 1.43 is valid whether the intersection is finite or not.
- A union of vector spaces is usually not a vector space.

Proposition

Let S be a subset of a vector space V. There exists a smallest vector subspace containing S.

Proof. Clearly there exist subspaces containing S, as for instance $S \subset V$. By proposition 1.43 we know that their intersection is also a vector subspace. This is the smallest one containing S in the sense of the inclusion. □

Definition (Span)

Let S be a subset of a vector space V. The smallest vector subspace containing S is called the vector subspace *spanned* by S and denoted $\text{span}(S)$.

Examples.

- If $S = \emptyset$ then $\text{span}(S) = \{0\}$;
- In the $\mathbb{R}$-vector space $\mathbb{C}$, the subspace spanned by $\{1, i\}$ is $\mathbb{C}$;

Proposition

Let $n \in \mathbb{N}^*$ and $S = \{s_1, s_2, \cdots, s_n\}$ a subset of a vector space V with n elements. The subspace spanned by S is the set of linear combinations of the s_i, $0 \leq i \leq n-1$.

Proof. Let $S' = \left\{\sum_{i=1}^{n} \lambda_i s_i, (\lambda_1, \lambda_2, \cdots, \lambda_n) \in \mathbb{K}^n\right\}$.

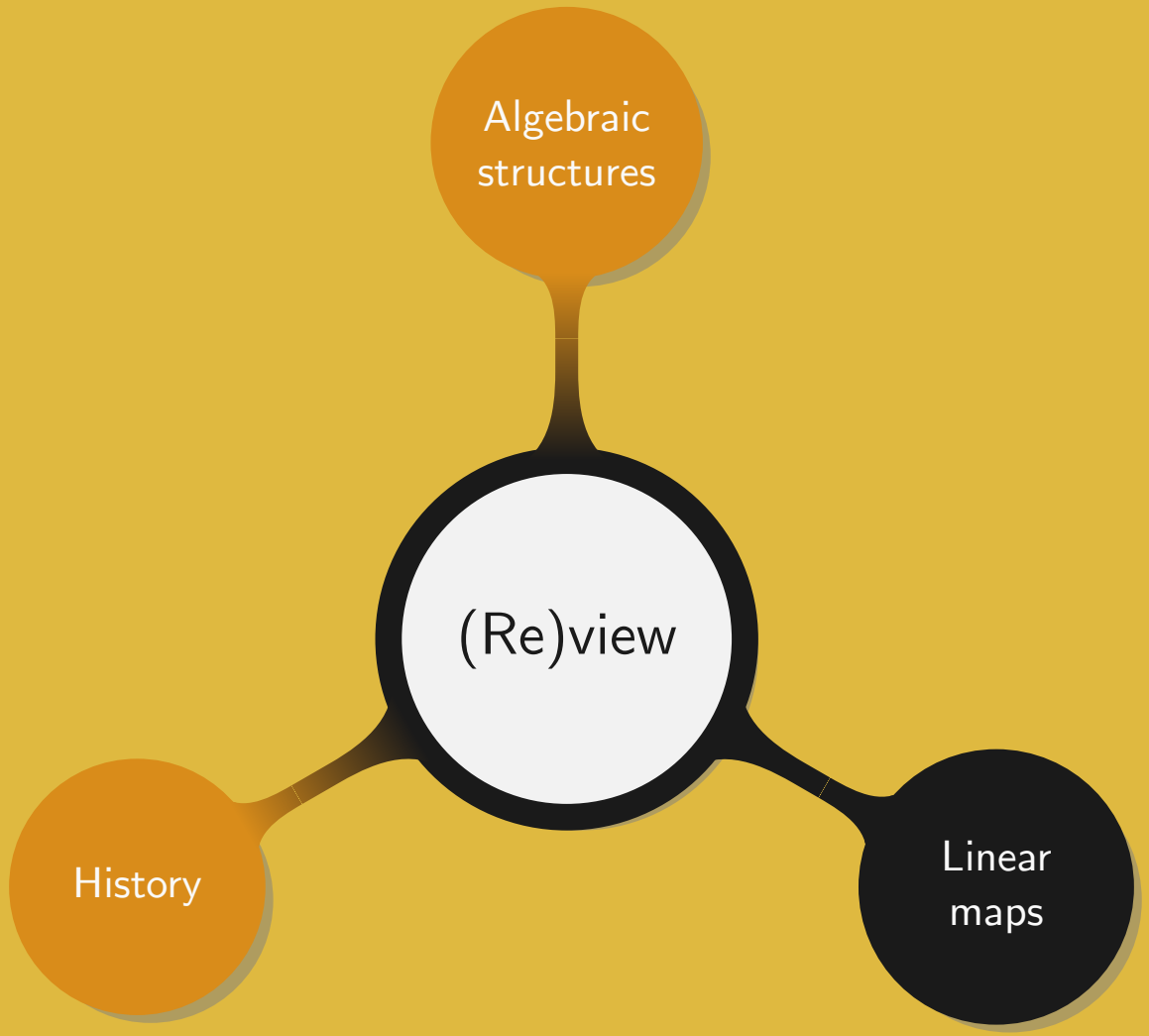
Observe that when taking all the λ_i to be 0, but $\lambda_k = 1$ we recover s_k. As this can be done for all $1 \leq k \leq n$, we have $S \subset S'$.

Now note that S' is a vector subspace of V. Indeed, it is not empty, as it contains $S \neq \emptyset$ and it is closed under linear combination,

$$\alpha \sum_{i=1}^{n} \lambda_i v_i + \beta \sum_{i=1}^{n} \mu_i v_i = \sum_{i=1}^{n} (\alpha\lambda_i + \beta\mu_i) v_i.$$

Let W be a subspace of V containing S. As it is closed under both $(+)$ and $(\cdot)$ and contains all the s_i, $S' \subset W$. Hence $S' = \text{span}(S)$. $\square$

Chapter organisation



Definitions

1. A *morphism* is a map preserving the mathematical structure between two sets;
2. Let $(S, *)$ and $(S', *')$ be two algebraic structures. A *homomorphism* is a function $f : S \to S'$ such that for any $x, y \in S$, $f(x * y) = f(x) *' f(y)$.
3. An *isomorphism* is a bijective homomorphism;
4. An *endomorphism* is a homomorphism from a set to itself;
5. An *automorphism* is a bijective endomorphism;

Example. For $x \in S$, $p, q \in \mathbb{N}$, saying that $x^p \cdot x^q = x^{p+q}$ is similar to saying that

$$\mathbb{N} \longrightarrow S$$
$$n \longmapsto x^n$$

is a homomorphism from $(\mathbb{N}, +)$ to $(S, \cdot)$.

Definitions

Let V and W be two $\mathbb{K}$-vector spaces.

1. A function f from V to W is called *linear map* if for any $u, v \in V$ and $\alpha, \beta \in \mathbb{K}$,
$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v).$$
2. A *linear form* or *linear functional* is a linear map from V to $\mathbb{K}$.

Exercise. Which of the following functions are linear maps?

- Differentiation
$$C^\infty(\mathbb{R}) \longrightarrow C^\infty(\mathbb{R})$$
$$f \longmapsto f';$$
- For $\alpha \in \mathbb{K}$ and $P \in \mathbb{K}[X]$, $P \mapsto P(\alpha)$;
- Integral on $[a, b] \subset \mathbb{R}$
$$C^\infty[a, b] \longrightarrow \mathbb{R}$$
$$f \longmapsto \int_a^b f(t)\,\mathrm{d}t;$$
- Translation of vector v;

Image of a subspace by a linear map

Remark. Linear maps are sometimes called vector space homomorphisms since they preserve the vector space structure. In the rest of the course we will abuse the notations and interchangeably use "homomorphism" and "linear map". In case of possible confusion we will specify the algebraic structure. Note that similarly an isomorphisms is a bijective linear map.

Proposition

Let V and W be vector spaces and f be a linear map from V to W.

- If V_0 is a subspace of V, then $f(V_0)$ is a subspace of W;
- If W_0 is a subspace of W, then $f^{-1}(W_0)$ is a subspace of V;

Proof. Let V_0 be a subspace of V, $W_0 = f(V_0) \subset W$, and $\alpha, \beta \in \mathbb{K}$.

Since $0 \in V_0$, $f(0) = 0 \in W_0$. Moreover for any $w, w' \in W_0$, there exist $v, v' \in V_0$ such that $f(v) = w$ and $f(v') = w'$. Thus we know that

$$\alpha w + \beta w' = \alpha f(v) + \beta f(v') = f(\alpha v + \beta v')$$

Image of a subspace by a linear map

belongs to W_0, since $\alpha v + \beta v'$ is in V_0. Hence W_0 is a vector subspace of W.
We now consider W_0, a subspace of W and show that $V_0 = f^{-1}(W_0)$ is a subspace of V.

Since $f(0) = 0 \in W_0$, we know that $0 \in V_0$. Moreover as for $u, v \in V_0$, $f(u), f(v) \in W_0$, meaning that $f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$ is also in W_0. This is true because W_0 is a subspace of W and as such $\alpha u + \beta v \in V_0$. Hence V_0 is a vector subspace of V. □

Definitions

Let f be a linear map between two vector spaces V and W.

1. The *kernel* of f is the subspace of V defined by
$$\ker f = f^{-1}(\{0\}) = \{x \in V, f(x) = 0\};$$
2. The *image* of f is the subspace of W defined by $\operatorname{im} f = f(V)$;

Theorem

Let f be a linear map from V to W. The following conditions are equivalent.

i. f is injective;
ii. $\ker f = \{0\}$;
iii. For any $x \in V$, $f(x) = 0$ implies $x = 0$;

Proof. (ii) $\Leftrightarrow$ (iii). Note that (iii) can be rewritten as $\ker f \subset \{0\}$. Recalling that by definition $\ker f$ is a subspace of V, $0 \in \ker f$. Thus (ii) and (iii) are equivalent.

(i) $\Rightarrow$ (ii). Assume f injective and take $x \in \ker f$. Then $f(x) = 0 = f(0)$, which means $x = 0$ by the injectivity of f.

(i) $\Leftarrow$ (ii). Assume $\ker f = \{0\}$ and take $x, y \in V$, such that $f(x) = f(y)$. Then by linearity we see that $0 = f(x) - f(y) = f(x-y)$, and $x - y \in \ker f$. This yields $x = y$ and thus f is injective. $\square$

Remark. To prove that a linear map is injective in suffices to use either characterization (ii) or (iii) from theorem 1.52

In the rest of the course we will denote the set of the linear maps from V to W, $\mathcal{L}(V, W)$, and write $\mathcal{L}(V)$ for the endomorphisms of V.

Proposition

Let f and g be in $\mathcal{L}(V, W)$, and λ and μ be two scalars. Then $\lambda f + \mu g$ is also in $\mathcal{L}(V, W)$.

Proof. Let $x, y \in V$ and $\alpha, \beta \in \mathbb{K}$.

$$\begin{aligned}(\lambda f + \mu g)(\alpha x + \beta y) &= \lambda f(\alpha x + \beta y) + \mu g(\alpha x + \beta y) \\ &= \lambda(\alpha f(x) + \beta f(y)) + \mu(\alpha g(x) + \beta g(y)) \\ &= \alpha\left((\lambda f + \mu g)(x)\right) + \beta\left((\lambda f + \mu g)(y)\right)\end{aligned}$$

This completes the proof. □

Proposition

1. If $f \in \mathcal{L}(U, V)$ and $g \in \mathcal{L}(V, W)$, then $g \circ f \in \mathcal{L}(U, W)$.
2. If f is an isomorphism from V to W, then f^{-1} is linear.

Proof. (1) For $x, y \in V$ and $\alpha, \beta \in \mathbb{K}$ we see that $g \circ f$ is linear since

$$\begin{aligned}(g \circ f)(\alpha x + \beta y) &= g\left(\alpha f(x) + \beta f(y)\right) \\ &= \alpha(g \circ f)(x) + \beta(g \circ f)(y).\end{aligned}$$

(2) Let $x, y \in W$, and $\alpha, \beta \in \mathbb{K}$. Since f is an isomorphism it is injective, so to prove that $f^{-1}(\alpha x + \beta y)$ and $\alpha f^{-1}(x) + \beta f^{-1}(y)$ are equal, it suffices to show that f maps them to a same element.

$$\begin{aligned}f\left(\alpha f^{-1}(x) + \beta f^{-1}(y)\right) &= \alpha f\left(f^{-1}(x)\right) + \beta f\left(f^{-1}(y)\right) \\ &= \alpha x + \beta y \\ &= f\left(f^{-1}(\alpha x + \beta y)\right).\end{aligned}$$

□

Proposition

$\mathcal{L}(V, W)$ is a $\mathbb{K}$-vector space.

Proof. Clearly the zero function is linear and by proposition 1.53 $\mathcal{L}(V, W)$ is closed under linear combination. Hence it is a subspace of $\mathcal{F}(V, W)$. □

Proposition

Let $f \in \mathcal{L}(U, V)$, $g \in \mathcal{L}(V, W)$, and $\alpha, \beta \in \mathbb{K}$.

1. For $u, v \in \mathcal{L}(V, W)$
$$(\alpha u + \beta v) \circ f = \alpha(u \circ f) + \beta(v \circ f).$$
2. For $u, v \in \mathcal{L}(U, V)$
$$g \circ (\alpha u + \beta v) = \alpha(g \circ u) + \beta(g \circ v).$$

Proof. (1) The result is straightforward when considering the definition of $\alpha u + \beta v$.

(2) This is a simple consequence of the linearity of g. □

Remark. Proposition 1.55 can be rephrased into

$$\begin{array}{ccc} \mathcal{L}(V, W) & \longrightarrow & \mathcal{L}(U, W) \\ u & \longmapsto & u \circ f \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathcal{L}(U, V) & \longrightarrow & \mathcal{L}(U, W) \\ u & \longmapsto & g \circ u \end{array}$$

are linear.

Proposition

$(\mathcal{L}(V), +, \circ)$ is a ring.

Proof. From the first proposition 1.55, $\mathcal{L}(V)$ is a vector space, implying that $(\mathcal{L}(V), +)$ is an abelian group, and by proposition 1.54 we also know that $\mathcal{L}(V)$ is closed under $(\circ)$.

$\mathcal{L}(V)$ also contains the identity function which is the unit for $(\circ)$, and $(\circ)$ is associative. Finally the distributivity of $(\circ)$ with respect to $(+)$ is a consequence of the second proposition 1.55. □

We denote by $\mathrm{GL}(V)$ the set of all the automorphism of V.

Proposition

$(\mathrm{GL}(V), \circ)$ is a group.

Proof. From proposition 1.54 we directly obtain that

- for any $f \in \mathrm{GL}(V)$, $f^{-1} \in \mathrm{GL}(V)$, i.e. symmetry is verified;
- the composition of two automorphisms is linear. As it is a bijection it is also in $\mathrm{GL}(V)$, meaning that it is closed under $(\circ)$.

Similarly to what happens for $\mathcal{L}(V)$, the identity function is in $\mathrm{GL}(V)$ and composition is associative. □

Definitions

Let V be a $\mathbb{K}$-vector space.

1. $\mathrm{GL}(V)$ is called the *general linear group* of V.
2. A subgroup of $\mathrm{GL}(V)$ is called *linear group*.

Remark. $\mathrm{GL}(V)$ is the group of the units of the ring $\mathcal{L}(V)$.

Definitions

Let U, V, W be three $\mathbb{K}$-vector spaces.

1. A *bilinear map* f, from $U \times V$ to W, is a map such that for any $(x, y) \in U \times V$, the following two maps are linear,

$$\begin{array}{ccc} V \longrightarrow W & & U \longrightarrow W \\ v \longmapsto f(x, v) & \text{and} & u \longmapsto f(u, y). \end{array}$$

2. A *bilinear form* on V is a bilinear map defined from $V \times V$ to $\mathbb{K}$.

Examples.

- The map

$$\begin{aligned} \mathbb{C} \times \mathbb{C} &\longrightarrow \mathbb{C} \\ (z, z') &\longmapsto zz' \end{aligned}$$

  is a bilinear map over the $\mathbb{C}$-vector space $\mathbb{C}^2$;

- The second proposition from 1.55 can be rephrased by saying that

$$\begin{aligned} \mathcal{L}(U, V) \times \mathcal{L}(V, W) &\longrightarrow \mathcal{L}(U, W) \\ (u, v) &\longmapsto v \circ u \end{aligned}$$

  is a bilinear map.

- The map

$$\begin{aligned} \mathcal{L}(V, W) \times V &\longrightarrow W \\ (v, x) &\longmapsto v(x) \end{aligned}$$

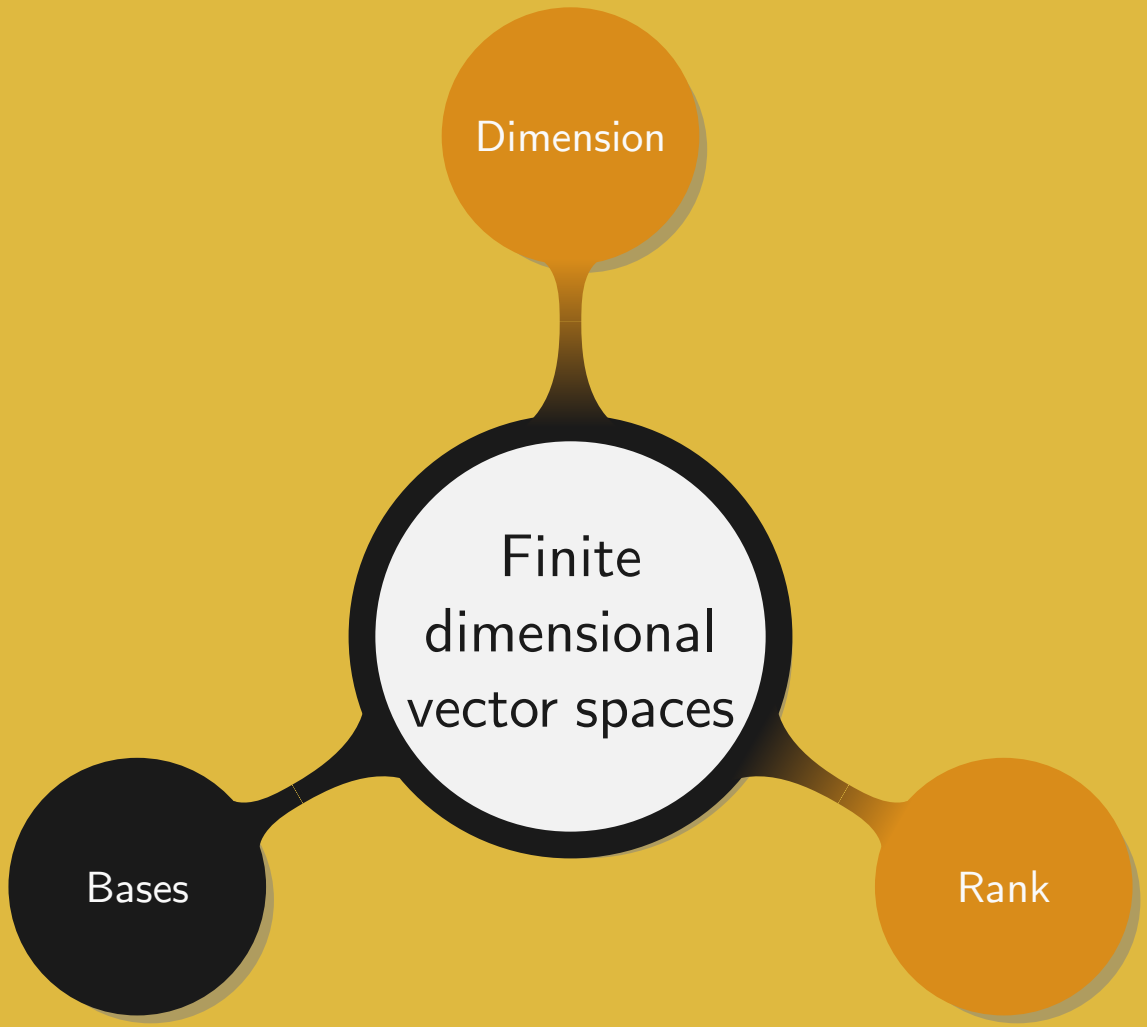
  is a bilinear map.

Key points

- What is a vector space?
- What is an endomorphism?
- What is the algebraic structure for the space of linear functions?
- What the kernel of a linear map?
- How rich is the vector space algebraic structure?

2. Finite dimensional vector spaces

Chapter organisation



Sum of subspaces

As noted in remark 1.44 the union of two vector spaces V and W is not a vector space. However we know that the smallest subspace containing both V and W is spanned by $V \cup W$ (definition 1.45).

Definition (Sum of subspaces)

Let V_1 and V_2 be two subspaces of a vector space V. The *sum* of V_1 and V_2, denoted $V_1 + V_2$ is defined by $\{v_1 + v_2, (v_1, v_2) \in V_1 \times V_2\}$.

Proposition

If V_1 and V_2 are two subspaces of a vector space V, then the sum $W = V_1 + V_2$ is the smallest subspace containing both V_1 and V_2.

Proof. Since for any $x \in V_1$, $x = x+0$, with $0 \in V_2$, we see that $V_1 \subset W$. Similarly $V_2 \subset W$. Besides this also shows that W is not empty.

By definition, for any $u_1, u_2 \in W$, there exist $x_1, x_2 \in V_1$ and $y_1, y_2 \in V_2$, such that

$$u_1 = x_1 + y_1 \text{ and } u_2 = x_2 + y_2.$$

Then for any $\alpha_1, \alpha_2 \in \mathbb{K}$ we have

$$\alpha_1 u_1 + \alpha_2 u_2 = (\alpha_1 x_1 + \alpha_2 x_2) + (\alpha_1 y_1 + \alpha_2 y_2),$$

with $\alpha_1 x_1 + \alpha_2 x_2 \in V_1$ and $\alpha_1 y_1 + \alpha_2 y_2 \in V_2$. Therefore $\alpha_1 u_1 + \alpha_2 u_2 \in W$, and W is a vector subspace of V.

We now show that W is the smallest subspace containing both V_1 and V_2. Let W' be a subspace containing V_1 and V_2. We know that for $w \in W$, there exist $x \in V_1$ and $y \in V_2$ such that $w = x + y$. Clearly x and y are also in W', such that $w \in W'$. This shows that $W \subseteq W'$. $\square$

Sum of subspaces

Examples.

- Any element of $\mathbb{K}^2$ can be written as $(x-y)(1,0)+y(1,1)$. So $\mathbb{K}^2 = \text{span}((1,0)) + \text{span}((1,1))$.
- As an obvious generalization of proposition 2.63, if V_i, $1 \leq i \leq n$, are subspaces of a vectors space V, then

$$\sum_{i=1}^{n} V_i = \left\{ \sum_{i=1}^{n} x_i, \forall i \in [\![1,n]\!], x_i \in V_i \right\}$$

  is the smallest subspace containing all the V_i.
- Let u be a linear map from V to W, and V_i, $1 \leq i \leq n$, are subspaces of V, then

$$u\left(\sum_{i=1}^{n} V_i\right) = \sum_{i=1}^{n} u(V_i).$$

Definition (Direct sum of subspaces)

Let V_1 and V_2 be two subspaces of a vector space V. Then V is called the *direct sum* of V_1 and V_2, denoted $V_1 \oplus V_2$ if $V = V_1 + V_2$ and $V_1 \cap V_2 = \{0\}$.

Proposition

A vector space V is the direct sum of two subspaces V_1 and V_2 if and only if for any $v \in V$, there exists a unique pair $(v_1, v_2) \in V_1 \times V_2$ such that $v = v_1 + v_2$.

Proof. ($\Rightarrow$) We simply need to show the unicity of the decomposition of any element of V into the sum of an element of V_1 and an element of V_2.

Assume this decomposition is not unique, i.e. there exists an element $v \in V$ for which there exist $v_1, v_1' \in V_1$ and $v_2, v_2' \in V_2$, such that $v = v_1 + v_2 = v_1' + v_2'$. Then we can write $v_1 - v_1' = v_2' - v_2$. Thus those two elements are in $V_1 \cap V_2$, which by definition is $\{0\}$.↯

($\Leftarrow$) Assume all elements of V can be written in a unique way as a sum of an element of V_1 and an element of V_2. Therefore we have $V = V_1 + V_2$ and we only need to prove that $V_1 \cap V_2 = \{0\}$. Observe that any $x \in V_1 \cap V_2$ can be written as $x = x + 0 = 0 + x$. However the unicity of the decomposition implies $x = 0$. Hence $V_1 \cap V_2 = \{0\}$. □

Remark. In general to prove that $V_1 \cap V_2 = \{0\}$ it suffices to show that taking $x \in V_1 \cap V_2$ implies $x = 0$, i.e. $V_1 \cap V_2 \subset \{0\}$. The other inclusion is trivial since $V_1 \cap V_2$ is a subspace of V, and as such always contains $\{0\}$.

Direct sum of subspaces

Examples.

- As an obvious generalization of proposition 2.66, if V is the direct sum of subspaces V_i, $1 \leq i \leq n$, then for any $x \in V$, there is a unique $(x_1, x_2, \cdots, x_n) \in \prod_{i=1}^{n} V_i$, such that $x = \sum_{i=1}^{n} x_i$, with $x_i \in V_i$.
- $\mathbb{K}^3$ can be written as a direct sum of $\mathbb{K}^2 \times \{0\}$ and $\text{span}((0,0,1))$.
- Let $V = \mathbb{K}[X]$ and $Q \in V \backslash \{0\}$. Then $V = \{R \in V, \deg R < \deg Q\} \oplus \{PQ, P \in V\}$.
- $\mathbb{K}^2 = \text{span}((1,0)) \oplus \text{span}((0,1)) = \text{span}((1,0)) \oplus \text{span}((1,1))$. In particular this highlights that the decomposition of a vector space into a direct sum of subspaces is not unique.

Example. Let $V = \mathcal{F}(\mathbb{R}, \mathbb{R})$ and $\mathcal{E}$ and $\mathcal{O}$ denote the sets of even and odd functions, respectively. Both $\mathcal{E}$ and $\mathcal{O}$ contain the zero function and are closed under linear combinations. We will show that any function f from V can be written in a unique way as the sum of $g \in \mathcal{E}$ and $h \in \mathcal{O}$.

If $f = g + h$, then for any $x \in \mathbb{R}$

$$\begin{cases} f(x) = g(x) + h(x) \\ f(-x) = g(-x) + h(-x) = g(x) - h(x). \end{cases}$$

So for any $x \in \mathbb{R}$,

$$g(x) = \frac{f(x) + f(-x)}{2} \quad \text{and} \quad h(x) = \frac{f(x) - f(-x)}{2}. \tag{2.1}$$

Therefore, if g and h exist, then their are uniquely determined by f. We now address the existence of those two functions.

Since we have already expressed g and h in term of f, (2.1), we only need to verify that $g \in \mathcal{E}$ and $h \in \mathcal{O}$. For any $x \in \mathbb{R}$

$$g(-x) = \frac{f(-x) + f(x)}{2} = g(x) \quad \text{and} \quad h(-x) = \frac{f(-x) - f(x)}{2} = -h(x).$$

This shows that $\mathcal{F}(\mathbb{R}, \mathbb{R}) = \mathcal{E} \oplus \mathcal{O}$.

A remarkable example is the exponential function which can be written as the sum of the hyperbolic cosine and hyperbolic sine functions, which are even and odd, respectively.

Definition

Let V be a vector space and $S \subset V$. If any element in V can be written as a linear combination of vectors in S, then S is called a *spanning set* for V.

Spanning set

Remark. If $\mathcal{S}$ is a spanning set for V, then any superset $\mathcal{S}' \supset \mathcal{S}$ is also a spanning set for V.

Examples.

- Considering $\mathbb{C}$ as an $\mathbb{R}$-vector space, $\{1, i\}$ is a spanning set for $\mathbb{C}$.
- Two non-parallel vectors from the plane span it.
- Any polynomial of $\mathbb{K}_n[X]$ can be written as $\sum_{i=0}^{n} a_i X^i$, with $a_i \in \mathbb{K}, 0 \leq i \leq n$. So $\{1, X, \cdots, x^n\}$ is a spanning set for $\mathbb{K}_n[X]$.

Proposition

Let $\mathcal{S}$ be a spanning set for V. A subset $\mathcal{S}'$ of V is a spanning set for V if and only if any element of $\mathcal{S}$ can be written as a linear combination of elements of $\mathcal{S}'$.

Proof. ($\Rightarrow$) If $\mathcal{S}'$ is a spanning set for V, then any of its elements can be written as a linear combination of vectors of $\mathcal{S}'$. In particular this is true for any element of $\mathcal{S} \subset V$.

($\Leftarrow$) If any element of $\mathcal{S}$ can be written as linear combinations of elements of $\mathcal{S}'$, this means that $S \subset \operatorname{span} \mathcal{S}'$. Moreover as we know that $V = \operatorname{span} \mathcal{S}$ is the smallest subspace containing $\mathcal{S}$, we conclude that $V \subset \operatorname{span} \mathcal{S}'$. Hence $\mathcal{S}'$ is a spanning set for V. □

Example. Let $j = e^{2i\pi/3}$. Then $\{1, j\}$ is a spanning set for $\mathbb{C}$ considered as an $\mathbb{R}$-vector space. To see it, it suffices to refer to the known spanning set $\{1, i\}$ and then write $1 = 1 \cdot 1 + 0 \cdot j$, and $i = \frac{1}{\sqrt{3}} \cdot 1 + \frac{2}{\sqrt{3}} j$.

Definition

A subset $\mathcal{I}$ of a vector space V is said to be *linearly independent* if no vector in $\mathcal{I}$ can be expressed as a linear combinations of others.

Proposition

Let V be a vector space and $\mathcal{I} = \{v_1, v_2, \cdots, v_n\} \subset V$. Then $\mathcal{I}$ is linearly independent if and only if the unique solution to

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$$

is $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.

Proof. ($\Rightarrow$) If $\mathcal{I}$ is linearly independent and $\sum_{i=1}^{n} \alpha_i v_i = 0$, with not all $\alpha_i = 0$, then we can express any of the $v_k = -\frac{1}{\alpha_k} \sum_{i=1, i\neq k}^{n} \alpha_i v_i$ as a linear combination of the other v_i.

($\Leftarrow$) Assume the unique solution is given by $\alpha_i = 0$, $1 \leq i \leq n$, and $\mathcal{I}$ is linearly dependent. From the linear dependency we know that there exist some β_i, $1 \leq i \leq n$, such that we can write for instance $v_k = \sum_{i=1, i\neq k}^{n} \beta_i v_i$, which can be transformed into $\sum_{i=1, i\neq k}^{n} \beta_i v_i - v_k = 0$.

Hence, $\mathcal{I}$ is linearly independent. □

Examples.

- A set with one element x is linearly independent if and only if $x \neq 0$. This is clear since
  - if $x = 0$, then $1 \cdot x = 0$, meaning that $\mathcal{I}$ is linearly dependent;
  - if $x \neq 0$, then for any λ, $\lambda x = 0$ implies $\lambda = 0$;
- In $\mathbb{C}$, as soon as the imaginary part of ω is not 0, then $\{1, \omega\}$ is linearly independent. Indeed for $a, b \in \mathbb{R}$ such that $a + b\omega = 0$ we observe that $b\omega = 0$, which implies $b = 0$. Subsequently we obtain $a = 0$
- The set $\{1, X, \cdots, X^n\}$ is linearly independent in $\mathbb{K}[X]$ since the polynomial $\sum_{i=0}^{n} \lambda_i X^i$ is equal to zero if and only if all the $\lambda_i = 0$.

Proposition

Any subset of a linearly independent set is linearly independent.

Proof. Let $\mathcal{I} = \{v_i, 1 \leq i \leq n\}$ be a linearly independent subset of a vector space V. We select p vectors among the n elements and use a permutation on v_i to reorder them into the subset $\mathcal{I}' = \{v_1, v_2, \cdots, v_p\}$. Let $\lambda_i \in \mathbb{K}$, $1 \leq i \leq p$, such that $\sum_{i=1}^{p} \lambda_i v_i = 0$. By defining $\lambda_i = 0$ for $p+1 \leq i \leq n$, we obtain $\sum_{i=0}^{n} \lambda_i v_i = 0$. However as $\mathcal{I}$ is linearly independent, it means that all the $\lambda_i = 0$. In particular $\lambda_1, \lambda_2, \cdots, \lambda_p$ are all zero, which shows the linear independence of $\mathcal{I}'$. □

Remarks.

- The contrapositive of proposition 2.74 states that any superset of a linearly dependent set is a linearly dependent set.
- A linearly independent set cannot contain the vector 0.
- A set with two parallel vectors is linearly dependent, however the converse if false. For instance in $\mathbb{C}$, $\{1, i, 1+i\}$ is not linearly independent, but no vector is parallel with any other.

Proposition

Let $\mathcal{I}$ be a linearly independent subset of a vector space V, and $v \in V$. Then $\mathcal{I} \cup \{v\}$ is a linearly dependent set if and only if v can be written as a linear combination of elements of $\mathcal{I}$.

Proof. ($\Leftarrow$) Let $\mathcal{I} = \{v_i, 1 \leq i \leq n\}$ and $\lambda_i, 1 \leq i \leq n$, be such that $v = \sum_{i=1}^{n} \lambda_i v_i$. Then $1 \cdot v + \sum_{i=1}^{n} (-\lambda_i) v_i = 0$, meaning that $\mathcal{I} \cup \{v\}$ is a linearly dependent set.

($\Rightarrow$) If $\mathcal{I} \cup \{v\}$ is linearly dependent, then there exist $\alpha, \lambda_i,\ 1 \leq i \leq n$, not all zeros, such that $\alpha v + \sum_{i=1}^{n} \lambda_i v_i = 0$. Therefore if $\alpha = 0$ the sum must be 0, with the λ_i not all zeros.
This cannot be the case since $\mathcal{I}$ is linearly independent. □

Theorem

Let $\mathcal{I}$ be a linearly independent set, and λ_i and μ_i, $1 \leq i \leq n$, be scalars. If

$$\sum_{i=1}^{n} \lambda_i v_i = \sum_{i=1}^{n} \mu_i v_i,$$

then for all $1 \leq i \leq n$, $\lambda_i = \mu_i$.

Proof. This is straightforward after reorganising the terms of the sums into

$$\sum_{i=1}^{n} (\lambda_i - \mu_i) v_i = 0.$$

□

Definition

A linearly independent spanning set for a vector space V is called a *basis* for V.

Examples.

- The empty set is a basis of $\{0\}$.
- $\{1, i\}$ is a basis for the $\mathbb{R}$-vector space $\mathbb{C}$.
- In the plane any two non-parallel vectors form a basis.
- The set $\mathcal{B} = \{e_i, 1 \leq i \leq n\}$ where e_i is the vector with ith coordinate equal to 1 and all others equal to 0, is a basis for $\mathbb{K}^n$. This is clearly seen from $\sum_{i=1}^{n} \lambda_i e_i = (\lambda_1, \lambda_2, \cdots, \lambda_n)$ which
  - means that any element of $\mathbb{K}^n$ can be written as a linear combination of vectors from $\mathcal{B}$;
  - can only be identically zero if all the $\lambda_i = 0$;

  This basis is called the *canonical basis* of $\mathbb{K}^n$.

Theorem

A set $\mathcal{B} = \{e_i, 1 \leq i \leq n\}$ is a basis for V, if and only if for any vector $v \in V$, there exists a unique set $\{\lambda_i \in \mathbb{K},\ 1 \leq i \leq n\}$, such that $v = \sum_{i=1}^{n} \lambda_i e_i$.

Proof. ($\Rightarrow$) The existence of the λ_i is a consequence of $\mathcal{B}$ being a basis, as it spans V. Their unicity was proven in theorem 2.77.

($\Leftarrow$) The existence of the decomposition shows that $\mathcal{B}$ spans V. Now suppose that $\sum_{i=1}^{n} \lambda_i e_i = 0$. Since $0 = \sum_{i=1}^{n} 0 \cdot e_i$, the unicity implies that $\lambda_i = 0$ for all $1 \leq i \leq n$. Thus $\mathcal{B}$ is linearly independent. □

Remark. The following map f is linear,

$$f : \mathbb{K}^n \longrightarrow V$$

$$(\lambda_i)_{1 \leq i \leq n} \longmapsto \sum_{i=1}^{n} \lambda_i v_i.$$

Besides, calling $S = \{v_i, 1 \leq i \leq n\}$,

- f is injective if and only if S is linearly independent in V;
- f is surjective if and only if S is spanning V;
- f is bijective if and only if S a basis for V;

Theorem

Let V and W be two $\mathbb{K}$-vector spaces. Given a basis $\mathcal{B} = \{e_i, 1 \leq i \leq n\}$ for V, and a set of vectors $\{f_i \in W, 1 \leq i \leq n\}$, there exists a unique linear map u defined from V to W such that for any $i \in [\![1, n]\!]$, $u(e_i) = f_i$.

Proof. The decomposition of $v \in V$ over $\mathcal{B}$ being unique (theorem 2.79) we can write $v = \sum_{i=1}^{n} \lambda_i e_i$, with $\lambda_i \in \mathbb{K}$, $1 \leq i \leq n$. Then we can define a map u from V to W, such that $u(v) = \sum_{i=1}^{n} \lambda_i f_i$.

We claim that u is the unique linear map such that $u(e_i) = f_i$.

To verify the linearity of u we take $x = \sum_{i=1}^{n} \lambda_i e_i$ and $y = \sum_{i=1}^{n} \mu_i e_i \in V$, with $\lambda_i, \mu_i \in \mathbb{K}$, for all $1 \leq i \leq n$, as well as two scalars α, β. We get

$$\begin{aligned} u(\alpha x + \beta y) &= u\left(\sum_{i=1}^{n} (\alpha\lambda_i + \beta\mu_i) e_i\right) \\ &= \alpha\left(\sum_{i=1}^{n} \lambda_i f_i\right) + \beta\left(\sum_{i=1}^{n} \mu_i f_i\right) \\ &= \alpha u(x) + \beta u(y). \end{aligned}$$

Furthermore, since for e_k all the $\lambda_i = 0$, but λ_k which is 1, we obtain that $u(e_i) = f_i$.

Note that beyond the existence this also shows the unicity. In fact two maps u_1 and u_2 matching on the e_i, also matches on any element of V.

□

Remark. As we have just seen, two linear maps u and v are equal if and only if they match on a basis of V. In particular $u = 0$ if and only if $u(e_i) = 0$ for all $i \in [\![1, n]\!]$.

Proposition

Let V and W be two vector spaces, and u be a linear map from V to W. Given a basis $\mathcal{B} \subset V$,

i. $u(\mathcal{B})$ spans $\operatorname{im} u$;
ii. u is surjective if and only if $u(\mathcal{B})$ spans W;
iii. u is injective if and only if $u(\mathcal{B})$ is linearly independent;
iv. u is bijective if and only if $u(\mathcal{B})$ is a basis for W;

Proof. Noting that (iv) is a combination of (ii) and (iii) we will only need to prove the other results.

Let $\mathcal{B} = \{e_1, e_2, \cdots, e_n\}$.

(i) Since $\mathcal{B}$ spans V we have $V = \mathbb{K}e_1 + \mathbb{K}e_2 + \cdots + \mathbb{K}e_n$. Then we can use the linearity of u to show that $u(\mathcal{B})$ spans $\operatorname{im} u$:

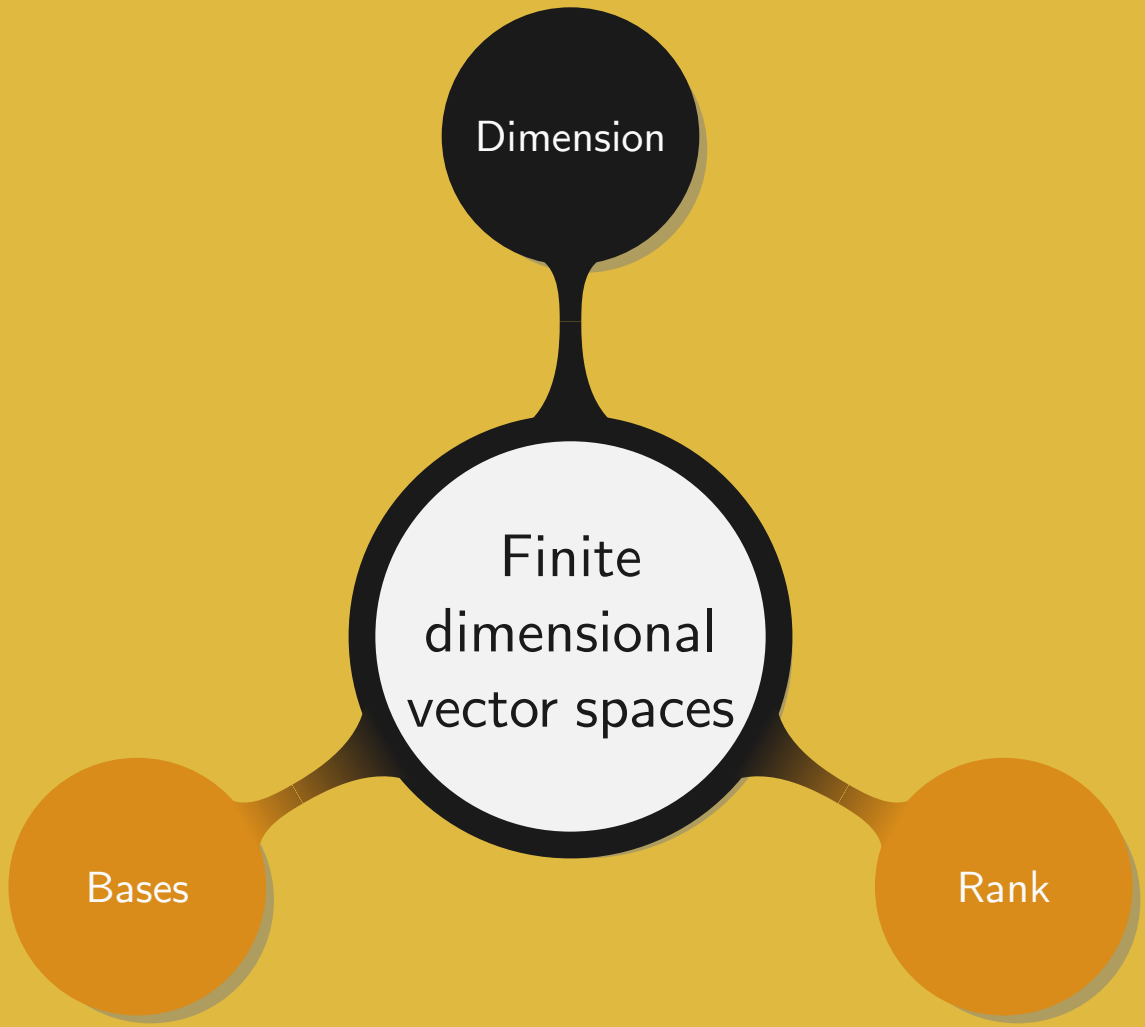
$$\operatorname{im} u = u(V) = \mathbb{K}u(e_1) + \mathbb{K}u(e_2) + \cdots + \mathbb{K}u(e_n).$$

(ii) u is surjective if and only if $\operatorname{im} u = W$, which from (i) means $u(\mathcal{B})$ spans W.

(iii) Suppose u injective and let λ_i, $1 \leq i \leq n$, be such that $\sum_{i=1}^{n} \lambda_i u(e_i) = 0$. By the linearity of u we see that $u\left(\sum_{i=1}^{n} \lambda_i e_i\right) = 0$, i.e. this sum belongs to the kernel of u. But as u is injective, this sum must be zero (theorem 1.52). Moreover as $\mathcal{B}$ is a basis, it is linearly independent and for all $i \in [\![1, n]\!]$, $\lambda_i = 0$. Hence $u(\mathcal{B})$ is linearly independent.

Conversely suppose $u(\mathcal{B})$ to be linearly independent. If we take $x \in \ker u$ then $u(x) = \sum_{i=1}^{n} \lambda_i u(e_i) = 0$. Thus by the linear independence of $u(\mathcal{B})$, all the λ_i are zero, and x is also zero. □

Chapter organisation



Existence of a basis

Definition

A *finite dimensional* vector space is a vector space with a spanning set featuring a finite number of elements. If it is not the case it is said to have *infinite dimension*.

Lemma

Let V be a finite dimensional vector space, and $\mathcal{S}$ be a spanning set for V. Then any linearly independent subset of $\mathcal{S}$ can be completed, using elements of $\mathcal{S}$, into a basis for V.

Proof. Intuitively, if we want a basis we need to extract the largest linearly independent subset of $\mathcal{S}$. This can be easily achieved at the cost of a small technicality.

Let $\mathcal{I}$ be a linearly independent subset of $\mathcal{S}$. Thus if we denote by n and p the cardinalities of $\mathcal{S}$ and $\mathcal{I}$, respectively, we have $p \leq n$.

Existence of a basis

We can then define the set of the cardinalities of all those $\mathcal{I}$. This set is in fact a non-empty subset of $\mathbb{N}$ bounded by n. Hence it has a largest element r, and we can define $\mathcal{I}_r$ as the set $\mathcal{I}$ with cardinal r.

Now that we have extracted the largest linearly independent subset $\mathcal{I}_r$ of $\mathcal{S}$, we need to ensure it still spans the whole space V. Using proposition 2.71 we only need to prove that any element in $\mathcal{S}$ can be written as a linear combination of elements in $\mathcal{I}_r$.

Two cases can arise. Either the element is in $\mathcal{S} \cap \mathcal{I}_r$ or in $\mathcal{S} \setminus \mathcal{I}_r$. The former case being trivially solved, we only address the latter one.

Recalling that $\mathcal{I}_r$ is the largest subset of $\mathcal{S}$ which is linearly independent it is clear that adding a new element s to it will make it linearly dependent. By proposition 2.76 this means that s can be expressed as a linear combination of elements of $\mathcal{I}_r$.

Finally as $\mathcal{I}_r$ is linearly independent and spans V, it is a basis for V. $\square$

Existence of a basis

Theorem

Any finite dimensional $\mathbb{K}$-vector space V has a finite basis. Moreover such a basis can be extracted from any set spanning V.

Proof. Based on lemma 2.85, as soon as we have a set $\mathcal{S}$ spanning V, as well as a linearly independent subset of $\mathcal{S}$, we can expand it to obtain a basis. Thus, noting that the empty set is a linearly independent subset of $\mathcal{S}$ directly yields the result. □

Remark. If $V = \{0\}$, the only linearly independent subset of V is the empty set, which also defines the only basis for V. However in the general case more than one basis exist for a vector space.

Proposition

Let $\mathcal{S}$ be a subset of V with n elements. Any subset of V with cardinal $n+1$, and whose vectors can be expressed as linear combinations of elements of $\mathcal{S}$, is linearly dependent.

Proof. We will prove the result by induction on the cardinality of $\mathcal{S}$.

The property is trivially true for $n = 0$, since a linear combination of 0 vector necessarily yields the null vector.

We now assume the result to be true up to $n-1$, and show it is also true for n. Let $\mathcal{S}_n = \{e_1, e_2, \cdots, e_n\}$ and $\mathcal{T}_n = \{f_1, f_2, \cdots, f_{n+1}\}$. By assumption, for all $j \in [\![1, n+1]\!]$, there exist λ_{ij} such that $f_j = \lambda_{1j}e_1 + \lambda_{2j}e_2 + \cdots + \lambda_{nj}e_n$.

In order to complete the proof we need to distinguish two cases.

First, if $\lambda_{nj} = 0$ for all $1 \leq j \leq n+1$, it means in particular that $\mathcal{T}_{n-1} = \{f_1, \cdots, f_n\}$ are a linear combination of $\{e_1, \cdots, e_{n-1}\}$. We can thus apply our induction hypothesis. Hence $\mathcal{T}_{n-1}$ is linearly dependent, and so is $\mathcal{T}_n$.

Second, if for instance $\lambda_{nn+1} \neq 0$, then we can define $g_j = f_j - \frac{\lambda_{nj}}{\lambda_{nn+1}} f_{n+1}$. In that case we see that

$$g_j = \sum_{i=1}^{n} \lambda_{ij} e_i - \frac{\lambda_{nj}}{\lambda_{n(n+1)}} \sum_{i=1}^{n} \lambda_{i(n+1)} e_i$$

has it e_n term which vanishes. Hence, the g_j can be written as a linear combinations of $e_1, e_2, \cdots, e_{n-1}$ and we can apply the induction hypothesis to get that the g_j are linearly dependent. But as g_j is defined as a linear combination of f_j and f_{n+1}, it means that the f_j are also linearly dependent. We can conclude by the induction principle that the result is verified. $\square$

Theorem

Let V be vector space of finite dimension. Then all the bases for V have same cardinality.

Proof. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two bases of V. Considering that $\mathcal{B}_1$ spans V and $\mathcal{B}_2$ is linearly independent we can conclude that $\operatorname{card} \mathcal{B}_2 \leq \operatorname{card} \mathcal{B}_1$. By symmetry we obtain $\operatorname{card} \mathcal{B}_1 \leq \operatorname{card} \mathcal{B}_2$, and hence the equality. □

Examples.

- The vector space $\{0\}$ has dimension 0.
- $\mathbb{K}$ is a $\mathbb{K}$-vector space of dimension 1.
- $\mathbb{K}[X]$ has infinite dimension.
- $\mathbb{K}_n[X]$ is a $\mathbb{K}$-vector space of dimension $n+1$.

Definition

Let V be a finite dimensional vector space. The cardinal of its bases is called the *dimension* of V and denoted $\dim V$.

Corollary

Let V be a vector space of dimension n. Then

- any linearly independent set has at most n vectors;
- any spanning set has at least n vectors;
- any set of more than n elements is linearly dependent;
- any with less than n elements does not span V;

Proof. The result is straightforward from lemma 2.85 and theorems 2.87 and 2.90. □

Proposition

A vector space V is of infinite dimension if and only if there exists an infinite sequence of vectors $(x_i)_{i\in\mathbb{N}} \in V$ such that for any $n \in \mathbb{N}$, $\{x_0, x_1, \cdots, x_n\}$ is a linearly independent set.

Proof. ($\Leftarrow$) If V has finite dimension n, then such infinite family cannot exist by corollary 2.91.

($\Rightarrow$) If V is of infinite dimension then we can recursively construct a sequence $(x_n)_n \in \mathbb{N}$.

Since V is of infinite dimension it is not restricted to $\{0\}$, and we can find an element $x_0 \neq 0$ in V. Assuming we now have $\mathcal{I}_n = \{x_0, x_1, \cdots, x_n\}$, V is still of infinite dimension, meaning that $\mathcal{I}_n$ cannot span it. Therefore there exists $x_{n+1} \notin \operatorname{span}\mathcal{I}_n$ but is in V. Hence $\mathcal{I}_{n+1} = \mathcal{I}_n \cup x_{n+1}$ is linearly independent.

□

Example. The most common example of infinite dimensional vector space is $\mathbb{K}[X]$. However it is simple to restrict the attention to a finite dimensional vector space by bounding the degree and considering $\mathbb{K}_n[X]$.

Theorem (Incomplete basis)

Any linearly independent subset of a finite dimensional vector space V can be completed into a basis for V.

Proof. The result is clear as soon as we can fall under the assumption of lemma 2.85. Therefore we select two subsets of V: $\mathcal{S}$ which spans V, and $\mathcal{I}$ which is linearly independent. Note that both subsets are finite since V is a finite dimensional vector space.

By adding the elements of $\mathcal{S}$ to $\mathcal{I}$ we obtain a set $\mathcal{S}'$ which spans V and contains a linearly independent subset. Thus we can apply lemma 2.85 and obtain the expected result. □

Theorem

Let $\mathcal{B}$ be a subset of a vector space V of dimension n. The following properties are equivalent:

i. $\mathcal{B}$ is a basis;
ii. $\mathcal{B}$ is linearly independent and is composed of n vectors;
iii. $\mathcal{B}$ spans V and is composed of n vectors;

Proof. (i) $\Rightarrow$ (ii) and (*iii*) A basis of V spans V and is linearly independent.

(i) $\Leftarrow$ (ii) If $\mathcal{B}$ is linearly independent, then it can be completed into a basis (theorem. 2.93), but as all basis have same cardinality (theorem. 2.90) $\mathcal{B}$ is a basis.

(i) $\Leftarrow$ (iii) If $\mathcal{B}$ spans V then we can extract a basis from it (theorem 2.87). Since such a basis has n elements, it means that $\mathcal{B}$ is a basis for V. $\square$

Remark. Theorem 2.94 is a fundamental result when it comes to proving that a subset of a finite dimensional vector space is a basis. Most often the best approach consists in proving that a subset with n vectors is linearly independent.

Example.

- For the $\mathbb{R}$-vector space $\mathbb{C}$, $\{1, \omega\}$ is linearly independent, so a basis, if and only if ω is not real.
- Two non-parallel vectors from a dimension 2 vector space V form a basis for V.

Proposition

Let V be an n-dimensional vector space, and V_0 be a subspace of V. Then V_0 has dimension at most n and $V = V_0$ if and only if V_0 has dimension n.

Proof. Let $\mathcal{I}$ be a linearly independent subset of V_0. Then $\mathcal{I}$ is linearly independent in V and by corollary 2.91 $\operatorname{card}\mathcal{I} \leq n$. Thus there exists a largest linearly independent subset $\mathcal{I}_n$ of V_0. By proposition 2.76, adding elements to $\mathcal{I}_n$ would make it linearly dependent. Hence $\mathcal{I}_n$ is a basis for V_0.

Clearly $\mathcal{I}_n$ will be a basis for V if and only if it is composed of n vectors and in that case $V_0 = V$. □

Examples.

- $\mathbb{K}$ is a 1-dimensional $\mathbb{K}$-vector space. So its subspaces are $\mathbb{K}$ and $\{0\}$, which are of dimensions 1 and 0, respectively.
- Let $V = \mathbb{R}^4$ and $V_0 = \{(x_1, x_2, x_3, x_4), x_1 + x_2 = x_3 + x_4 = 0\}$. The fact than $x \in V$ can be written $(x_1, -x_1, -x_4, x_4)$, hints us that V_0 is a 2-dimensional subspace of $\mathbb{R}^4$. It then suffices to prove for instance that $(1, -1, 0, 0)$ and $(0, 0, -1, 1)$ form a basis for V_0.

Proposition

For any finite n-dimensional vector space V, and any p-dimensional subspace V_1 of V, there exists a subspace V_2 such that $V = V_1 \oplus V_2$.

Proof. Let V be a finite n-dimensional vector space and V_1 be a p-dimensional subspace of V.

From theorems 2.87 and 2.93 we can find a basis $\mathcal{B}_{V_1} = \{e_i, 1 \leq i \leq p\}$ for V_1 and complete it using $\mathcal{B}_{V_2} = \{e_i, p+1 \leq i \leq n\}$ to obtain a basis for V, i.e. $\mathcal{B}_{V_1} \cup \mathcal{B}_{V_2}$ defines a basis for V. So we simply need to verify that $V = V_1 \oplus V_2$, with V_2 the set spanned by $\mathcal{B}_{V_2}$.

First note that $V = V_1 + V_2$ since any $x \in V$ can be written

$$x = \sum_{i=1}^{n} \lambda_i e_i = \sum_{i=1}^{p} \lambda_i e_i + \sum_{i=p+1}^{n} \lambda_i e_i, \text{ with } \lambda_i \in \mathbb{K}.$$

Second note that if we take $x \in V_1 \cap V_2$, then x can be written as

$$x = \sum_{i=1}^{p} \lambda_i e_i = \sum_{i=p+1}^{n} \mu_i e_i, \text{ with } \lambda_i, \mu_i \in \mathbb{K},$$

which leads to

$$\sum_{i=1}^{p} \lambda_i e_i + \sum_{i=p+1}^{n} (-\mu_i) e_i = 0, \text{ with } \lambda_i, \mu_i \in \mathbb{K}.$$

Since $\{e_1, e_2, \cdots, e_n\}$ is linearly independent it means that all the λ_i and μ_i are equal to 0. Thus $V_1 \cap V_2 = \{0\}$ and we can conclude that $V = V_1 \oplus V_2$. □

Theorem

Let V be a finite n-dimensional vector space. A vector space W is isomorphic to V if and only if it has dimension n.

Proof. ($\Rightarrow$) If there exists an isomorphism from V to W then by proposition 2.82 it will map a basis for V to a basis for W. Hence $\dim W = n$.

($\Leftarrow$) Since $\dim V = \dim W$ by theorem 2.80 there exists a unique linear map u mapping a basis for V to a basis for W. Then by proposition 2.82, u is an isomorphism. $\square$

Corollary

Any n-dimensional $\mathbb{K}$-vector space is isomorphic to $\mathbb{K}^n$.

Theorem

Let V and W be two finite dimensional vector spaces. Then the dimension of $V \times W$ is equal to $\dim V + \dim W$.

Dimension of a product of vector spaces

Proof. Let $\mathcal{B}_V = \{e_i, 1 \leq i \leq n\}$ and $\mathcal{B}_W = \{f_j, 1 \leq j \leq p\}$ be some bases for V and W, respectively. We will prove that $\mathcal{B} = \{(e_i, 0), (0, f_j), 1 \leq i \leq n, 1 \leq j \leq p\}$ defines a basis for $V \times W$.

Any element in $V \times W$ can be decomposed into

$$\begin{aligned}(x, y) &= (x, 0) + (0, y) \\ &= \left(\sum_{i=1}^{n} \lambda_i e_i, 0\right) + \left(0, \sum_{j=1}^{p} \mu_j f_j\right) \\ &= \sum_{i=1}^{n} \lambda_i (e_i, 0) + \sum_{j=1}^{p} \mu_j (0, f_j),\end{aligned}$$

showing that $\mathcal{B}$ spans $V \times W$.

To prove that $\mathcal{B}$ is linearly independent we take $\lambda_i, \mu_j \in \mathbb{K}$ with $1 \leq i \leq n$ and $1 \leq j \leq p$ such that $\sum_{i=1}^{n} \lambda_i (e_i, 0) + \sum_{j=1}^{p} \mu_j (0, f_j) = (0, 0)$, and ensure all the λ_i and μ_j are zero.

Dimension of a product of vector spaces

This implies that each of the two sums must be 0. But since $\mathcal{B}_V$ and $\mathcal{B}_W$ are linearly independent it means that all the λ_i, μ_j are zero. Hence $\mathcal{B}$ is also linearly independent, and this completes the proof as it has $n + p$ elements which is equivalent to saying that $V \times W$ has dimension $n + p$ (theorem 2.94).

□

Example. Noting that $\mathbb{R} \times \mathbb{R}$ is a 2-dimensional vector space over $\mathbb{R}$, we recursively obtain that $\mathbb{R}^n$ is an n-dimensional $\mathbb{K}$-vector space.

Theorem

Let V be a finite dimensional vector space, and V_1 and V_2 be two subspaces such that $V = V_1 \oplus V_2$. Then

- $\dim V = \dim V_1 + \dim V_2$;
- given $\mathcal{B}_1 = \{e_i, 1 \leq i \leq p\}$ a basis for V_1 and $\mathcal{B}_2 = \{e_j, p + 1 \leq j \leq n\}$, $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is a basis for V;

Dimension of a direct sum of vector spaces

Proof. For the first point it suffices to notice that

$$\varphi : V_1 \times V_2 \longrightarrow V$$
$$(x_1, x_2) \longmapsto x_1 + x_2$$

is an isomorphism. Indeed φ is clearly linear and since $V_1 \cap V_2 = \{0\}$, $x_1 + x_2$ is uniquely defined by x_1 and x_2. Therefore by theorem 2.99 we obtain $\dim V = \dim V_1 + \dim V_2$.

By definition, $\mathcal{B}$ has n elements and $\dim V = n$. Thus we only need to prove that $\mathcal{B}$ spans V.

Any $x = x_1 + x_2 \in V$ can be decomposed using φ^{-1} into $(x_1, x_2) = \left(\sum_{i=1}^{p} \lambda_i e_i, \sum_{j=p+1}^{n} \lambda_j e_j\right)$ for some λ_i and μ_j. Henceforth $\varphi(x_1, x_2) = \sum_{i=1}^{n} \lambda_i e_i$. In other words any $x \in V$ can be written as a linear combination of elements in $\mathcal{B}$, which is a basis for V. $\square$

Theorem (Grassmann formula)

Let V_1 and V_2 be two subspaces of a finite dimensional $\mathbb{K}$-vector space V. Then $\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2)$.

Proof. Let V_2' be such that $V_2' \oplus (V_1 \cap V_2) = V_2$. Then by theorem 2.101,

$$\dim V_2 = \dim(V_1 \cap V_2) + \dim V_2'. \tag{2.2}$$

We will now verify that $V_1 + V_2 = V_1 \oplus V_2'$. It is trivial that V_1 and V_2' are subspaces of $V_1 + V_2$. Then note that since $V_2' \subset V_2$, any $x \in V_1 \cap V_2'$ also belongs to $V_1 \cap V_2$, so that by definition of V_2', $x = 0$.

For any $x \in V_1 + V_2$ we can find $x_1 \in V_1$ and $x_2 \in V_2$ such that $x = x_1 + x_2$. Recalling that $V_2 = V_2' \oplus (V_1 \cap V_2)$ we can further decompose x_2 into $x_2' + x_2''$, with $x_2' \in V_2'$ and $x_2'' \in V_1 \cap V_2$.

Therefore we have $x = x_1 + x_2'' + x_2'$, with $x_1 + x_2'' \in V_1$ and $x_2' \in V_2'$. We have just showed than $V_1 + V_2 = V_1 + V_2'$. Hence by theorem 2.101 we obtain

$$\dim(V_1 + V_2) = \dim(V_1 \oplus V_2') = \dim V_1 + \dim V_2'. \tag{2.3}$$

Finally after combining (2.2) and (2.3) we get the expected result. □

Remark. Let V_1 and V_2 be two subspaces of a finite dimensional vector space V. To prove that $V = V_1 \oplus V_2$ it suffices to show that $\dim V = \dim V_1 + \dim V_2$, and either of the following two points

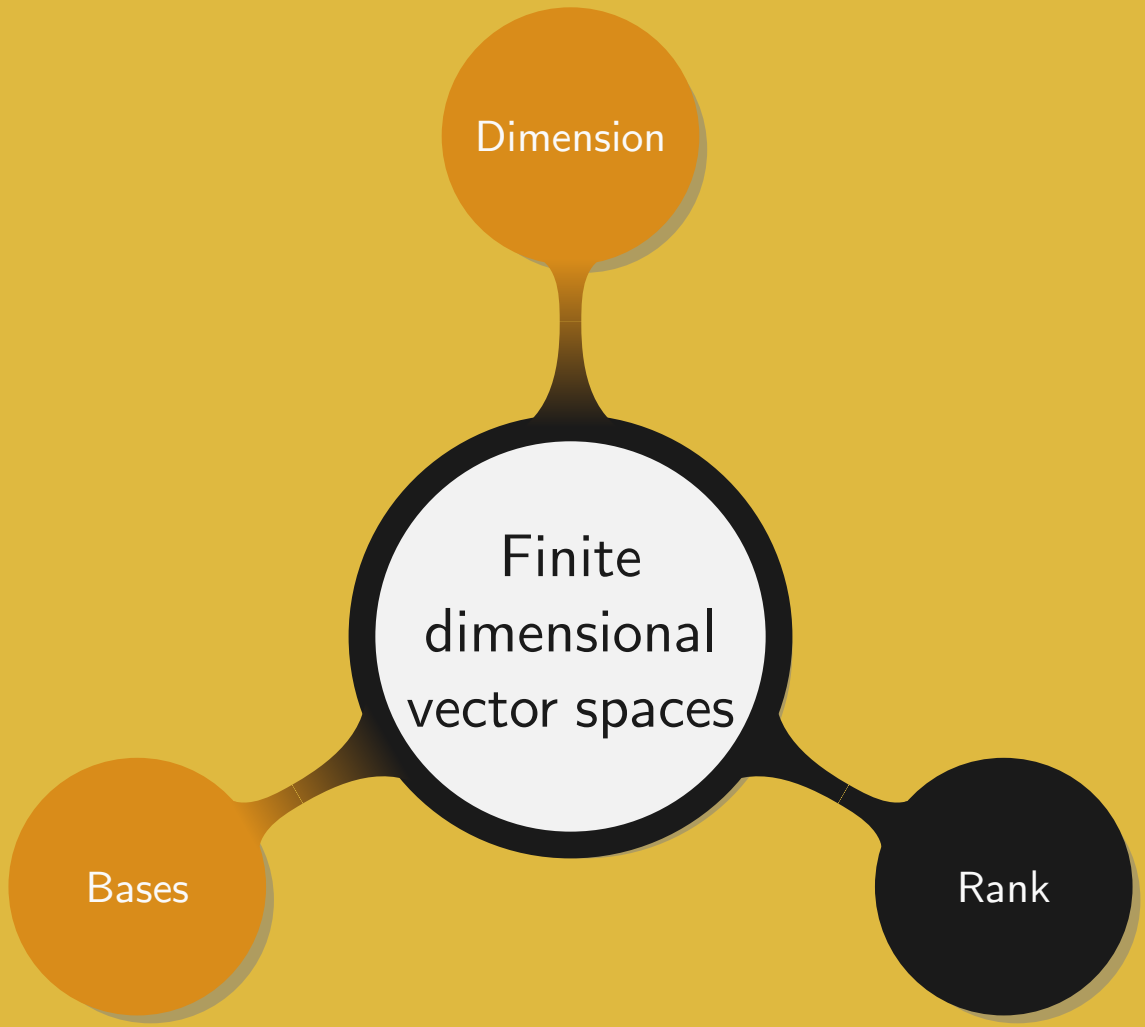
- $V_1 \cap V_2 = \{0\}$: by theorem 2.103,

  $$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 = \dim V,$$

  which shows that $V_1 + V_2$ is a subspace of V with same dimension;
- $V_1 + V_2 = V$: by theorem 2.103 we see that $V_1 \cap V_2 = \{0\}$ since

  $$\dim(V_1 \cap V_2) = \dim V_1 + \dim V_2 - \dim(V_1 + V_2) = 0;$$

Chapter organisation



Definition of the rank

Definitions

1. For a subset S of a vector space V we define the *rank* of S as the dimension of the subspace of V spanned by S.
2. Let V and W be two $\mathbb{K}$-vector spaces and $u \in \mathcal{L}(V, W)$. The rank of u corresponds to the dimension of $\operatorname{im} u$ when it is finite.

Remarks. Let $S = \{x_1, \cdots, x_n\}$ a subset of an n-dimensional vector space V.

- Since S spans $\operatorname{span} S$, a basis for $\operatorname{span} S$ can be extracted. Hence $\operatorname{rank} S$ is at most $\dim V$, with equality if and only if S is linearly independent.
- If S contains r linearly independent vectors, then its rank is at least r, with equality if and only if those r vectors span S.

Definition of the rank

- For a finite n-dimensional space W, the rank of any linear map $u : V \to W$ is less or equal to n, with equality if and only if u is surjective.
- Let V be an n-dimensional vector space with basis $\mathcal{B} = \{e_i, 1 \leq i \leq n\}$. From proposition 2.82 we know that if u is a linear map from V to W then $\operatorname{im} u$ is spanned by $u(\mathcal{B})$. Thus the rank of u is equal to the rank of $u(\mathcal{B})$, i.e. less or equal to n with equality if and only if $u(\mathcal{B})$ forms a basis for $\operatorname{im} u$, or in other words equality happens if and only if u is injective.

Lemma

Let V and W be two vector spaces and u be a linear map from V to W. If V_0 a subspace such that $V = V_0 \oplus \ker u$, then u induces an isomorphism from V_0 to $\operatorname{im} u$.

Proof. The map $v : V_0 \to \operatorname{im} u$, $x \mapsto u(x)$, is well defined since the image of any vector from $V \supset V_0$, belongs to $\operatorname{im} u$. Since it is a restriction of u it is clearly linear. Thus we are only left with proving its bijective property.

First observe that v is injective. We know that $\ker v = \{x \in V_0, u(x) = 0\} = V_0 \cap \ker u$. But as $V = V_0 \oplus \ker u$ this means that $\ker v = \{0\}$. Hence by theorem 1.52, v is injective.

Second we show that v is also surjective. Let $y \in \operatorname{im} u$. Then there exists $x \in V$ such that $u(x) = y$, but as $V = V_0 \oplus \ker u$ we can decompose x into $x_1 + x_2$ with $x_1 \in V_0$ and $x_2 \in \ker u$. Thus we obtain

$$y = u(x) = u(x_1 + x_2) = u(x_1) + u(x_2) = u(x_1) = v(x_1),$$

which shows the surjectivity of v. □

Theorem (Rank theorem)

Let V be a finite dimensional vector space and u be a linear map from V into a vector space W. Then $\dim V = \operatorname{rank} u + \dim \ker u$.

Proof. Let V_0 be a subspace of V such that $V = V_0 \oplus \ker u$. From lemma 2.107 we know that V_0 and $\operatorname{im} u$ have same dimension since they are isomorphic. It directly follows that

$$\dim V = \dim V_0 + \dim \ker u = \dim \operatorname{im} u + \dim \ker u.$$

□

Remark. Note that using the rank theorem (2.109) we can recover Grassmann formula (2.103). It suffices to consider the linear map

$$f : V_1 \times V_2 \longrightarrow V_1 + V_2$$
$$(x_1, x_2) \longmapsto x_1 + x_2.$$

By definition its image is $V_1 + V_2$ and its kernel $\{(x, -x), x \in V_1 \cap V_2\}$. But since $\ker f$ is isomorphic to $V_1 \cap V_2$ by the map $x \mapsto (x, -x)$, this yields $\dim(V_1 \cap V_2) = \dim \ker f$.

Theorem

Let V and W be two finite n-dimensional $\mathbb{K}$-vector spaces, and u be a linear map from V to W. Then the following properties are equivalent.

i. u is injective;
ii. u is surjective;
iii. u is bijective;

Proof. It is clear that it suffices to prove that (i) $\Leftrightarrow$ (ii). By assumption and using the rank theorem (2.109), we have

$$\dim V = \dim W = \dim \operatorname{im} u + \dim \ker u. \tag{2.4}$$

Characterisation of isomorphisms

By theorem 1.52 we know that u is injective if and only if $\ker u = \{0\}$, which by equation (2.4) is equivalent to saying that $\dim W = \dim \operatorname{im} u$. Hence $W = \operatorname{im} u$ which proves the surjectivity of u. □

Remark. From the previous theorem we directly get that if u is an endomorphism over a finite dimensional vector space, then it is equivalent to say that u is injective, surjective, or bijective.

This is however not correct as soon as V is not of finite dimension:

- Over $\mathbb{R}[X]$, $P(X) \mapsto XP(X)$ is injective but not surjective;
- Over $\mathbb{R}[X]$, differentiation is surjective but not injective;

Example. A very useful result from numerical analysis states that given $n+1$ points (x_i, y_i), $0 \leq i \leq n$, there exists a unique polynomial of degree at most n passing through all of them. This polynomial is called the Lagrange interpolation polynomial. Its existence and unicity is in fact a direct consequence of theorem 2.110.

To see it, we consider the x_i as $n+1$ scalars distinct two by two. A polynomial of degree less or equal to n admitting all the x_i as roots is null. This means that the linear map

$$\begin{aligned} f : \mathbb{K}_n[X] &\longrightarrow \mathbb{K}^{n+1} \\ P &\longmapsto \big(P(x_0), P(x_1), \cdots, P(x_n)\big) \end{aligned}$$

is injective. But as $\mathbb{K}_n[X]$ and $\mathbb{K}^{n+1}$ have same dimension, we know that f is an isomorphism (theorem 2.110). This shows that for any set of $n+1$ scalars y_i, there exists a unique polynomial of degree at most n such that $P(x_i) = y_i$, for all $0 \leq i \leq n$.

Proposition

Let U, V, and W be three finite dimensional vector spaces and $u \in \mathcal{L}(U, V)$ and $v \in \mathcal{L}(V, W)$.

- If u is an isomorphism, then $\operatorname{rank}(v \circ u) = \operatorname{rank} v$;
- If v is an isomorphism, then $\operatorname{rank}(v \circ u) = \operatorname{rank} u$;

Proof. We start by observing that

$$\operatorname{im}(v \circ u) = \{v(u(x)), x \in U\} = \{v(y), y \in \operatorname{im} u\} = v(\operatorname{im} u).$$

If u is bijective, then $\operatorname{im} u = V$ and as such $\operatorname{im}(v \circ u) = \operatorname{im} v$, yielding $\operatorname{rank}(v \circ u) = \operatorname{rank} v$.

If v is bijective, then v induces an isomorphism from $\operatorname{im} u$ to $v(\operatorname{im} u)$. Since those two subspaces have same dimension, $\operatorname{rank} u = \operatorname{rank}(v \circ u)$.

□

Definition

An $(n-1)$-dimensional subspace of a finite n-dimensional $\mathbb{K}$-vector space V is called a *hyperplane* of V.

Proposition

Let $\mathcal{H}$ be a hyperplane of a finite n-dimensional vector space V. Then the following properties are equivalent.

i. $\dim \mathcal{H} = n - 1$;

ii. There exists a vectorial line $\mathcal{L}$ such that $V = \mathcal{L} \oplus \mathcal{H}$;

iii. There exists a non-zero linear form f such that $\mathcal{H} = \ker f$

Proof. (i) $\Rightarrow$ (ii) If $\mathcal{H}$ is an $(n-1)$-dimensional subspace of V, then there exists $\mathcal{L}$ such that $V = \mathcal{H} \oplus \mathcal{L}$, and we have $\dim \mathcal{L} = \dim V - \dim \mathcal{H} = 1$.

(ii) $\Rightarrow$ (iii) Let $p : V \to \mathcal{L} \oplus \mathcal{H}$, be a projection. It maps an $x \in V$ to the unique $y \in \mathcal{L}$ such that $x = y + z$, with $z \in \mathcal{H}$.

By construction it is clear that $\mathcal{H} \subset \ker p$. Conversely let $x \in V$ such that $p(x) = 0$. When decomposing x we obtain $x = p(x) + z = z$, with $z \in \mathcal{H}$. Hence $\ker p \subset \mathcal{H}$, and we obtain $\ker p = \mathcal{H}$.
Moreover p is linear since for any $\alpha, \beta \in \mathbb{K}$, and $x, y \in V$

$$p(\alpha x + \beta y) = \alpha x_1 + \beta y_1 = \alpha p(x) + \beta p(y),$$

with x_1 and $y_1 \in \mathcal{L}$.
In other words, p is a non-zero linear map whose kernel is $\mathcal{H}$. Moreover as $\mathcal{L}$ is a vectorial line on $\mathbb{K}$, it can be expressed as $\mathbb{K}d$, and we can define a linear form f such that $p(x) = f(x)d$. In this phrasing, $\mathcal{H}$ is the kernel of f.

(iii) $\Rightarrow$ (i) We apply the rank theorem (2.109) to the linear form f and obtain $\dim V = \dim \operatorname{im} f + \dim \mathcal{H}$. As f is a non-zero linear form its image is a subspace of $\mathbb{K}$, different from $\{0\}$, i.e. its dimension is 1. Hence $\dim \mathcal{H} = n - 1$. $\square$

Hyperplane

Examples.

- In dimension 3, the hyperplanes are the vectorial planes.
- In dimension 2, the hyperplanes are the vectorial lines.

Remarks.

- Any space $V_0 \subsetneq V$ is in at least one hyperplane. Indeed a basis $\{e_1, \cdots, e_p\}$ of V_0 can be completed into a basis $\{e_1, \cdots e_p, \cdots, e_n\}$ of V. The hyperplane defined by $\text{span}\{e_1, \cdots, e_{n-1}\}$ works.
- If a subspace V_0 strictly contains a hyperplane then $V_0 = V$, since its dimension is strictly larger than $n-1$ and smaller or equal to n.
- If a subspace $\mathcal{H} \subsetneq V$ is such that for any subspace W, $\mathcal{H} \subset W \subset V$ implies either $W = \mathcal{H}$ or $W = V$, then $\mathcal{H}$ is a hyperplane of V.

Theorem

Let f and g be two non-zero linear forms such that $\ker f = \ker g$. Then there exists $\lambda \in \mathbb{K}^*$, such that $f = \lambda g$.

Proof. By proposition 2.114, we know the existence of a vectorial line $\mathcal{L}$ such that $\ker f \oplus \mathcal{L} = V$. By construction if we take $e_n \in \mathcal{L}$, we know that $g(e_n) \neq 0$, since $\ker f = \ker g$, and we can define $\lambda = \frac{f(e_n)}{g(e_n)} \neq 0$.
Then as any vector $x \in V$ can be written $x = h + \alpha e_n$, with $h \in \ker f$ and $\alpha \in \mathbb{K}$, we get

$$g(x) = g(h) + \alpha g(e_n) = \alpha g(e_n)$$
$$f(x) = f(h) + \alpha f(e_n) = \alpha f(e_n) = \lambda g(x)$$

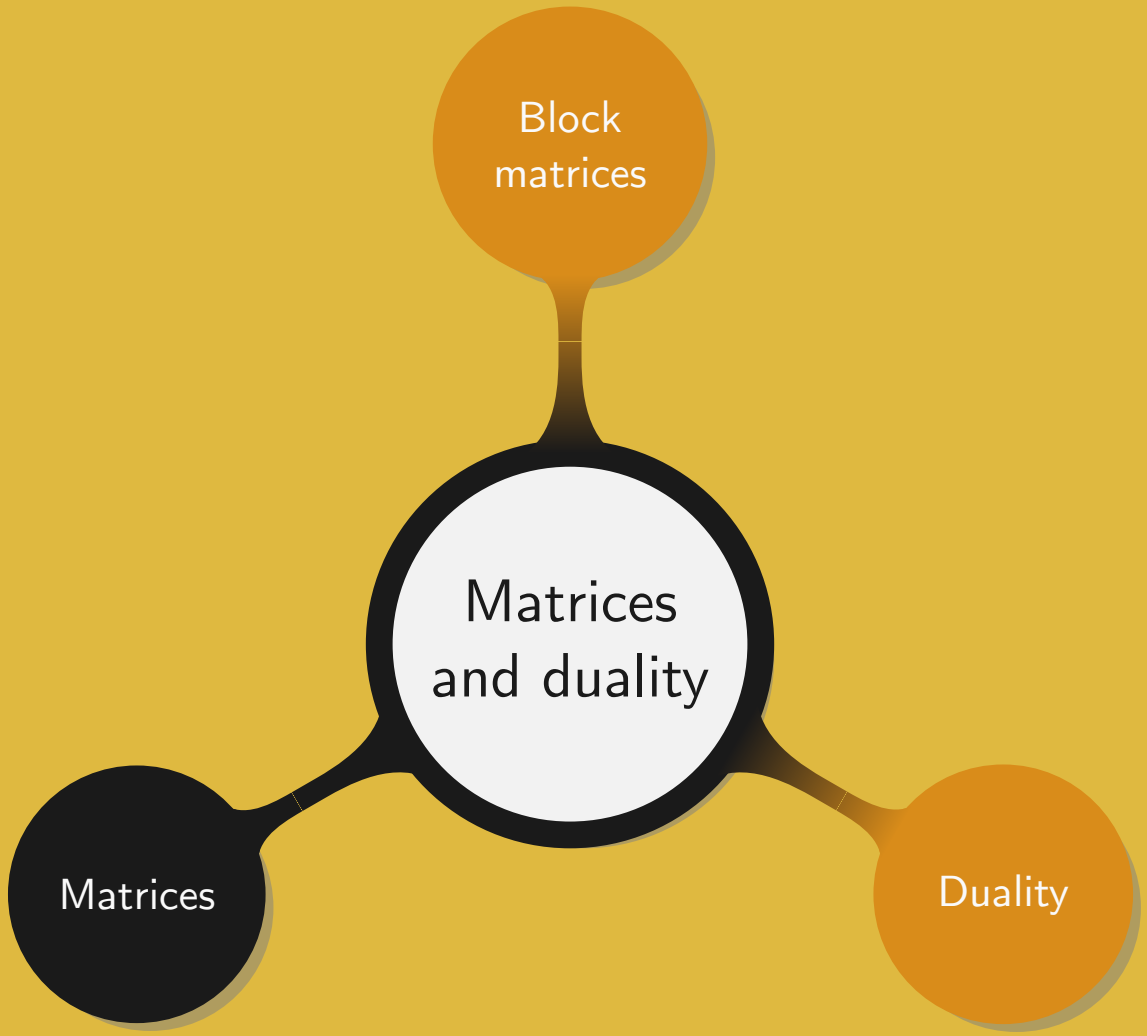
This shows that $f = \lambda g$. □

Key points

- What are a spanning set and a linear independent set?
- What is a basis?
- Are bases of finite or infinite dimension? Give examples.
- What is the rank of a function or of a set?
- How to prove that a linear map is an isomorphism?

3. Matrices and duality

Chapter organisation



Definition

A *matrix* with n *rows*, p *columns*, and *coefficients* in $\mathbb{K}$, is a map from $[\![1, n]\!] \times [\![1, p]\!]$ to $\mathbb{K}$.

Remarks. Common representations for a matrix A:

- Similar to a sequence: $A = \left(a_{i,j}\right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$;
- As an array:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,j} & \cdots & a_{1,p} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,j} & \cdots & a_{2,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i,1} & a_{i,2} & \cdots & a_{i,j} & \cdots & a_{i,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,j} & \cdots & a_{n,p} \end{pmatrix};$$

Remarks. A few remarks regarding common notations.

- $(a_{1,j}, \cdots, a_{n,j})$ is called jth *column vector*;
- $(a_{i,1}, \cdots, a_{i,p})$ is called ith *row vector*;
- A *column matrix* is a matrix with a single column. In that case we omit the column index and only write $(a_i)_{1 \leq i \leq n}$;
- A *row matrix* is a matrix with a single row. In that case we omit the row index and only write $(a_j)_{1 \leq j \leq p}$;
- A *submatrix* of A is the restriction of A to $I \times J$, where $I \subset [\![1, n]\!]$ and $J \subset [\![1, p]\!]$;
- The *transpose* B of A, denoted $A^\top$, is the matrix with n rows and p columns, where $b_{i,j} = a_{j,i}$, for all $1 \leq i \leq n$ and $1 \leq j \leq p$;

Special matrices

- If $p = n$, then A is said to be a *square matrix*;
- For A a square matrix:
  - If $\forall i, j$ such that $i > j$, $a_{i,j} = 0$, then A is *upper triangular*;
  - If $\forall i, j$ such that $i < j$, $a_{i,j} = 0$, then A is *lower triangular*;
  - A is upper triangular if and only if its transpose is lower triangular;
  - If $\forall i, j$ such that $i \neq j$, $a_{i,j} = 0$, then A is *diagonal*;
  - A diagonal matrix with all its diagonal elements equal is *scalar*;
  - If $A^\top = A$, then A is *symmetric*;
  - If $A^\top = -A$, where $-A$ is A with all its elements mulitplied by -1, then A is *antisymmetric*;

Notations for special sets of matrices

Commonly used notations:

- The $n \times p$ matrices with coefficients in $\mathbb{K}$: $\mathcal{M}_{n,p}(\mathbb{K})$;
- The square matrices with coefficients in $\mathbb{K}$: $\mathcal{M}_n(\mathbb{K})$;
- The upper triangular matrices of $\mathcal{M}_n(\mathbb{K})$: $\mathcal{T}_n(\mathbb{K})$;
- The diagonal matrices of $\mathcal{M}_n(\mathbb{K})$: $\mathcal{D}_n(\mathbb{K})$;
- The symmetric matrices of $\mathcal{M}_n(\mathbb{K})$: $\mathcal{S}_n(\mathbb{K})$;
- The antisymmetric matrices of $\mathcal{M}_n(\mathbb{K})$: $\mathcal{A}_n(\mathbb{K})$;
- The $n \times n$ scalar matrix with elements equal to 1: I_n;

So far we have mainly focused on vector spaces and linear maps, proving various fundamental and important results. Then the introduction of the notions of dimension and of basis paved the way for future applications.

Matrix of a linear map

We now want to consider what we have previously achieved from a new perspective, which might be easier to use and apply when it comes to representing and sharing information with computers. We therefore turn our attention to how to translate the previous results in term of matrices.

Let V and W be two $\mathbb{K}$-vector spaces of finite dimension p and n, respectively. Let $\mathcal{B}_V = \{e_1, \cdots, e_p\}$ be a basis for V and $\mathcal{B}_W = \{f_1, \cdots, f_n\}$ be a basis for W. A linear map u, from V to W, is uniquely determined by the vectors $u(e_j)$, $1 \leq j \leq p$, which in turn can be uniquely decomposed over the basis $\mathcal{B}_W$.

Those notations will be used for the rest of the chapter.

Definition

The *matrix* of u with respect to the basis $\mathcal{B}_V$ and $\mathcal{B}_W$, is the matrix whose jth column, $1 \leq j \leq p$, is composed of the $u(e_j)$ expressed in $\mathcal{B}_W$. In compact form it can be written $M_{\mathcal{B}_V,\mathcal{B}_W}(u) = (a_{i,j})_{\substack{1\leq i\leq n\\ 1\leq j\leq p}}$, with $u(e_j) = \sum_{i=1}^{n} a_{i,j} f_i$, for all $1 \leq j \leq p$.

This analytical representation of u is very useful in practice as for any $x \in V$ expressed in the basis $\mathcal{B}_V$ it is easy to determine the decomposition of $y = u(x)$ in $\mathcal{B}_W$. Noting the decomposition of x and y, $(x_1, \cdots, x_p)$ and $(y_1, \cdots, y_n)$, respectively, we obtain

$$\sum_{i=1}^{n} y_i f_i = u\left(\sum_{j=1}^{p} x_j e_j\right) = \sum_{j=1}^{p} x_j u(e_j).$$

Then replacing $u(e_j)$ by its definition we can further get

$$\sum_{i=1}^{n} y_i f_i = \sum_{j=1}^{p} x_j \left(\sum_{i=1}^{n} a_{i,j} f_i \right) = \sum_{j=1}^{p} \sum_{i=1}^{n} x_j a_{i,j} f_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{p} x_j a_{i,j} \right) f_i.$$

Finally by the unicity of the decomposition into base $\mathcal{B}_W$, we see that $y_i = \sum_{j=1}^{p} a_{i,j} x_j$. In other words y_i can be expressed using the elements of the ith row of the matrix A of u.

Conversely, to any matrix $A = (a_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}$ one can associate a linear map u which maps a vector $x = (x_1, \cdots, x_p)$ in $\mathcal{B}_V$ to a vector $y = (y_1, \cdots, y_n)$ in $\mathcal{B}_W$.

Remarks.

- When u is an endomorphism, the two bases are often chosen identical.
- When u is a linear form, the scalar 1 is often used as basis, and A is a row matrix.

Example. Let V_1 and V_2 be two subspaces of V such that $V_1 \oplus V_2 = V$. We call *symmetry* with respect to V_1 and parallel to V_2 the function s which maps any element $v = v_1 + v_2$, with $v_1 \in V_1$ and $v_2 \in V_2$, to $s(v) = v_1 - v_2$.

In $\mathbb{R}^2$, we consider the symmetry with respect with the line V_1 of equation $y + 2x = 0$ and parallel to $y - x = 0$. We want to determine its matrix in the canonical basis $\mathcal{B}$ of $\mathbb{R}^2$.

First note that taking two vectors on the lines $y + 2x = 0$ and $y - x = 0$ will define a basis (example 2.95). For instance $\mathcal{B}_s = \{(1, -2), (1, 1)\}$ is a basis of $\mathbb{R}^2$, and the matrix of s in $\mathcal{B}_s$ is $M_s = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. This is clearly seen as a vector v decomposed into $v_1 + v_2$ will have its v_1 component unchanged while the v_2 one will "point" in the opposite direction, when s is applied to it. Hence, this gives the two column vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ composing M_s.

We have the matrix M_s of s in the basis $\mathcal{B}_s$, but we need it in the canonical basis. From proposition 2.82, we know the existence of an isomorphism transforming a basis into a basis. That is, there exists a linear bijection u such that $u(\mathcal{B}) = \mathcal{B}_s$ and $u^{-1}(\mathcal{B}_s) = \mathcal{B}$. Hence given a vector v in the canonical basis $\mathcal{B}$, it suffices to use u to map it to a vector v' in the basis $\mathcal{B}_s$, apply the symmetry s to v', and finally map back $s(v')$ to $\mathcal{B}$. In other words we need to determine the matrix of the linear map $u^{-1} \circ s \circ u$.

To know how to "render" the inner composition law for functions in term of matrix, we first need to investigate the algebraic structure of $\mathcal{M}_{n,p}$.

The set $\mathcal{M}_{n,p}$ is naturally endowed with a structure of vector space for the inner law $(+)$ and the outer law $(\cdot)$ defined by

$$\lambda \cdot \left(a_{i,j}\right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}} + \mu \cdot \left(b_{i,j}\right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}} = \left(\lambda \cdot a_{i,j} + \mu \cdot b_{i,j}\right)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}}.$$

The unit element for $(+)$ is the matrix uniquely composed of zeros.

Theorem

Let V and W be two $\mathbb{K}$-vector spaces of dimensions p and n, with bases $\mathcal{B}_V$ and $\mathcal{B}_W$, respectively. The map

$$\begin{aligned} \varphi : \mathcal{L}(V,W) &\longrightarrow \mathcal{M}_{n,p}(\mathbb{K}) \\ u &\longmapsto M_{\mathcal{B}_V,\mathcal{B}_W}(u), \end{aligned}$$

is an isomorphism of vector spaces.

Proof. Let $\mathcal{B}_V = \{e_1, \cdots, e_p\}$ and $\mathcal{B}_W = \{f_1, \cdots, f_n\}$. Let u and v be two linear maps from V to W and λ and μ be two scalars. Then

$$A = \left(a_{i,j}\right)_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} = M_{\mathcal{B}_V,\mathcal{B}_W}(u)$$

$$B = \left(b_{i,j}\right)_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} = M_{\mathcal{B}_V,\mathcal{B}_W}(v)$$

$$C = \left(c_{i,j}\right)_{\substack{1\leq i\leq n \\ 1\leq j\leq p}} = M_{\mathcal{B}_V,\mathcal{B}_W}(\lambda u + \mu v)$$

As in the matrix of a linear map the scalar $c_{i,j}$ corresponds to the ith component of $\lambda u + \mu v$ in the basis $\mathcal{B}_W$, and

$$\begin{aligned}(\lambda u + \mu v)(e_j) &= \lambda u(e_j) + \mu v(e_j) \\ &= \sum_{i=1}^{n} \left(\lambda a_{i,j} + \mu b_{i,j}\right) f_i,\end{aligned}$$

it shows that $c_{i,j} = \lambda a_{i,j} + \mu b_{i,j}$. Moreover since $M_{\mathcal{B}_V,\mathcal{B}_W}(\lambda u + \mu v) = C = \lambda A + \mu B = \lambda M_{\mathcal{B}_V,\mathcal{B}_W}(u) + \mu M_{\mathcal{B}_V,\mathcal{B}_W}(v)$, we get the linearity of φ.

Finally the bijection directly follows from the definition of a matrix. Indeed since for any matrix there exists a unique linear map this yields the result. □

At this stage we would like to enrich the algebraic structure of the set of matrices by importing the ring structure of $\mathcal{L}(V)$ (proposition 1.56).

Matrix of a linear map

Let U, V, and W be three vector spaces with bases $\mathcal{B}_U = \{u_1, \cdots, u_r\}$, $\mathcal{B}_V = \{v_1, \cdots, v_q\}$, and $\mathcal{B}_W = \{w_1, \cdots, w_p\}$, respectively.

We then define the linear maps $f \in \mathcal{L}(U, V)$ and $g \in \mathcal{L}(V, W)$, and denote $M_f = \left(a_{j,k}\right)_{\substack{1 \leq j \leq q \\ 1 \leq k \leq r}}$ the matrix of f with respect to $\mathcal{B}_U$ and $\mathcal{B}_V$, and $M_g = \left(b_{i,j}\right)_{\substack{1 \leq j \leq p \\ 1 \leq j \leq q}}$ the matrix of g with respect to $\mathcal{B}_V$ and $\mathcal{B}_W$.

From theorem 3.130, there must exist some isomorphisms between $\mathcal{L}(U, V)$ and $\mathcal{M}_{q,r}(\mathbb{K})$, $\mathcal{L}(V, W)$ and $\mathcal{M}_{p,q}(\mathbb{K})$, and $\mathcal{L}(U, W)$ and $\mathcal{M}_{p,r}(\mathbb{K})$. We therefore need to construct a "counter-part" inner-law which is compatible with $(\circ)$. We will call this operation *matrix multiplication*.

Matrix of the composition of two functions

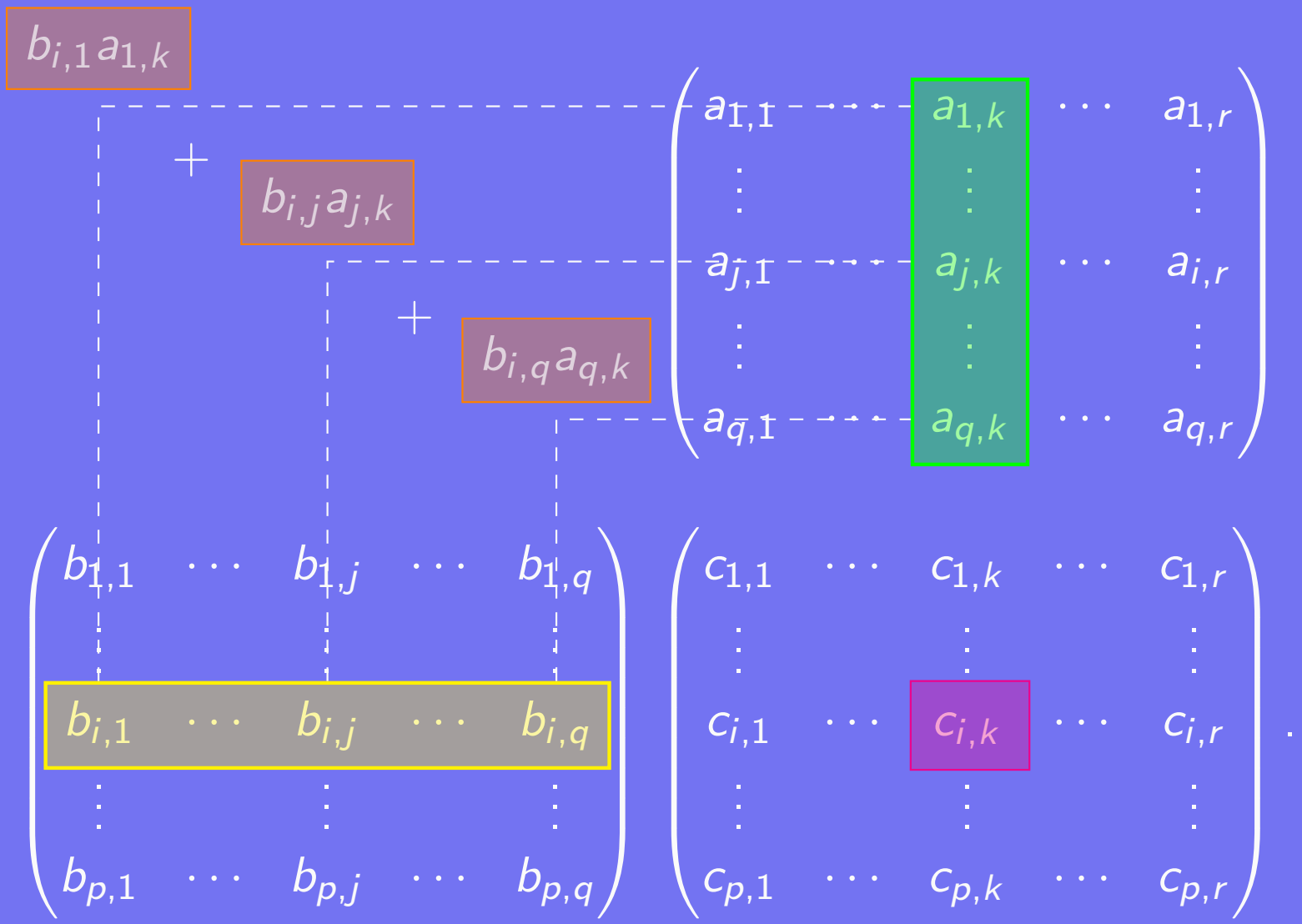
We want to determine the matrix of $g \circ f$ with respect to the bases $\mathcal{B}_U$ and $\mathcal{B}_W$. Therefore we expect a matrix with p rows and r columns. For any $u_k \in \mathcal{B}_U$, we have

$$\begin{aligned} g \circ f(u_k) &= g\left(\sum_{j=1}^{q} a_{j,k} v_j\right) = \sum_{j=1}^{q} a_{j,k} g(v_j) \\ &= \sum_{j=1}^{q} a_{j,k} \left(\sum_{i=1}^{p} b_{i,j} w_i\right) = \sum_{j=1}^{q} \sum_{i=1}^{p} a_{j,k} b_{i,j} w_i \\ &= \sum_{i=1}^{p} \left(\sum_{j=1}^{q} b_{i,j} a_{j,k}\right) w_i. \end{aligned}$$

If we call this matrix $M_{g\circ f} = \left(c_{i,k}\right)_{\substack{1\leq i\leq p \\ 1\leq k\leq r}}$, then $c_{i,k} = \sum_{j=1}^{q} b_{i,j} a_{j,k}$.

Matrix multiplication

Hence, matrix multiplication is defined as



Theorem

Let V be an n-dimensional $\mathbb{K}$-vector space and $\mathcal{B}$ be a basis of V.

- $\left(\mathcal{M}_n(\mathbb{K}), +, \times\right)$ is a ring.
- The map

$$\begin{aligned} \varphi : \mathcal{L}(V) &\longrightarrow \mathcal{M}_n(\mathbb{K}) \\ u &\longmapsto M_{\mathcal{B}}(u) \end{aligned}$$

  is a ring isomorphism.

Proof. From theorem 3.130 we know that $\left(\mathcal{M}_n(\mathbb{K}), +, \cdot\right)$ is a vector space, so in particular $\left(\mathcal{M}_n(\mathbb{K}), +\right)$ is an abelian group.

By construction $(\times)$ is closed, associative, and distributive with respect to $(+)$. Moreover the matrix of the identity map is composed of 1 on the diagonal and 0 everywhere else. It is the unit element for $\mathcal{M}_n(\mathbb{K})$.

Finally φ is a ring isomorphism since it is a vector space isomorphism and by construction $M(\mathrm{Id}_v) = I_n$ and $M_{\mathcal{B}}(u \circ v) = M_{\mathcal{B}}(u) \times M_{\mathcal{B}}(v)$. □

Remarks.

- The ring isomorphism from theorem 3.135 allows us to import all the algebraic structure from $\mathcal{L}(V, W)$ into $\mathcal{M}_{n,p}(\mathbb{K})$. In particular since two maps are not necessarily commutative, then two matrices do not always commute.
- Given two linear maps $u, v \in \mathcal{L}(V)$, $u \circ v = \mathrm{Id}$ does not always imply $v \circ u = \mathrm{Id}$. Indeed it fails as soon as V is not a finite dimensional vector space. For instance over $\mathbb{R}[X]$, if we take differentiation for u and anti-differentiation with constant term 0 for v, then we see that for any $P \in \mathbb{R}[X]$, $(u \circ v)(P) = P$. However when considering $v \circ u$ the constant term gets lost.
- Since we have defined matrices with respect to finite dimensional spaces, saying that $AB = I$ is that same as saying that $BA = I$, where A, B are two matrices.
- If $u : V \to W$ is a linear map and M is the matrix representing u, then we denote by $\operatorname{null} M$ the set $\{v \in V, Mv = 0\} = \ker u$.

Example. We can now complete example 3.128 by determining the matrix M of $u^{-1} \circ s \circ u$, where u is the linear map transforming elements expressed in the canonical basis $\mathcal{B}$ into elements written in $\mathcal{B}_s$. Moreover we know that the matrix of s in $\mathcal{B}_s = \{(1,-2),(1,1)\}$ is $M_s = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

The matrix representing the change of basis from $\mathcal{B}_s$ to the canonical basis $\mathcal{B}$ is given by $M_{u^{-1}} = \left(\begin{smallmatrix} 1 & 1 \\ -2 & 1 \end{smallmatrix}\right)$. Furthermore, since we have a ring isomorphism between $\mathcal{L}(V)$ and $\mathcal{M}_n(\mathbb{K})$, the fact that u^{-1} is invertible means that $M_{u^{-1}}$ is also invertible, i.e. there exists $M_u \in \mathcal{M}_n(\mathbb{K})$ such that $M_{u^{-1}} \times M_u = I_n$.

To determine $M_{u^{-1}}$ we simply express the vectors of $\mathcal{B}$ in basis $\mathcal{B}_s$. In particular observe that $\frac{1}{3}(1,-2)+\frac{2}{3}(1,1) = (1,0)$ and $-\frac{1}{3}(1,-2)+\frac{1}{3}(1,1) = (0,1)$. As a result $M_u = \frac{1}{3}\left(\begin{smallmatrix} 1 & -1 \\ 2 & 1 \end{smallmatrix}\right)$.

Finally we have $M_{u^{-1} \circ s \circ u} = M_{u^{-1}} \times M_s \times M_u$, i.e.

$$M_{u^{-1} \circ s \circ u} = -\frac{1}{3}\begin{pmatrix} 1 & 2 \\ 4 & -1 \end{pmatrix}.$$

Definition

For two bases $\mathcal{B}$ and $\mathcal{B}'$, the *transition matrix* from $\mathcal{B}$ to $\mathcal{B}'$, denoted $M_{\mathcal{B}\to\mathcal{B}'}$, has its jth column composed of the jth vector of $\mathcal{B}$ expressed in $\mathcal{B}'$.

Proposition

Let V and W be two finite dimensional vector spaces and u be a linear map from V to W. If

- $\mathcal{B}_V$ and $\mathcal{B}'_V$ are bases of V and $P = P_{\mathcal{B}'_V\to\mathcal{B}_V}$,
- $\mathcal{B}_W$ and $\mathcal{B}'_W$ are bases of W and $Q = Q_{\mathcal{B}'_W\to\mathcal{B}_W}$,
- $M = M_{\mathcal{B}_V,\mathcal{B}_W}(u)$ and $M' = M'_{\mathcal{B}'_V,\mathcal{B}'_W}(u)$,

then $M' = Q^{-1}MP$.

Proof. Let x and y be two vectors of V and W, respectively. We denote X and Y the column matrices representing x and y in $\mathcal{B}_V$ and $\mathcal{B}_W$, respectively. We denote X' and Y' the column matrices representing x and y in $\mathcal{B}'_V$ and $\mathcal{B}'_W$, respectively.

Then we have $X = PX'$ and $Y = QY'$. In particular when $y = u(x)$ we can write $Y = MX$ and get $QY' = MPX'$. Recalling that Q is the matrix representing a change of basis it is invertible and we finally obtain $Y' = Q^{-1}MPX'$, which shows that the matrix of u with respect to bases $\mathcal{B}'_V$ and $\mathcal{B}'_W$ is $Q^{-1}MP$. □

Corollary

Let u be an endomorphism on a finite dimensional vector space V, and $\mathcal{B}$ and $\mathcal{B}'$ be two bases of V. If we call M and M' the matrices of u in bases $\mathcal{B}$ and $\mathcal{B}'$, respectively, then $M' = P^{-1}MP$, with P the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$.

Example. We revisit example 3.137 in terms of matrices, and consider the matrix representing the change of basis from $\mathcal{B}_s$ to $\mathcal{B}$ given by $P = \left(\begin{smallmatrix} 1 & 1 \\ -2 & 1 \end{smallmatrix}\right)$, i.e. each column corresponds to a vector of $\mathcal{B}_s$ expressed in $\mathcal{B}$. Similarly we can determine P^{-1} by expressing the vectors of $\mathcal{B}$ in $\mathcal{B}_s$, which yields $P^{-1} = \frac{1}{3}\left(\begin{smallmatrix} 1 & -1 \\ 2 & 1 \end{smallmatrix}\right)$.

Recalling that $M_s = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ is the matrix of s in $\mathcal{B}_s$ and applying corollary 3.139 we see that the matrix M of s in the canonical basis is given by $M = PM_sP^{-1}$. This results in

$$M = \begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \frac{1}{3}\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} = -\frac{1}{3}\begin{pmatrix} 1 & 2 \\ 4 & -1 \end{pmatrix}.$$

Definition

Two matrices M and M' in $\mathcal{M}_n(\mathbb{K})$ are said to be *similar* if there exists a matrix $P \in \mathrm{GL}_n(\mathbb{K})$ such that $M' = P^{-1}MP$

Theorem

Let V and W be two finite dimensional vector spaces. Then $\mathcal{L}(V, W)$ has dimension $\dim V \dim W$.

Proof. The result is clear since $\mathcal{M}_{n,p}(\mathbb{K})$ is of dimension np and by theorem 3.130, $\mathcal{L}(V, W)$ and $\mathcal{M}_{n,p}(\mathbb{K})$ are isomorphic. □

Examples.

- $\mathcal{T}_n(\mathbb{K})$ is an $\frac{n(n+1)}{2}$-dimensional vector subspace of $\mathcal{M}_n(\mathbb{K})$;
- $\mathcal{D}_n(\mathbb{K})$ is an n-dimensional vector subspace of $\mathcal{M}_n(\mathbb{K})$;

Proposition

Matrix transpose is an isomorphism from $\mathcal{M}_{n,p}(\mathbb{K})$ to $\mathcal{M}_{p,n}(\mathbb{K})$.

Proof. This is trivial since transposition is linear and for a matrix M we have $(M^\top)^\top = M$. □

Remark. Matrix transpose is a symmetry. Its invariants are the elements of $\mathcal{S}_n(\mathbb{K})$ and the set of its elements "pointing" in the opposite direction is $\mathcal{A}_n(\mathbb{K})$. Hence $\mathcal{S}_n(\mathbb{K}) \oplus \mathcal{A}_n(\mathbb{K}) = \mathcal{M}_n(\mathbb{K})$.

In term of dimension we see that $\dim \mathcal{S}_n(\mathbb{K}) = n(n+1)/2$, implying that $\dim \mathcal{A}_n(\mathbb{K}) = n(n-1)/2$, since $\dim M_n(\mathbb{K}) = n^2$.

Proposition

Let $A \in \mathcal{M}_{p,q}(\mathbb{K})$ and $B \in \mathcal{M}_{q,r}(\mathbb{K})$. Then $(AB)^\top = B^\top A^\top$.

Proof. If we call $C = AB$, then we know that C has p rows and r columns, hence $C' = (AB)^\top$ has r rows and p columns. Besides

$$c'_{k,i} = c_{i,k} = \sum_{j=1}^{q} a_{i,j} b_{j,k}.$$

Since $A' = A^\top \in \mathcal{M}_{q,p}(\mathbb{K})$ and $B' = B^\top \in \mathcal{M}_{r,q}(\mathbb{K})$, the product $D = B^\top A^\top$ features r rows and p columns. Then the relation

$$d_{k,i} = \sum_{j=1}^{q} b'_{k,j} a'_{j,i} = \sum_{j=1}^{q} b_{j,k} a_{i,j} = c'_{k,i}$$

shows that the matrices $(AB)^\top$ and $B^\top A^\top$ are equal. □

Proposition

The set $\mathcal{T}_n(\mathbb{K})$ of the upper triangular matrices is a vector subspace and a subring of $\mathcal{M}_n(\mathbb{K})$.

Proof. It is clear that it is a subspace of $\mathcal{M}_n(\mathbb{K})$ containing I_n.

To see that the product of two upper triangular matrices is an upper triangular matrix, let A and B be in $\mathcal{T}_n(\mathbb{K})$.

For any $i, k \in [\![1, n]\!]$, such that $i > k$ we have for $C = AB$

$$c_{i,k} = \sum_{j=1}^{n} a_{i,j} b_{j,k}. \tag{3.1}$$

Observe that if $j > k$, then $b_{j,k} = 0$ since B is upper triangular, and if $j \leq k$, then $j < i$ and $a_{i,j} = 0$. Thus (3.1) is 0, and C is upper triangular. Hence $\mathcal{T}_n(\mathbb{K})$ is closed for $(\times)$. □

Exercise. Show that the set of the lower triangular matrices is a vector subspace and subring of $\mathcal{M}_n(\mathbb{K})$.

Proposition

The set $\mathcal{D}_n(\mathbb{K})$ of the diagonal matrices is a vector subspace and a commutative subring of $\mathcal{M}_n(\mathbb{K})$.

Proof. $\mathcal{D}_n(\mathbb{K})$ is clearly a subspace since it is the intersection of the sets of the upper and lower matrices.

The product is commutative since multiplying two diagonal matrices together is the same as multiplying them "element by element". □

Proposition

Let $A \in \mathcal{M}_n(\mathbb{K})$. If for any $X \in \mathcal{M}_n(\mathbb{K})$, $AX = 0$ implies $X = 0$, then A is invertible.

Proof. This result simply means that the linear map associated to A is injective. But as the dimension is finite it is also bijective. □

Remark. The set of all the invertible matrices of $\mathcal{M}_n(\mathbb{K})$ is noted $\mathrm{GL}_n(\mathbb{K})$.

Proposition

If $A \in \mathrm{GL}_n(\mathbb{K})$, then $A^\top \in \mathrm{GL}_n(\mathbb{K})$ and $(A^\top)^{-1} = \left(A^{-1}\right)^\top$.

Proof. It suffices to observe that using proposition 3.142 we have

$$\left(A^{-1}\right)^\top A^\top = \left(AA^{-1}\right)^\top = I_n^\top = I_n.$$

□

Example. Construction of $\mathbb{C}$. For $a, b \in \mathbb{R}$ we define $M(a,b) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, and define the set $\mathcal{C} = \left\{M(a,b), (a,b) \in \mathbb{R}^2\right\} = \mathbb{R}I + \mathbb{R}J$, with

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We immediately see that $\mathcal{C}$ is a 2-dimensional vector subspace of $\mathcal{M}_2(\mathbb{R})$ and that $\{I, J\}$ defines a basis.

Construction of $\mathbb{C}$

A quick calculation yields $J^2 = -I$, which in turn leads to

$$M(a,b)M(c,d) = M(ac - bd, ad + bc).$$

This show that in $\mathcal{C}$ multiplication is stable and commutative. Moreover as the unit element I is in $\mathcal{C}$ we can conclude that it is a subring of $\mathcal{M}_2(\mathbb{R})$.

The equality $M(a,b)M(a,b)^\top = (a^2 + b^2)I$ shows that all non-zero elements of $\mathcal{C}$ are invertible for $(\times)$. Hence it is a field.

Finally we define the injective ring homomorphism $\lambda \mapsto \lambda I$ from $\mathbb{R}$ to $\mathcal{C}$. Hence $\mathbb{R}$ is isomorphic to a subfield of $\mathcal{C}$. By identifying each real λ to its corresponding matrix λI and setting $i = J$, we obtain a field isomorphic to $\mathbb{C}$, i.e. $\mathcal{C}$ can be identified to $\mathbb{C}$.

Remark. In the above construction the complex number conjugation corresponds to the matrix transpose.

Rank of a matrix

Similarly to how we defined the rank of a linear map (definition 2.106) we now introduce the concept of rank of a matrix.

Definition

Let $A \in \mathcal{M}_{n,p}(\mathbb{K})$, we call *rank* of A the rank of the set defined by the column vectors of A. It is also called the *column space* of A and in that case is denoted $\operatorname{col} A$.

Proposition

Let V and W be two finite dimensional vector spaces, with bases $\mathcal{B}_V$ and $\mathcal{B}_W$, respectively. The rank of a linear map $u \in \mathcal{L}(V, W)$ is equal to the rank of the matrix $M_{\mathcal{B}_V, \mathcal{B}_W}(u)$.

Proof. First note that the rank of any subset S of vectors of V is equal to the rank of the matrix representing S in the base $\mathcal{B}_V$.

Then for any $u \in \mathcal{L}(V, W)$ we write $M_{\mathcal{B}_V, \mathcal{B}_W}(u) = M_{\mathcal{B}_W}(u(\mathcal{B}_V))$, and obtain $\operatorname{rank} u = \dim \operatorname{im} u = \dim \operatorname{span} u(\mathcal{B}_V)$. □

Theorem

A matrix $M \in \mathcal{M}_{n,p}(\mathbb{K})$ has rank r, if and only if there exists $Q \in \mathrm{GL}_n(\mathbb{K})$ and $P \in \mathrm{GL}_p(\mathbb{K})$ such that $M = QJ_rP$, with

$$J_r = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \cdots & 0 & 0 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} r \\ \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} n-r \end{matrix}$$

$$\underbrace{\phantom{\begin{matrix}a&a\end{matrix}}}_{r} \quad \underbrace{\phantom{\begin{matrix}a&a\end{matrix}}}_{p-r}$$

Proof. ($\Rightarrow$) Let M be a matrix with rank r, and u be its corresponding linear map defined from $\mathbb{K}^p$ to $\mathbb{K}^n$. Thus from proposition 3.148 we know that u has rank r, and by the rank theorem (2.109) $\dim \ker u = p - r$. Then we can construct a basis of $\mathbb{K}^p$, by choosing a basis of $\ker u$, $\{v_{r+1}, \cdots, v_p\}$, and completing it using a basis of a subspace V_0 in direct sum with $\ker u$. Hence we have obtained $\mathcal{B}_V = \{v_1, \cdots, v_r, v_{r+1}, \cdots, v_p\}$ a basis of $\mathbb{K}^p$.

We define $w_i = u(v_i)$, for all $1 \leq i \leq r$. By lemma 2.107 we know that u induces an isomorphism from V_0 on $\operatorname{im} u$, implying that the image $\{w_1, \cdots, w_r\}$ of a basis of V_0 is linearly independent. It can therefore be completed into a basis $\mathcal{B}_W = \{w_1, \cdots, w_r, w_{r+1}, \cdots, w_n\}$ of W (theorem 2.93).

By construction the matrix $M_{\mathcal{B}_V, \mathcal{B}_W}(u)$ is J_r. The result is then clear as soon as P is taken as the matrix transforming the canonical basis of $\mathbb{K}^p$ into $\mathcal{B}_V$ and Q as the matrix transforming $\mathcal{B}_W$ into the canonical basis of $\mathbb{K}^n$.

($\Leftarrow$) Conversely, let $M = QJ_rP$ and call f the automorphism from $\mathbb{K}^p$ associated to P, g the automorphism of $\mathbb{K}^n$ associated to Q, and h the linear map from $\mathbb{K}^p$ to $\mathbb{K}^n$ associated to J_r.

In this context QJ_rP is the matrix of the linear map $g \circ h \circ f$. Since both f and g are isomorphism we can apply proposition 2.112 and conclude that $g \circ h \circ f$ has rank r, or in other words M has rank r. □

Corollary

Let $A \in \mathcal{M}_{n,p}(\mathbb{K})$. Then the rank of $A^\top$ is equal to the rank of A.

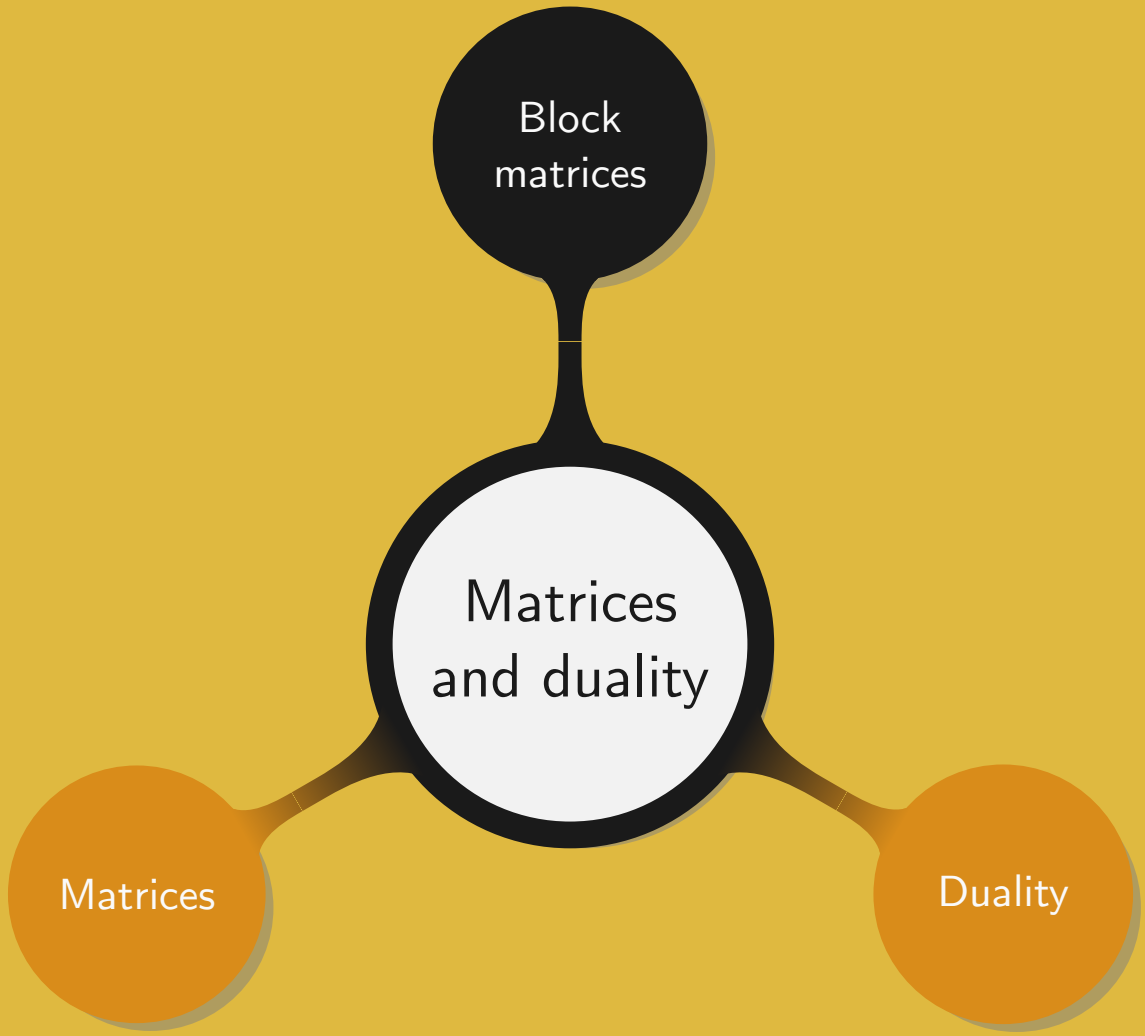
Proof. Let r be the rank of A. From theorem 3.149 we know the existence of $P \in \mathrm{GL}_p(\mathbb{K})$ and $Q \in \mathrm{GL}_n(\mathbb{K})$ such that $A = QJ_rP$. Then by proposition 3.142 we get $A^\top = P^\top J_r^\top Q^\top$.

$P^\top$ and $Q^\top$ are invertible since they are the transposes of invertible matrices, and $J_r^\top = J_r$. Thus we can apply theorem 3.149 to complete the proof. □

Remarks. Let $A \in \mathcal{M}_{n,p}(\mathbb{K})$.

- Since the ranks of a matrix and of its transpose are equal, we can say that the rank of A is equal to
  - the rank of its column vectors;
  - the rank of its rows vectors;
  - any linear map it can represent;
- As the rank of A is less or equal to n and less or equal to p, it is less or equal to $\min(n, p)$.
- An upper triangular matrix is invertible if and only if all its diagonal elements are non-zero.

Chapter organisation



Matrix setup

When dealing with complex systems it is not uncommon for matrices to feature a very large number of rows and columns. In such a case splitting it into blocks can ease calculations, especially if patterns emerge from the matrix. In this part of the chapter we will consider how to perform common matrix operations in term of blocks instead of elements.

In this section we will use the following notations.

- $A = (a_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq p}} \in \mathcal{M}_{n,p}(\mathbb{K})$, with $n, p \in \mathbb{N}^*$;
- $s, t \in \mathbb{N}^*$ such that $\sum_{i=0}^{s} n_i = n$ and $\sum_{j=0}^{t} p_j = p$, with all the $n_i, p_j \in \mathbb{N}^*$, and $n_0 = p_0 = 0$;
- $\sigma_k = \sum_{i=0}^{k} n_i$, for $k \in [\![0, s]\!]$;
- $\tau_l = \sum_{j=0}^{l} p_j$, for $l \in [\![0, t]\!]$;
- Upper and lower case letters will refer to blocks and elements, respectively;

Matrix setup

Using the previous notations we group the elements of A by block

$$\left(\begin{array}{c|c|c|c|c}
\begin{matrix} a_{1,1} & \cdots & a_{1,\tau_1} \\ \vdots & & \vdots \\ a_{\sigma_1,1} & \cdots & a_{\sigma_1,\tau_1} \end{matrix} & \begin{matrix} a_{1,\tau_1+1} & \cdots & a_{1,\tau_2} \\ \vdots & & \vdots \\ a_{\sigma_1,\tau_1+1} & \cdots & a_{\sigma_1,\tau_2} \end{matrix} & \cdots & \begin{matrix} a_{1,\tau_{t-1}+1} & \cdots & a_{1,p} \\ \vdots & & \vdots \\ a_{\sigma_1,\tau_{t-1}+1} & \cdots & a_{\sigma_1,p} \end{matrix} \\
\hline
\begin{matrix} a_{\sigma_1+1,1} & \cdots & a_{\sigma_1+1,\tau_1} \\ \vdots & & \vdots \\ a_{\sigma_2,1} & \cdots & a_{\sigma_2,\tau_1} \end{matrix} & \begin{matrix} a_{\sigma_1+1,\tau_1+1} & \cdots & a_{\sigma_1+1,\tau_2} \\ \vdots & & \vdots \\ a_{\sigma_2,\tau_1+1} & \cdots & a_{\sigma_2,\tau_2} \end{matrix} & \cdots & \begin{matrix} a_{\sigma_1+1,\tau_{t-1}+1} & \cdots & a_{\sigma_1+1,p} \\ \vdots & & \vdots \\ a_{\sigma_2,\tau_{t-1}+1} & \cdots & a_{\sigma_2,p} \end{matrix} \\
\hline
\vdots & \vdots & \ddots & \vdots \\
\hline
\begin{matrix} a_{\sigma_{s-1}+1,1} & \cdots & a_{\sigma_{s-1}+1,\tau_1} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,\tau_1} \end{matrix} & \begin{matrix} a_{\sigma_{s-1}+1,\tau_1+1} & \cdots & a_{\sigma_{s-1}+1,\tau_2} \\ \vdots & & \vdots \\ a_{n,\tau_1+1} & \cdots & a_{n,\tau_2} \end{matrix} & \cdots & \begin{matrix} a_{\sigma_{s-1}+1,\tau_{t-1}+1} & \cdots & a_{\sigma_{s-1}+1,p} \\ \vdots & & \vdots \\ a_{n,\tau_{t-1}+1} & \cdots & a_{n,p} \end{matrix}
\end{array}\right).$$

Definition

For $k \in [\![1, s]\!]$ and $l \in [\![1, t]\!]$ the matrix

$$B_{k,l} = \begin{pmatrix} a_{\sigma_{k-1}+1,\tau_{l-1}+1} & \cdots\cdots & a_{\sigma_{k-1}+1,\tau_l} \\ \vdots & & \vdots \\ a_{\sigma_k,\tau_{l-1}+1} & \cdots\cdots & a_{\sigma_k,\tau_l} \end{pmatrix}$$

of $\mathcal{M}_{n_k,p_l}$ is called the (k,l)th *block of A with respect to the decomposition into n_i and p_j*, where

$$A = \begin{pmatrix} \boxed{B_{1,1}} & \cdots & \boxed{B_{1,t}} \\ \vdots & & \vdots \\ \boxed{B_{s,1}} & \cdots & \boxed{B_{s,t}} \end{pmatrix} \begin{matrix} \updownarrow n_1 \\ \\ \updownarrow n_s \end{matrix}.$$

$$\underset{p_1}{\longleftrightarrow} \qquad \underset{p_t}{\longleftrightarrow}$$

Remarks.

- If A is a square matrix, then usually s and t are chosen to be equal and $(n_1, \cdots, n_s) = (p_1, \cdots, p_s)$. In that case the blocks B_{kk} are called *diagonal blocks*, and they are square.
- Let V be a finite n-dimensional $\mathbb{K}$-vector space, and V_0 be a p-dimensional subspace of V. An endomorphism $u \in \mathcal{L}(V)$ is closed on V_0 if and only if there exists a basis $\mathcal{B} = \{e_1, \cdots, e_n\}$ such that

$$\begin{cases} (e_1, \cdots, e_p) & \text{is a basis of } V_0 \\ M_{\mathcal{B}}(u) & \text{can be written } \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}, \end{cases}$$

where $A \in \mathcal{M}_{n,p}(\mathbb{K})$. It is the matrix in $\{e_1, \cdots, e_p\}$, of the endomorphism induced by u on V_0.

Internal and external compositions by blocks

When considering $\mathcal{M}_{n,p}(\mathbb{K})$ as a vector space we want to know how to perform the basic $(+)$ and $(\cdot)$ operations.

Proposition

If A and B are two block matrices decomposed in a similar way, then for any $\lambda \in \mathbb{K}$, $\lambda A + B$ is a block matrix decomposed similarly to A and B, and in particular

$$\lambda \begin{pmatrix} A_{1,1} & \cdots & A_{1,t} \\ \vdots & & \vdots \\ A_{s,1} & \cdots & A_{s,t} \end{pmatrix} + \begin{pmatrix} B_{1,1} & \cdots & B_{1,t} \\ \vdots & & \vdots \\ B_{s,1} & \cdots & B_{s,t} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1}+B_{1,1} & \cdots & \lambda A_{1,t}+B_{1,t} \\ \vdots & & \vdots \\ \lambda A_{s,1}+B_{s,1} & \cdots & \lambda A_{s,t}+B_{s,t} \end{pmatrix}$$

Figuring out how to perform matrix multiplication by blocks is slightly more complex due to our initial construction of matrix multiplication. This is however possible, and in practice it in fact saves much time to work with blocks rather than elements.

Proposition

Let $A \in \mathcal{M}_{n,p}$ and $B \in \mathcal{M}_{p,q}$, be two block matrices with block decomposition $(n_1, \cdots, n_s) \times (p_1, \cdots, p_t)$ and $(n'_1, \cdots, n'_{s'}) \times (p'_1, \cdots, p'_{t'})$, respectively. In order for A and B to be compatible for the product we expect $s' = t$ and $(n'_1, \cdots, n'_{s'}) = (p_1, \cdots, p_t)$. Then AB is given by

$$\begin{pmatrix} \sum_{j=1}^{s'} A_{1,j} B_{j,1} & \cdots & \sum_{j=1}^{s'} A_{1,j} B_{j,t'} \\ \vdots & & \vdots \\ \sum_{j=1}^{s'} A_{s,j} B_{j,1} & \cdots & \sum_{j=1}^{s'} A_{s,j} B_{j,t'} \end{pmatrix},$$

and the blocks are of size $(n_1, \cdots, n_s) \times (p'_1, \cdots, p'_{t'})$.

Proof. Let $(i, j') \in [\![1, n]\!] \times [\![1, q]\!]$. Then their exists a unique $(k, l') \in [\![1, s]\!] \times [\![1, t']\!]$ such that $n_0 + \cdots + n_{k-1} + 1 \leq i \leq n_0 + \cdots + n_k$ and $p'_0 + \cdots + p'_{l'} + 1 \leq j' \leq p'_0 + \cdots + p'_{l'}$.

The element (i, j') of AB is given by

$$\sum_{j=1}^{p} a_{i,j} b_{j,j'} = \sum_{j=1}^{p_1} a_{i,j} b_{j,j'} + \sum_{j=p_1+1}^{p_1+p_2} a_{i,j} b_{j,j'} + \cdots + \sum_{j=p_1+\cdots+P_{t-1}+1}^{p} a_{i,j} b_{j,j'}.$$

But in fact each sum corresponds to the products of two blocks. In particular in the above sum we recognise the elements of $A_{k,1} B_{1,l'}$, $A_{k,2} B_{2,l'}$, and $A_{k,s'} B_{s',l'}$ located at position

$$\left(i - (n_0 + \cdots + n_{k-1}), j' - (p'_0 + \cdots + p'_{l'-1})\right).$$

□

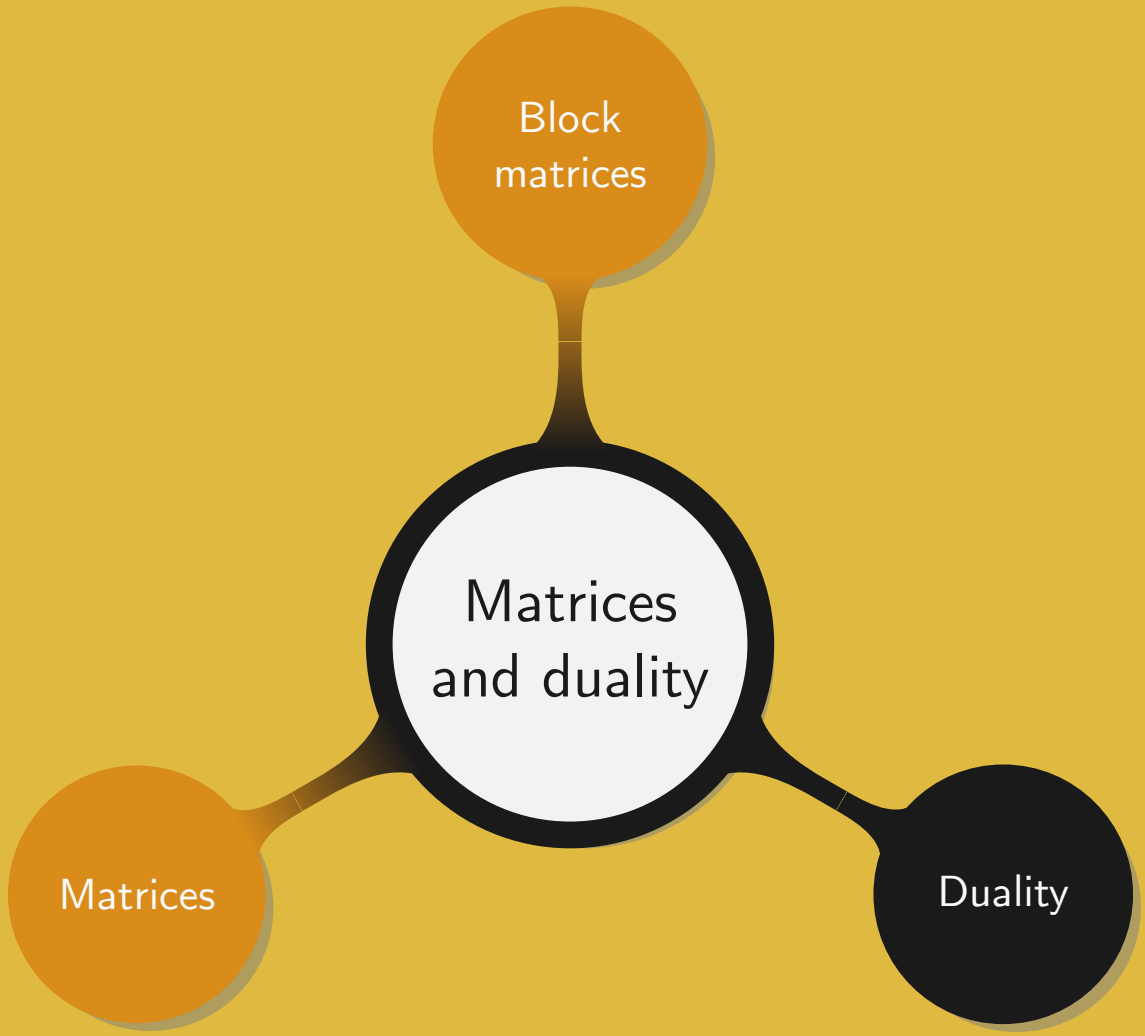
Remark. When performing block multiplications it is important to follow the "correct order", i.e. not change the order of the blocks in the multiplications. A block can be viewed as a matrix, and as such two bocks generally do not commute.

Proposition

For any block decomposition,

$$\begin{pmatrix} A_{1,1} & \cdots & A_{1,t} \\ \vdots & & \vdots \\ A_{s,1} & \cdots & A_{s,t} \end{pmatrix}^{\top} = \begin{pmatrix} A_{1,1}^{\top} & \cdots & A_{s,1}^{\top} \\ \vdots & & \vdots \\ A_{1,t}^{\top} & \cdots & A_{s,t}^{\top} \end{pmatrix}.$$

Chapter organisation



Duality

Duality is a broad concept appearing in various fields on mathematics. From an informal and general perspective when a unique relation can be established between two spaces of objects, using a common property that they share, then the objects of one space are said to be the duals of the objects in the other space. As a result a duality can be expressed as a bijection between the two spaces. While duality can take many different forms, it always features two properties:

- *Symmetry:* if a is the dual of b, then b is the dual of a;
- *Idempotence:* the dual of the dual of a is a;

Definition (Dual space)

Let V be a $\mathbb{K}$-vector space. The set of the linear forms $V^* = \mathcal{L}(V, \mathbb{K})$ is called the *dual* of V.

Remark. The dual is sometimes equivalently defined as $V^* = \mathrm{Hom}(V, \mathbb{K})$, i.e. it is the set of the homomorphisms from V into $\mathbb{K}$.

Remarks. Let V be a vector space.

- V^* is a vector space.
- From proposition 2.114, $\mathcal{H}$ is a hyperplane of V if and only if there exists $\varphi \in V^* \setminus \{0\}$ such that $\mathcal{H} = \ker \varphi$. In particular we see that the hyperplanes of V are the kernels of the non-zero linear forms over V. We say that the relation $\varphi(x) = 0$ is an *equation* of the hyperplane $\mathcal{H}$.
- The map

$$\begin{aligned} V \times V^* &\longrightarrow \mathbb{K} \\ (x, \varphi) &\longmapsto \varphi(x) \end{aligned}$$

is a bilinear form called the *evaluation* of φ on x. It is commonly denoted $\langle x, \varphi \rangle$.

Theorem

Let $\mathcal{B} = \{b_1, \cdots, b_n\}$ be a basis of V. For all $1 \leq i \leq n$ we define the linear form f_i such that for any $j \in [\![1, n]\!]$,

$$f_i : V \longrightarrow \mathbb{K}$$

$$b_j \longmapsto \delta_{i,j} \quad \text{with } \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

The set $\mathcal{B}^* = \{f_1, \cdots, f_n\}$ is a basis of V^*.

Proof. By construction the f_i, $1 \leq i \leq n$, all belong to V^*. Then we can reason by equivalence. Since the dimension is finite, $\mathcal{B}^*$ is a basis if and only if any linear form φ can be decomposed using $\lambda_1, \cdots, \lambda_n \in \mathbb{K}$, into $\varphi = \sum_{i=1}^{n} \lambda_i f_i$.

This is true if and only for any $j \in [\![1, n]\!]$, $\varphi(b_j) = \left(\sum_{i=1}^{n} \lambda_i f_i\right)(b_j)$, or equivalently if and only if for all $j \in [\![1, n]\!]$, $\varphi(b_j) = \lambda_j$. Hence any $\varphi \in V^*$ can be written $\varphi = \sum_{i=1}^{n} \varphi(b_i) f_i$. □

Corollary

Using the previous notations, $\dim V^* = \dim V$.

Remarks.

- As V and V^* have same finite dimension, this corollary shows the existence of an isomorphism between them. Similarly there is an isomorphism between V^* and $V^{**} = (V^*)^*$. Hence, V and V^{**} are isomorphic, meaning that they can be identified.
- The isomorphism between V and V^* is not a *canonical map*, i.e. it does not arise "naturally" from the construction or definition.
- When the dimension is infinite, in the general case the homomorphism from V to V^{**} is injective but not necessarily surjective.

Definition

Let $\mathcal{B} = \{b_1, \cdots, b_n\}$ be a basis of a vector space V. The basis $\mathcal{B}^* = \{f_1, \cdots, f_n\}$, with $f_i(b_j) = 1$ if $i = j$, and 0 otherwise, is called the *dual basis* of $\mathcal{B}$. Besides, f_i is called the ith *coordinate function* with respect to $\mathcal{B}$.

Proposition

Let $\mathcal{B} = \{b_1, \cdots, b_n\}$ be a basis of V, and $\mathcal{B}^* = \{f_1, \cdots, f_n\}$ be its dual. Then

$$\forall \varphi \in V^*,\ \varphi = \sum_{i=1}^{n} \varphi(b_i) f_i \quad \text{and} \quad \forall x \in V,\ x = \sum_{i=1}^{n} f_i(x) b_i.$$

Proof. The first equality was shown in the proof of theorem 3.165, and the second one renders the concept of coordinate function. □

Proposition

Let $\mathcal{B}$ be a basis of V and $\mathcal{B}^*$ be its dual. For $x \in V$ and $\varphi \in V^*$, then $\varphi(x) = M_{\mathcal{B}^*}(\varphi)^\top M_{\mathcal{B}}(x)$.

Proof. Let $\mathcal{B} = \{b_1, \cdots, b_n\}$ be a basis of V, and $\mathcal{B}^* = \{f_1, \cdots, f_n\}$ be a basis of V^*.

Since $\varphi = \sum_{i=1}^n \varphi(b_i) f_i$, we have $M_{\mathcal{B}^*}(\varphi) = \begin{pmatrix} \varphi(b_1) \\ \vdots \\ \varphi(b_n) \end{pmatrix}$. Then for $x \in V$,

$$\varphi(x) = \varphi\left(\sum_{i=1}^n x_i b_i\right) = \sum_{i=1}^n x_i \varphi(b_i) = \left(\varphi(b_1), \cdots, \varphi(b_n)\right) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

□

Remark. Calling $\mathcal{B}_0 = \{1\}$ the canonical basis of $\mathbb{K}$, viewed as a $\mathbb{K}$-vector space, the previous proposition simply says that for any linear form φ, $M_{\mathcal{B}^*}(\varphi) = \left(M_{\mathcal{B},\mathcal{B}_0}(\varphi)\right)^\top$.

Observe that to apply a linear map f to a column vector v we right multiply the matrix representation of f by v. The essence of the previous proposition is to say that taking the transpose of v leads to a row vector, i.e. $v^\top \in V^*$. As a result $v^\top$ can then be applied, or right multiplied in term of matrices, by some other matrix or vector.

Example. In physics, a force is represented as a column vector F. However if we are interested in how it affects the displacement of some object, we look at how F "acts" on it by calculating a scalar called the work. In that case F is interpreted as an element of the dual space.

Following this idea, we want to translate the notion of matrix transpose in term of linear map. Note that this make sense as there is a ring isomorphism between $\mathcal{M}_{n,p}(\mathbb{K})$ and $\mathcal{L}(V, W)$ (theorems 3.135).

A key observation to define the transpose of a linear map, is that the counterpart to matrix left multiply by a transpose should be left function composition by an element of the dual space.

Definition

Let V and W be two vector spaces and V^* and W^* be their dual, respectively. For a linear map $f \in \mathcal{L}(V, W)$, we define $f^\top$ the *transpose* of f as

$$\begin{aligned} f^\top : W^* &\longrightarrow V^* \\ w^* &\longmapsto w^* \circ f. \end{aligned}$$

Remarks.

- Expressed in term of evaluation map (remark 3.164), this is equivalent to $\langle u, f^\top(w^*)\rangle = \langle f(u), w^*\rangle$, for all $u \in V$ and $w^* \in W^*$.
- The transpose $f^\top$ of f is sometimes referred to as the *dual map* of f.

At this stage we want to ensure that

$$\begin{aligned} \top : \mathcal{L}(V, W) &\longrightarrow \mathcal{L}(W^*, V^*) \\ f &\longmapsto f^\top \end{aligned}$$

is linear and consistent with matrix transpose, i.e. the matrix of $f^\top$ is the transpose of the matrix of f.

First we address the linearity of $\top$. Clearly, since f and w^* are linear their composition is also linear, i.e. $f^\top$ is a linear map. Using the linearity of the maps, we take $u \in V$, $\alpha, \beta \in \mathbb{K}$, and $f, g \in \mathcal{L}(V, W)$, to obtain

$$\begin{aligned} \top\left((\alpha f + \beta g)(u)\right) &= (\alpha f + \beta g)^\top(u) = \langle u, (\alpha f + \beta g)^\top(w^*)\rangle \\ &= \langle (\alpha f + \beta g)(u), w^*\rangle = \alpha\langle f(u), w^*\rangle + \beta\langle g(u), w^*\rangle \\ &= \alpha\langle u, f^\top(w^*)\rangle + \beta\langle u, g^\top(w^*)\rangle \\ &= \alpha f^\top + \beta g^\top = \alpha\top(f) + \beta\top(g). \end{aligned}$$

Before constructing the matrix of $f^\top$, quickly observe that

$$\top\big((f\circ g)(u)\big) = \langle u, (f\circ g)^\top(w^*)\rangle = \langle f\circ g(u), w^*\rangle$$
$$= \langle g(u), f^\top(w^*)\rangle = \langle u, g^\top\circ f^\top(w^*)\rangle.$$

Then by the bilinearity of the evaluation map, we get $(f\circ g)^\top - g^\top\circ f^\top = 0$, which is consistent with the transpose of a matrix product (proposition 3.142).

For V and W, we fix two bases $\mathcal{B}_V = \{v_1, \cdots, v_p\}$ and $\mathcal{B}_W = \{w_1, \cdots, w_n\}$, respectively, and write $M_{\mathcal{B}_V,\mathcal{B}_W}(f) = (a_{i,j})_{\substack{1\leq i\leq n\\ 1\leq j\leq p}}$, with all the $a_{i,j}\in\mathbb{K}$.

Applying theorem 3.165, we build the dual bases $\mathcal{B}_{V^*} = \{v_1^*, \cdots, v_p^*\}$ and $\mathcal{B}_{W^*} = \{w_1^*, \cdots, w_n^*\}$. Then observing that f determines $f^\top$ as a linear map from W^* into V^* (definition 3.170), we get $M_{\mathcal{B}_{W^*},\mathcal{B}_{V^*}}(f^\top) = (b_{j,i})_{\substack{1\leq j\leq p\\ 1\leq i\leq n}}$, with all the $b_{j,i}\in\mathbb{K}$.

As a result of both definition and proposition 3.167, for $i \in [\![1, n]\!]$ and $j \in [\![1, p]\!]$,

$$(f^\top w_i^*)(v_j) = \left(\sum_{j=1}^{p} b_{j,i} v_j^*\right)(v_j) = b_{j,i}.$$

On the other hand, if we directly calculate it, for $i \in [\![1, n]\!]$ and $j \in [\![1, p]\!]$, we get

$$(f^\top w_i^*)(v_j) = w_i^*(f(v_j)) = w_i^*\left(\sum_{i=1}^{n} a_{i,j} w_i\right) = a_{i,j}.$$

This shows that for all $(i, j) \in [\![1, n]\!] \times [\![1, p]\!]$, $a_{i,j} = b_{j,i}$, or in other words $M_{\mathcal{B}_{W^*}, \mathcal{B}_{V^*}}(f^\top) = M_{\mathcal{B}_V, \mathcal{B}_W}(f)^\top$.

Theorem

Let $\mathcal{B}$ and $\mathcal{B}'$ be two bases of a vector space V and M be the matrix for the change of basis from $\mathcal{B}'$ to $\mathcal{B}$. Then the transition matrix from $\mathcal{B}'^*$ to $\mathcal{B}^*$ is $(M^{-1})^\top$.

Proof. Let $\mathcal{B} = \{b_1, \cdots, b_n\}$, and $\mathcal{B}' = \{b'_1, \cdots, b'_n\}$. The matrix M is composed of $m_{i,j}$, and we call $P = (p_{i,j})_{1 \leq i,j \leq n}$ the matrix from $\mathcal{B}'^* = \{f'_1, \cdots, f'_n\}$ to $\mathcal{B}^* = \{f_1, \cdots, f_n\}$. Then for any $j, k \in [\![1, n]\!]$,

$$f'_j(b'_k) = \left(\sum_{i=1}^{n} p_{i,j} f_i\right)\left(\sum_{l=1}^{n} m_{l,k} b_l\right)$$
$$= \sum_{i=1}^{n}\sum_{l=1}^{n} p_{i,j} m_{l,k} \delta_{i,l} = \sum_{i=1}^{n} p_{i,j} m_{i,k}.$$

Noting that $f'_j(b'_k) = \delta_{j,k}$, we have $P^\top M = I_n$, and $P = (M^{-1})^\top$. □

Example. Let $\mathcal{B} = \{(2,1,4),(3,2,3),(-1,-1,2)\}$. Show that $\mathcal{B}$ is a basis for $\mathbb{R}^3$, and determine its dual basis.

The matrix M mapping the canonical basis to $\mathcal{B}$ is

$$\begin{pmatrix} 2 & 3 & -1 \\ 1 & 2 & -1 \\ 4 & 3 & 2 \end{pmatrix}.$$

This is meaningful since saying "$\mathcal{B}$ is a basis" is equivalent to saying "u is an isomorphism", i.e. M has an inverse. In fact we have

$$\begin{pmatrix} 2 & 3 & -1 \\ 1 & 2 & -1 \\ 4 & 3 & 2 \end{pmatrix} \begin{pmatrix} 7 & -9 & -1 \\ -6 & 8 & 1 \\ -5 & 6 & 1 \end{pmatrix} = I_3,$$

which means that $\mathcal{B}$ is a basis of $\mathbb{R}^3$. Besides, by theorem 3.174, $\mathcal{B}^* = \{(7,-9,-1)^\top,(-6,8,1)^\top,(-5,6,1)^\top\}$ is a basis of the dual of $\mathbb{R}^3$.

Example. Let $x_0, \cdots, x_n$ be $n+1$ points, distinct two-by-two. For each $i \in [\![0, n]\!]$ we define

$$\ell_i(X) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{X - x_j}{x_i - x_j}.$$

The set $\mathcal{B} = \{\ell_i, i \in [\![0, n]\!]\}$ defines a basis for $\mathbb{K}_n[X]$. Since $\mathbb{K}_n[X]$ is an $(n+1)$-dimensional space and $\mathcal{B}$ is composed of $n+1$ vectors, $\mathcal{B}$ is a basis if it is linearly independent. Let $\lambda_0, \cdots, \lambda_n \in \mathbb{K}$ such that $\sum_{i=0}^{n} \lambda_i \ell_i = 0$. Noting that $\ell_i(x_j) = \delta_{ij}$, we get

$$0 = \left(\sum_{i=0}^{n} \lambda_i \ell_i \right)(x_j) = \sum_{i=0}^{n} \lambda_i \ell_i(x_j) = \lambda_j.$$

Now that we know that $\mathcal{B}$ is a basis for $\mathbb{K}_n[X]$ we want to determine its dual basis $\mathcal{B}^*$. As $\mathcal{B}$ is a basis, any $P \in \mathbb{K}_n[X]$ can be decomposed on it, i.e. there exist $\alpha_0, \cdots, \alpha_n \in \mathbb{K}$ such that $P = \sum_{i=0}^{n} \alpha_i \ell_i$.

In particular when evaluating P in x_j, for $0 \leq j \leq n$, we get $P(x_j) = \sum_{i=0}^{n} \alpha_i \ell_i(x_j) = \alpha_j$. Hence $P = \sum_{i=0}^{n} P(x_i)\ell_i$. Therefore, by theorem 3.165, the dual basis is composed of ℓ_i^*, $i \in [\![0, n]\!]$, that we need to determine.

For all $i \in [\![0, n]\!]$, $\ell_i^*(P) = \sum_{j=0}^{n} P(x_j)\ell_i^*(\ell_j) = P(x_i)$. Hence for all $i \in [\![0, n]\!]$, ℓ_i^* is simply the evaluation of P in x_i

$$\begin{aligned} \ell_i^* : \mathbb{K}_n[X] &\longrightarrow \mathbb{K} \\ P &\longmapsto P(x_i). \end{aligned}$$

Remark. Let V be a finite dimensional vector space, and $\beta(V)$ and $\beta(V^*)$ denote the sets of the bases of V and V^*, respectively. Based on theorem 3.165 we can define a bijection

$$\begin{aligned} d : \beta(V) &\longrightarrow \beta(V^*) \\ \mathcal{B} &\longmapsto \mathcal{B}^* \end{aligned}$$

which maps a basis to its dual.

Bijection between a basis and its dual

We start by showing the surjectivity of d. Since V is a finite dimensional space it admits at least one basis $\mathcal{B}_0$. Let $\mathcal{F}$ be a basis of V^*. Then we call Q the matrix of the change basis from $\mathcal{F}$ to $\mathcal{B}_0^*$, and define $P = (Q^{-1})^\top$. We can thus determine a basis $\mathcal{B}$ such that the matrix transforming $\mathcal{B}$ into $\mathcal{B}_0$ is P. Hence, by theorem 3.174 the matrix from $\mathcal{B}^*$ to $\mathcal{B}_0^*$ is $(P^{-1})^\top$, which is exactly Q (proposition 3.146). This shows that $\mathcal{F} = \mathcal{B}^* = d(\mathcal{B})$, and d is surjective.

To prove the injectivity we take two bases of V, $\mathcal{B}_1$ and $\mathcal{B}_2$ such that $\mathcal{B}_1^* = \mathcal{B}_2^*$. The matrix P from $\mathcal{B}_2$ to $\mathcal{B}_1$ verifies $(P^{-1})^\top = I_n$. Hence $P = I_n$ and $\mathcal{B}_1 = \mathcal{B}_2$.

Definition

Let $\mathcal{F}$ be a basis of V^*. The unique basis $\mathcal{B}$ such that $\mathcal{B}^* = \mathcal{F}$ is called the *predual basis* of $\mathcal{F}$. The bases $\mathcal{B}$ and $\mathcal{F}$ are said to be *dual of each others*.

Definitions

1. Let S be a subset of V. The *annihilator* of S in V^* is
$$S^\perp = \{\varphi \in V^*,\ \forall s \in S,\ \varphi(s) = 0\}.$$
2. Let L be a subset of V^*. The *annihilator* of L in V is
$${}^\circ L = \{x \in V,\ \forall \varphi \in L,\ \varphi(x) = 0\}.$$

Remarks.

- For any $\varphi \in V^*$ and any subset S of V, saying that $\varphi \in S^\perp$ is equivalent to saying that the restriction of φ to S is equal to 0, or that $S \subset \ker \varphi$.
- For any subset L of V^*, ${}^\circ L = \bigcap_{\varphi \in L} \ker \varphi$.

Annihilator and orthogonal complement

The notation of the annihilator is not without reminding us of the *orthogonal complement* $S^\perp$ of a subset S in V, defined by $S^\perp = \{x \in V : \forall v \in S, \langle v, x\rangle = 0\}$. This definition assumes V to be an inner product space, i.e. a vector space together with an inner product.[1] Interestingly the way to denote the inner product $\langle \cdot, \cdot \rangle$ is also similar to how we denoted our evaluation map $\langle x, \varphi \rangle$ in remark 3.164.

In fact this makes sense since the inner product is a symmetric, positive definite bilinear form. In particular, given $x \in V$, for any $v \in V$, $\langle v, x\rangle = \langle x, v\rangle \in \mathbb{R}$. Besides it can be shown that the linear form $\varphi : V \to V^*$ defined, for all $x \in V$ by $\varphi(v) = \langle v, x\rangle$, with $\langle \cdot, \cdot \rangle$ the inner product, is an isomorphism. Hence the orthogonal complement of S can be identified with its annihilator. More generally we can identify a vector space with its dual without even choosing a basis!

[1] An inner product can be seen as a generalization of the dot product. A more formal definition is given in homework 4, for the real numbers.

Dimension of orthogonal spaces

This highlights how notions related to the inner product can be rephrased in term of dual space. Using the dual space saves the choice of an inner product and the addition of an extra structure to the vector space.

Theorem

Let V be a finite n-dimensional vector space.

1. For any subspace V_0 of V, $\dim V_0^\perp = \dim V - \dim V_0$.
2. For any subspace L of V^*, $\dim {}^\circ L = \dim V - \dim L$.

Proof. Both demonstrations being similar we will only prove 1.

Let p be the dimension of V_0. By the incomplete basis theorem (2.93), we know that a basis $\mathcal{B}_0 = \{b_1, \cdots, b_p\}$ of V_0 can be completed into a basis $\mathcal{B} = \{b_1, \cdots, b_p, b_{p+1}, \cdots, b_n\}$ of V.

Let $\varphi \in V^*$. Then by proposition 3.167 we can write it $\varphi = \sum_{i=1}^n \varphi(b_i) f_i$, where $\{f_i, i \in [\![1, n]\!]\}$ is a basis of V^*.

Dimension of orthogonal spaces

As we want to determine dim $V_0^\perp$, we take $\varphi \in V_0^\perp$. Thus $\varphi \in (\text{span}\,\mathcal{B}_0)^\perp$, and by definition of orthogonality $\varphi(b_i) = 0$, for all $i \in [\![1, p]\!]$. Hence the only components left are along $f_{p+1}, \cdots, f_n$.

This shows that $\{f_{p+1}, \cdots, f_n\}$ spans $V_0^\perp$. However as this set is included in $\mathcal{B}^*$, it is linearly independent and as such is a basis for $V_0^\perp$. This yields the result

$$\dim V_0^\perp = n - p = \dim V - \dim V_0.$$

□

Corollary

Let V_0 be a subspace of V, and L be a subspace of V^*.

1. $^\circ(V_0^\perp) = V_0$;
2. $(^\circ L)^\perp = L$;
3. $^\circ(V^*) = \{0\}$;

Proof. Noting that for any $v \in V_0$ and $\varphi \in V_0^\perp$ we have $\varphi(v) = 0$, we see that $V_0 \subset {}^\circ(V_0^\perp)$. Then looking at the dimension we observe

$$\dim{}^\circ(V_0^\perp) = n - dim V_0^\perp = n - (n - \dim V_0) = \dim V_0.$$

The second point is proven similarly. For the third one remark that

$$\dim^\circ(V^*) = n - \dim V^* = n - \dim V = 0.$$

□

Corollary

Let V and W be two vector spaces and $f \in \mathcal{L}(V, W)$. Then $\ker f^\top = (\operatorname{im} f)^\perp$.

Proof. From remark 3.170 we know that $\langle u, f^\top(w^*)\rangle = \langle f(u), w^*\rangle$, for all $u \in V$ and $w^* \in W^*$. Hence $\ker f^\top \subset (\operatorname{im} f)^\perp$. Conversely, if we take $w^* \in (\operatorname{im} f)^\perp$, then $f^\top(w^*) \in V^\perp$, which by theorem 3.181 is $\{0\}$. □

Remark. A direct consequence of this corollary is that if f is injective then $f^\top$ is surjective and vice-versa.

Definition (Codimension)

Let V_0 be subspace of vector space V. If V_0 is in direct sum with at least one finite dimensional subspace V_1 of V, then it is said to be of *finite codimension*, and the *codimension* of V_0 is the dimension of V_1.

Remarks.

- There is no assumption on V or V_0 being finite dimensional spaces. In fact the codimension is especially useful when working with infinite dimensional spaces as it can allow to bring back a "concept" of finite dimension.
- When two subspaces V_0 and V_1 are in direct sum in V, we say that they are mutually *complementary* and we can speak of V_1 as the *complement* of V_0 in V.

Digression on the codimension

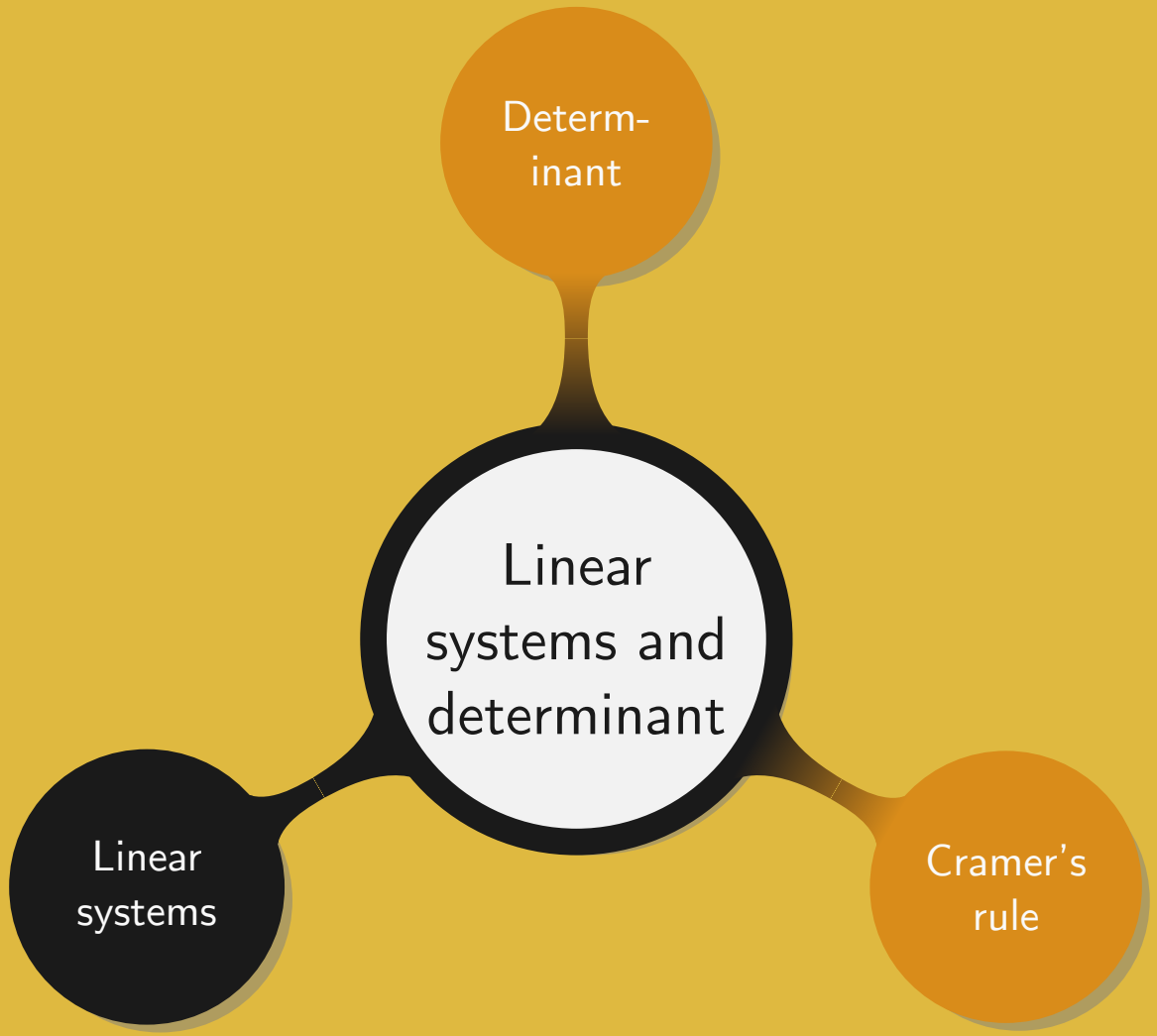
- All the complements of a subspace $V_0 \subset V$ have same dimension, if it is finite it is the codimension of V_0. This is clear as all the complements of V_0 are isomorphic, so if one has finite dimension, they all have.
- A consequence of theorem 2.101 is that $\dim V = \dim V_0 + \operatorname{codim} V_0$. Note that in this context V is a finite dimensional space.
- The rank theorem (2.109) assumes V to be a finite dimensional space but does not require anything on W. Now if W has finite dimension and we do not have this assumption on V, then $\ker u$ has finite codimension and the equality $\operatorname{rank} u = \operatorname{codim} \ker u$ still holds.
- Theorem 3.181 can be rephrased in term of codimension:
  1. For any subspace V_0 of V, $\dim V_0^{\perp} = \operatorname{codim} V_0$;
  2. For any subspace L of V^*, $\dim {}^{\circ}L = \operatorname{codim} L$;

Key points

- What is $\mathcal{M}_{n,p}(\mathbb{K})$ and what is its algebraic structure?
- What is the connection between linear maps and matrices?
- What is the rank of a matrix?
- How useful are block matrices, why?
- Explain the concept of duality

4. Linear systems and determinant

Chapter organisation



From advanced to elementary

As explained in the introduction of the course, linear equations appeared as an important incentive towards the study of linear algebra. In this chapter we want to relate our previous studies to linear equations and matrices.

In the previous chapter we investigated the algebraic structure of matrices using more advanced knowledge from algebra. In particular we said that $\mathcal{M}_{n,p}(\mathbb{K})$ was a vector space and that $\mathcal{M}_n(\mathbb{K})$ had a ring structure. Furthermore using an isomorphism we could transform a matrix representation in a basis into another representation in another basis. To do so we relied on "advanced" ring properties.

We now jump back in time and see how the same can be achieved by only applying elementary operations on the row of a matrix. We will prove the equivalence of the two approaches and see how this more basic idea can help in solving systems of linear equations.

Elementary row and column operations

We call *elementary row operation* on the rows of a matrix one of the following operations.

- Add a multiple of a row to another row;
- Multiply a row by a non-zero constant;
- Swap two rows;

For the sake of simplicity we will use the following notations to represent basic row operations.

- Add λR_j to row R_i: $R_i \leftarrow R_i + \lambda R_j$;
- Multiply row R_i by λ: $\mathbb{R}_i \leftarrow \lambda R_i$;
- Swap row R_i and R_j: $R_i \leftrightarrow R_j$;

Remarks.

- We can similarly define *elementary column operations*.
- In term of notation we simply replace the R by a C;

Elementary row and column operations

If a matrix A is *reduced* to a matrix B, using only elementary operations then it is possible to reduce B to A. For instance we have the following correspondence.

From A to B	From B to A
$R_i \leftarrow R_i + \lambda R_j$	$R_j \leftarrow R_i - \lambda R_j$
$R_i \leftarrow \lambda R_i$	$R_i \leftarrow \lambda^{-1} R_i$
$R_i \leftrightarrow R_j$	$R_i \leftrightarrow R_j$

Proposition

Let A and B be two matrices, such that A is reduced to B using a finite number of elementary operations. Then $\operatorname{rank} A = \operatorname{rank} B$.

Proof. Without any lose of generality we can assume a single elementary operations on the rows. Indeed it otherwise suffices to take the transpose to obtain an operation on the rows instead of the columns. Besides if after one operation the two matrices still have the same rank, applying more similar operations will keep it unchanged.

Let V and W be the subspaces spanned by the rows of A and B, respectively. Since B is reduced from A it means that all the rows of B can be written as linear combinations of rows of A, which shows that $W \subset V$. But as A can be reduced from B, then we also have $V \subset W$. This shows that $V = W$ and as such they have same dimension. □

In fact one can notice that an elementary row operation simply consists in the left multiplication by an invertible matrix, while an elementary column operation consists in the right multiplication by an invertible matrix.

Matrix product and elementary operations

We denote by $E_{i,j}$ the matrix whose coefficients are all zero but the one at position (i, j) which is 1, and proceed with finding a matrix equivalent for each elementary row operation.

Elementary operation	Matrix P
$R_i \leftarrow R_i + \lambda R_j$	$I_n + \lambda E_{i,j}$
$R_i \leftarrow \lambda R_i$	$I_n - E_{i,i} + \lambda E_{i,i}$
$R_i \leftrightarrow R_j$	$I_n - E_{i,i} - E_{j,j} + E_{i,j} + E_{j,i}$

By proposition 4.191 the matrix P has same rank as the identify matrix, and as such it must be invertible. The transpose leaving the rank unchanged the result is also valid for elementary column operations.

Proposition

Let $\alpha \in \mathbb{K}^*$ such that

$$\operatorname{rank} \left(\begin{array}{c|c} \alpha & * \cdots \cdots * \\ \hline 0 & \\ \vdots & A' \\ 0 & \end{array}\right) = 1 + \operatorname{rank} A'.$$

Proof. We denote by A the matrix from the proposition, and call W the space spanned by the rows $\{R_1, \cdots, R_n\}$. Since $\alpha \neq 0$, the intersection of the subspaces $\operatorname{span}\{R_1\}$ and $\operatorname{span}\{R_2, \cdots, R_n\}$ is $\{0\}$. Hence they are in direct sum, and as a result $\dim W = 1 + \dim \operatorname{span}\{R_2, \cdots, R_n\}$.

This shows the result since the subspace spanned by the columns of $\left(\begin{array}{c|c} 0 & \\ \vdots & A' \\ 0 & \end{array}\right)$ is the same as the one spanned by the columns of A'. □

Remark. This proposition is very important as it permits to determine the rank of any matrix. Indeed let $A \in \mathcal{M}_{n,p}(\mathbb{K})$.

- If $A = 0$, then its rank is evidently 0;
- If A features at least one non-zero coefficient we can move it in position $(1, 1)$ using elementary operations on the rows and columns of A, i.e. without altering the rank. Then it suffices to use an appropriate factor to multiply the first row and deduct it from all others, to ensure the first column only contains 0 for the last $n - 1$ rows.
- The process can be repeated on the submatrix A' of size $(n - 1) \times (p - 1)$ until obtaining the 0 matrix or any other matrix whose rank is easily calculated.

Example. Determine the rank of the matrix $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ a & b & c & d \end{pmatrix}$.

By applying the elementary row operations $R_2 \leftarrow R_2 - 2R_1$, $R_3 \leftarrow R_3 - 3R_1$, and $R_4 \leftarrow R_4 - aR_1$, we get

$$A' = \begin{pmatrix} -1 & -2 & -3 \\ -2 & -4 & -6 \\ b-2a & c-3a & d-4a \end{pmatrix}.$$

Thus $\operatorname{rank} A = 1 + \operatorname{rank} A'$, and we can repeat the same process this time with $R_2 \leftarrow R_2 + 2R_1$ and $R_3 \leftarrow R_3 - (b-2a)R_1$ to obtain $\operatorname{rank} A = 2 + \operatorname{rank} \begin{pmatrix} 0 & 0 \\ a+c-2b & 2a+d-3b \end{pmatrix}$. Hence if $a + c = 2b$ and $2a + d = 3b$, then $\operatorname{rank} A = 2$, and otherwise it is 3.

Row echelon form

Remark. In proposition 4.194, α is taken in $\mathbb{K}$. Starting from the matrix obtained using the method described in remark 4.195, extra elementary operations can be applied to ensure that the leading coefficient on each line is 1. We say that a matrix A is in *row echelon form* if it satisfies the following properties.

- A row that is not entirely composed of zeros, has its first non-zero entry in that row, equal to 1;
- Rows entirely composed of zeros, can only occur at the bottom of the matrix;
- In any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs further to the right than the leading 1 above;

If each column of A that contains a leading 1 has zeros everywhere else, then A is said to be in *reduced row echelon form*.

Theorem

Any matrix in $\mathrm{GL}_n(\mathbb{K})$ can be transformed into an invertible matrix from $\mathcal{T}_n(\mathbb{K})$ using elementary operations.

Proof. Let $A \in M_n(\mathbb{K})$. We will proceed by induction on n.

If $n = 1$, then A is already in $\mathcal{T}_n(\mathbb{K})$.

We now suppose the property to be true for $n-1$, and prove it for a matrix $A \in \mathrm{GL}_n(\mathbb{K})$. Since A is invertible its first column is non-zero, and we can suppose that $a_{1,1} \neq 0$, otherwise we apply elementary permutations on the rows to ensure it is the case. By applying the strategy described in remark 4.195 we can write A in the same form as in proposition 4.194, and by this result we have $n = \operatorname{rank} A = 1 + \operatorname{rank} A'$. In this context A' is invertible and A is now upper triangular, since A' is also upper triangular by our induction hypothesis.

By the induction principle the theorem is therefore correct. □

Remark. Following the strategy described in the proof provides a simple method to transform an invertible matrix into an upper triangular one. But it also proves that a matrix is invertible, since a triangular matrix without any 0 on the diagonal is invertible (remark 3.152).

Example. We want to know if $A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{pmatrix}$ is invertible.

Following Gauss elimination we successively obtain

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 5 & 9 \\ 0 & 3 & 9 & 19 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 3 & 10 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which shows that A is invertible.

Remark. If A is not invertible then either one of the submatrix A' will feature a 0 zero column or the last element on the diagonal will be 0.

Definition (Linear system)

A *linear system* with n equations and p unknowns $x_1, \cdots, x_p$ is a system of equations such that

$$\begin{cases} a_{1,1}x_1 + \cdots + a_{1,j}x_j + \cdots + a_{1,p}x_p = b_1 \\ \vdots \\ a_{i,1}x_1 + \cdots + a_{i,j}x_j + \cdots + a_{i,p}x_p = b_i \\ \vdots \\ a_{n,1}x_1 + \cdots + a_{n,j}x_j + \cdots + a_{n,p}x_p = b_n \end{cases},$$

with $(a_{i,j})_{\substack{i\in[\![1,n]\!] \\ j\in[\![1,p]\!]}}$ are in $\mathbb{K}$ and $b_1, \cdots, b_n \in \mathbb{K}$.

Commonly used notations:

- $A = (a_{i,j})_{\substack{i \in [\![1,n]\!] \\ j \in [\![1,p]\!]}}$ is called the *matrix of the system*;
- The *solution of the system* is a p-tuple $(x_1, \cdots, x_p)$ of $\mathbb{K}^p$ satisfying the n equations;
- The *rank of the system* corresponds to the rank of A;
- If all the b_i are 0, then the system is said to be *homogeneous*;
- A linear system with at least one solution it is said to be *compatible* or *consistent*;

Example. The system

$$\begin{cases} x_1 - x_2 = 1 \\ x_2 - x_3 = 1 \\ x_3 - x_1 = 1 \end{cases}$$

is not compatible in $\mathbb{R}$ or $\mathbb{C}$: if (x_1, x_2, x_3) was a solution, then by summation of the three equations we would have $0 = 3$.

Definition

Let V be a vector space. A subset $A \subset V$ is an *affine subspace* of V if there exists $\Omega \in V$ and a vector subspace V_0 such that $A = \Omega + V_0$. In that case we say that A is an affine space *directed* by V_0, and *passing through* Ω. The elements of A are called *points*.

If non-empty, the set of the solutions of a linear system S is an affine subspace of $\mathbb{K}^p$ directed by the vector subspace of the solutions of S_0 the homogeneous system associated to S. In particular this means that

- S_0 is a vector subspace of $\mathbb{K}^p$;
- if $(x_1, \cdots, x_p)$ and $(y_1, \cdots, y_p)$ are two solutions of S then their difference is also a solution for S;
- if $(x_1, \cdots, x_p)$ is a solution for S and $(y_1, \cdots, y_p)$ is a solution for S_0, then their sum is a solution for S;
- if S is compatible, its solutions are given by the addition of a particular solution of S to the solutions of S_0;

Matrix interpretation: if A is the matrix of S then we can set $X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$ and $B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ and describe the system as $AX = B$.

Conversely given $A \in \mathcal{M}_{n,p}(\mathbb{K})$ and $B \in \mathcal{M}_{n,1}(\mathbb{K})$, finding a matrix $X \in \mathcal{M}_{p,1}(\mathbb{K})$ such that $AX = B$ can be rephrased in term of linear system.

Based on this interpretation it is clear that if A is invertible, then there exists a unique solution to S, given by $X = A^{-1}B$.

Vectorial interpretation: let $C_1, \cdots, C_p$ be the column vectors of A and B be the vector $(b_1, \cdots, b_n) \in \mathbb{K}^n$. Then the system can be written $\sum_{j=1}^{p} x_j C_j = B$.

Conversely if $v_1, \cdots, v_p$ and b are $p+1$ vectors of an n-dimensional vector space, then finding p scalars $x_1, \cdots, x_p$ such that $\sum_{j=1}^{p} x_j v_j = b$, is equivalent to solving a linear system with n equations and p unknown expressed in a basis $\mathcal{B}$.

Based on this interpretation it is clear that

- S is compatible if and only if $B \in \text{span}\{C_1, \cdots, C_p\} \subset \mathbb{K}^n$;
- the rank of S is equal to the rank of $\{C_1, \cdots, C_p\}$;
- if $\{C_1, \cdots, C_p\}$ is linearly independent, then S has at most one solution;
- if $n = p$ and $\{C_1, \cdots, C_n\}$ is a basis for $\mathbb{K}^n$, then for any B the system has a unique solution corresponding to the components of B in the basis $\{C_1, \cdots, C_n\}$;

Linear map interpretation: let u be the linear map from $\mathbb{K}^p$ to $\mathbb{K}^n$, canonically associated to the matrix A of a system S. If $x = (x_1, \cdots, x_p)$ and $b = (b_1, \cdots, b_n)$ then we can write S as $u(x) = b$.

Conversely for two vector spaces V and W with dimension p and n, respectively, as well as a linear map from V to W and a vector $b \in W$, finding x such that $u(x) = b$ can be written as a linear system with n equations and p unknown.

Based on this interpretation it is clear that

- the system is compatible if and only if $b \in \operatorname{im} u$;
- if u is injective, then the system has at most one solution;
- if u is bijective, then the system has exactly one solution;
- if x_0 is a solution for S, then all its solution are of the form $x_0 + \ker u$;
- the rank of S is equal to the rank of u;
- if S is a homogeneous system with rank r, then the set of all its solutions is a $(p - r)$-dimensional vector subspace of $\mathbb{K}^p$;

Dual space interpretation: let $\varphi_1, \cdots, \varphi_n$, be n linear forms over $\mathbb{K}^n$ canonically associated to the rows of A. If $x = (x_1, \cdots, x_p)$ and $b = (b_1, \cdots, b_n)$, then S can be written as $\varphi_1(x) = b_1, \cdots, \varphi_n(x) = b_n$.

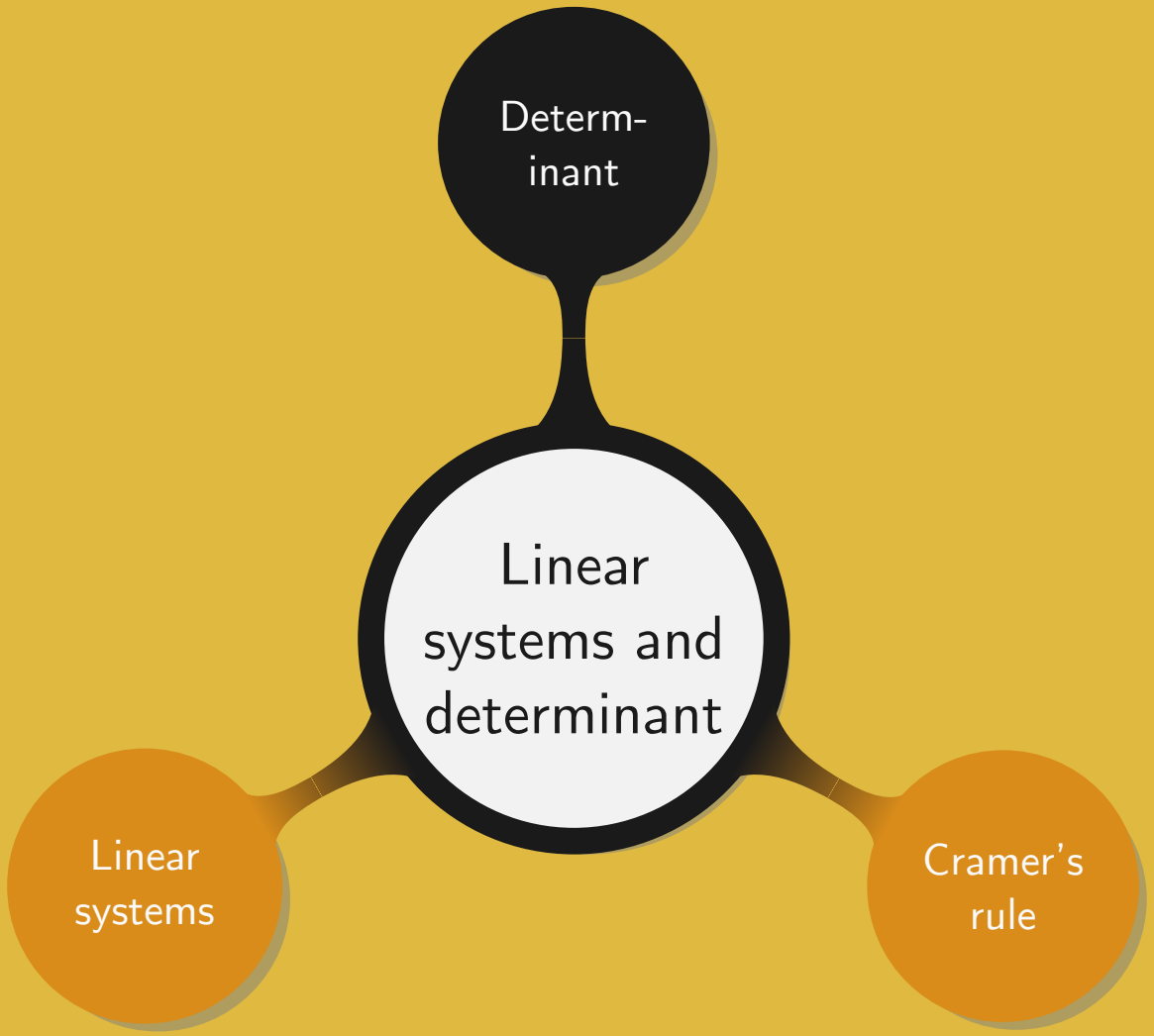
Conversely given a p-dimensional $\mathbb{K}$-vector space V, as well as n linear forms over V, finding x such that $\varphi_1(x) = b_1, \cdots, \varphi_n(x) = b_n$, is equivalent to solving a linear system with n equations and p unknowns.

Based on this interpretation it is clear that

- the set of the solutions of S_0 is the intersection of the kernels of the linear forms φ_i, i.e. of n vectorial hyperplanes;
- the set of the solutions of S is the intersection of n affine hyperplanes;
- the intersection of n vectorial hyperplanes is a $(p - n)$-dimensional vector subspace;

Remark. In fact linear forms, as elements of the dual space, generalize the notion of linear equations from $\mathbb{R}^n$ to arbitrary vector spaces. This is especially helpful when no canonical basis is available.

Chapter organisation



Definition (Symmetric group)

Let $\mathcal{S}_n$ be the set of the *permutations* from $[\![1, n]\!]$, i.e. the set of the bijection from $[\![1, n]\!]$ in itself. Endowed with the internal composition law of maps, $(\mathcal{S}_n, \circ)$ is called a *symmetric group*.

Example. For $i \neq j \in [\![1, n]\!]$ the map τ defined by $\tau(i) = j$, $\tau(j) = i$, and for any $a \notin \{i, j\}$, $\tau(x) = x$, is an involution, so a permutation. It is a called *transpose*, and is denoted (i, j).

Remarks.

- We denote (σ, τ) the composition of the permutations σ and τ and call it the product of σ and τ if written $\sigma\tau$.
- If $n = 1$, then $\mathcal{S}_1 = \{\mathrm{Id}\}$. In the rest of the chapter $n \geq 2$.
- If $n \geq 3$, then $\mathcal{S}_n$ is not commutative.

Proposition

Any permutation from $[\![1, n]\!]$ is a product of transposes.

Proof. We prove the result by induction on k using the hypothesis: any permutation of $[\![1, n]\!]$ which fixes $k, k+1, \cdots, n$ can be written as a product of transposes.

For $k = 1$ we simply need to find a product of transposes fixing all the elements, i.e. corresponding to the identity permutation. Taking the transpose twice works.

We assume our assumption to be true up to k and prove it for $k + 1$. Let σ be a permutation of $[\![1, n]\!]$ fixing $k + 1, \cdots, n$. If $\sigma(k) = k$ then is means k is also fixed and we we get the result by applying the recursion hypothesis.

Otherwise, since $k+1, \cdots, n$ are all fixed it means that k is permuted with an element strictly smaller than k. If we call that transpose τ_0 we see that $\sigma' = \tau_0\sigma$ leaves $k, k+1, \cdots, n$ unchanged. Applying our induction hypothesis to σ', we write it as a product of transposes $\tau_1 \cdots \tau_p$, for some integer p. Hence $\sigma = \tau_0\sigma' = \tau_0\tau_1 \cdots \tau_p$.

This proves our hypothesis at the rank $k+1$, and we can conclude by applying the induction principle. □

Definitions

Let $\sigma \in \mathcal{S}_n$.

1. A pair (i, j) of elements in $[\![1, n]\!]$ is an *inversion* of σ if $i < j$ and $\sigma(i) > \sigma(j)$. The number of inversions of σ is denoted $I(\sigma)$.
2. The *signature* of σ is the number $\varepsilon(\sigma) = (-1)^{I(\sigma)}$.
3. A permutation is said to be *even* if its signature is 1, and *odd* otherwise.

Proposition

The signature of a transpose is -1.

Proof. Let $\tau = (k, l)$ be a transpose with $k < l$. The inversions of τ are the pairs (k, i) and (i, l) with $k < i < l$, and (k, l). Thus $l(\sigma)$ is odd. □

Theorem

If σ and τ are two permutations of $[\![1, n]\!]$, then $\varepsilon(\sigma\tau) = \varepsilon(\sigma)\varepsilon(\tau)$, and ε is a group homomorphism from $\mathcal{S}_n$ to $(\{-1, 1\}, \times)$.

Proof. The set $A = \{(i, j) \in [\![1, n]\!], i < j\}$ can be split into four disjoint subsets.

We define n_1, n_2, n_3, and n_4 to be the cardinals of the subsets

$$A_1 = \{(i,j) \in A,\ \tau(i) < \tau(j) \text{ and } \sigma(\tau(i)) < \sigma(\tau(j))\},$$
$$A_2 = \{(i,j) \in A,\ \tau(i) < \tau(j) \text{ and } \sigma(\tau(i)) > \sigma(\tau(j))\},$$
$$A_3 = \{(i,j) \in A,\ \tau(i) > \tau(j) \text{ and } \sigma(\tau(i)) < \sigma(\tau(j))\},$$
$$A_4 = \{(i,j) \in A,\ \tau(i) > \tau(j) \text{ and } \sigma(\tau(i)) > \sigma(\tau(j))\}.$$

By construction, it is clear that the number of inversions $I(\tau)$ and $I(\sigma\tau)$ are $n_3 + n_4$ and $n_2 + n_4$, respectively.

The maps $f : A_2 \to A$, $f(i,j) = (\tau(i), \tau(j))$ and $g : A_3 \to A$, $g(i,j) = (\tau(j), \tau(i))$ are both injective and their images are given by

$$f(A_2) = \{(k,l) \in A,\ \tau^{-1}(k) < \tau^{-1}(l) \text{ and } \sigma(k) > \sigma(l)\}$$
$$g(A_3) = \{(k,l) \in A,\ \tau^{-1}(k) > \tau^{-1}(l) \text{ and } \sigma(k) > \sigma(l)\}.$$

Since those two sets are disjoint and their union is the set of the inversions of σ, we see that $I(\sigma) = n_2 + n_3$. Finally we have $I(\sigma\tau) - I(\sigma) - I(\tau) = (n_2 + n_4) - (n_2 + n_3) - (n_3 + n_4) = -2n_3$.

This proves that $I(\sigma\tau)$ and $I(\sigma) + I(\tau)$ have same parity, which allows us to conclude that $\varepsilon(\sigma\tau) = \varepsilon(\sigma)\varepsilon(\tau)$. □

Corollary

If a permutation σ is written as a product of transposes $\sigma = \tau_1 \cdots \tau_p$, then $\varepsilon(\sigma) = (-1)^p$.

Remark. The decomposition of σ into a product of transposes is not unique. However its parity is an invariant.

Definition (Alternating group)

The subgroup of $\mathcal{S}_n$ containing all the even permutations is called *alternating group* and denoted $\mathcal{A}_n$.

Example. We consider $\mathcal{S}_3 = \{\mathrm{Id}, (1,2), (1,3), (2,3), (1,2,3), (3,2,1)\}$. The 3-cycle $(1,2,3)$ can be written as the product of the transposes $(1,2)(2,3)$. Similarly $(3,2,1)$ can be written as the product of two transposes. Hence $\mathcal{A}_3 = \{\mathrm{Id}, (1,2,3), (3,2,1)\}$.

Proposition

Let τ be an odd permutation. Then $\mathcal{A}_n\tau = \{\sigma\tau,\ \sigma \in \mathcal{A}_n\}$ is the set of the odd permutations.

Proof. If $\sigma \in \mathcal{A}_n$, then $\varepsilon(\sigma\tau) = \varepsilon(\sigma)\varepsilon(\tau) = -1$, i.e. $\sigma\tau$ is odd. Conversely if σ is odd then $\sigma\tau^{-1}$ is even and $(\sigma\tau^{-1})\tau \in \mathcal{A}_n\tau$. □

Definition (multi-linear map)

Let $V_1, \cdots V_p$ and W be some vector spaces. A map $f : V_1 \times \cdots \times V_p \to W$ is said to be *p-linear* if for $1 \leq j \leq p$, and for any set of vectors $\{v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_p\}$ with $v_i \in V_i$, the map

$$V_j \to W$$
$$x_j \mapsto f(v_1, \cdots, v_{j-1}, x_j, v_{j+1}, \cdots v_p)$$

is linear. When $W = \mathbb{K}$, f is called a *p-linear form*.

Examples.

- The set of the p-linear maps from $V_1 \times \cdots \times V_p$ in W is a subspace of $\mathcal{F}(V_1, \cdots, V_p, W)$. We denote it $\mathcal{L}_p(V_1, \cdots, V_p, W)$. If all the V_i are equal we write $\mathcal{L}_p(V, W)$.
- The map $f : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, $f((x_1, x_2), (y_1, y_2)) = x_1 y_1 + x_2 y_2$ is a 2-linear form, also called bilinear form.

Definitions

Let $f \in \mathcal{L}_p(V, W)$ and $v = (v_1, \cdots, v_p) \in V^p$. For $\sigma \in \mathcal{S}_p$ we write $f^\sigma(v)$ to denote $f(v_{\sigma(1)}, \cdots, v_{\sigma(p)})$. We say that

- f is *antisymmetric* if $f^\sigma = \varepsilon(\sigma) f$;
- f is an *alternating multi-linear map* if for any v with $v_i = v_j$, $i \neq j$, we have $f(v) = 0$;

Theorem

Let f be a multi-linear map. Then f is antisymmetric if and only for any transpose τ, $f^\tau = -f$. Moreover any alternating multi-linear map is antisymmetric, and if the characteristic of $\mathbb{K}$ is not 2, then the converse is also true.

Proof. If only one transpose is applied then clearly by corollary 4.213 we will have $f^\tau = -f$. Conversely if for any transpose $f^\tau = -f$, then we get $f^\sigma = \varepsilon(\sigma)f$, since any permutation σ can be written as a product of transposes (proposition 4.209).

We now want to show that any alternating multi-linear map is antisymmetric. Let $j, k \in [\![1, p]\!]$, such that $j < k$, and $x_1, \cdots, x_p \in V$. We define the map

$$\begin{aligned} g : V^2 &\longrightarrow W \\ (y, z) &\longmapsto f(x_1, \cdots, x_{j-1}, y, x_{j+1}, \cdots, x_{k-1}, z, x_{k+1}, \cdots, x_p). \end{aligned}$$

By the p-linearity of f we have

$$g(x_j + x_k, x_j + x_k) = g(x_j, x_j) + g(x_j, x_k) + g(x_k, x_j) + g(x_k, x_k).$$

Since f is an alternating map, so is g. Thus $g(x_j, x_j)$, $g(x_k, x_k)$ and $g(x_j, x_k) + g(x_k, x_j)$ are all 0. Calling τ the transpose (j, k), $f^\tau = -f$, which based on what we proved earlier means that f is antisymmetric.

If now f is antisymmetric then $f^\tau = -f$. Then keeping the previous notations and choosing x_k such that $x_k = x_j$ we get $2f(x_1, \cdots, x_p) = 0$. Hence if we are not in characteristic 2, i.e. 2 is not a zero divisor, then $f(x_1, \cdots, x_p) = 0$ and f is an alternating multi-linear map. □

Theorem

Let f be an alternating p-linear map from V^p to W. If $S = \{v_1, \cdots, v_p\}$ is a linearly dependent set of vectors from V then $f(v_1, \cdots, v_p) = 0$.

Proof. As S is linearly dependent there exists i such that $v_i = \sum_{k \neq i} \lambda_k v_k$. Thus

$$f(v_1, \cdots, v_p) = f(v_1, \cdots, v_{i-1}, \sum_{k \neq i} \lambda_k v_k, v_{i+1}, \cdots, v_p)$$

$$= \sum_{k \neq i} \lambda_k f(u_1, \cdots, v_{i-1}, v_k, v_{i+1}, \cdots, v_p).$$

□

This sum is null since f is an alternating map and as such each of the term is 0. □

Corollary

Let f be an alternating p-linear map from V^p to W, and $v = (v_1, \cdots, v_p) \in V^p$. Then $f(v)$ remains unchanged if a linear combination of the v_j is added to one of them.

Proposition

Let $f \in \mathcal{L}_p(V, W)$. Then $A(f) = \sum_{\sigma \in \mathcal{S}_p} \varepsilon(\sigma) f^\sigma$ is an alternating p-linear map from V^p to W.

Proof. Let $j \neq k \in [\![1, p]\!]$, and τ be the transpose (j, k). Using proposition 4.214 we see that there are as many even and odd permutations. Hence there exist a bijection $\sigma \to \tau\sigma$, between $\mathcal{A}_n$ and $\mathcal{S}_n \setminus \mathcal{A}_n$.

As a result we can restrict the sum $A(f)$ to even permutations and account for the odd ones using the bijection $\sigma \to \tau\sigma$. If σ is even then $\varepsilon(\sigma)f^\sigma = f^\sigma$, and otherwise $\varepsilon(\sigma)f^\sigma = -f^{\tau\sigma'}$, for some even σ'. Thus,

$$\sum_{\sigma\in\mathcal{S}_p} \varepsilon(\sigma)f^\sigma = \sum_{\sigma\in\mathcal{A}_p} (f^\sigma - f^{\tau\sigma}).$$

Let $x = (x_1, \cdots, x_p) \in V^p$ such that $x_j = x_k$. Then for $\sigma \in \mathcal{A}_p$ we set $u_\sigma = f^\sigma(x) - f^{\tau\sigma}(x)$. With these notations $A(f)(x)$ corresponds to the sum of the u_σ.

Recalling that $\tau = (j, k)$, observe that $u_\sigma = 0$. Indeed,

- if $\sigma(i) \in \{j, k\}$, then $x_{\sigma(i)} = x_{\tau\sigma(i)} = x_j = x_k$;
- if $\sigma(i) \notin \{j, k\}$, then $\sigma(i) = (\tau\sigma)(i)$;

This shows that $A(f)$, which is the sum of all the u_σ, is an alternating map.

□

Let V be a finite n-dimensional $\mathbb{K}$-vector space and $\mathcal{B} = \{b_1, \cdots, b_n\}$ be one of its bases. Denoting its dual basis $\mathcal{B}^* = \{f_1, \cdots, f_n\}$, we define f as the n-linear form

$$f : V^n \longrightarrow \mathbb{K}$$

$$(x_1, \cdots, x_n) \longmapsto \prod_{j=1}^{n} \langle x_j, f_j \rangle.$$

We apply $A(\cdot)$, as expressed in proposition 4.219, to f and call $A(f)$ the *determinant* in $\mathcal{B}$. It is denoted $\det_{\mathcal{B}}$, i.e.

$$A(f)(x_1, \cdots, x_n) = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} \langle x_j, f_j \rangle = \det_{\mathcal{B}}(x_1, \cdots, x_n).$$

Note that each of the x_i is a vector in V and as such features n components. When evaluating $\det_{\mathcal{B}}$ on $\mathcal{B}$, using theorem 3.165, we get

$$\det_{\mathcal{B}}(b_1, \cdots, b_n) = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} \delta_{j,\sigma(j)} = 1. \tag{4.1}$$

Equation (4.1) is clear as soon as one notices that Id is the only permutation not yielding 0 and $\varepsilon(\mathrm{Id}) = 1$. For all other permutations at least one of the $\delta_{j,\sigma(j)}$ is 0, thus cancelling the whole product.

We want to prove that the determinant is the unique multi-linear form meeting equation (4.1). In order to better understand what is happening we first consider the case of a map $g \in \mathcal{A}_2(V)$, with V a 2-dimensional vector space. Defining $\lambda_{ij} = \langle x_j, f_i \rangle$, with $1 \leq i, j \leq 2$ we can write

$$\begin{aligned} g(x_1, x_2) &= g(\lambda_{1,1} b_1 + \lambda_{1,2} b_2, \lambda_{2,1} b_1 + \lambda_{2,2} b_2) \\ &= \lambda_{1,1} g(b_1, \lambda_{2,1} b_1 + \lambda_{2,2} b_2) + \lambda_{1,2} g(b_2, \lambda_{2,1} b_1 + \lambda_{2,2} b_2) \\ &= \lambda_{1,1}\lambda_{2,1} g(b_1, b_1) + \lambda_{1,1}\lambda_{2,2} g(b_1, b_2) \\ &+ \lambda_{1,2}\lambda_{2,1} g(b_2, b_1) + \lambda_{2,1}\lambda_{2,2} g(b_2, b_2). \end{aligned}$$

However as g is an alternating multi-linear map, $g(b_1, b_1) = g(b_2, b_2) = 0$, which leads to

$$g(x_1, x_2) = \lambda_{1,1}\lambda_{2,2} g(b_1, b_2) + \lambda_{1,2}\lambda_{2,1} g(b_2, b_1).$$

Now observe that g is only applied to permutations of b_1 and b_2, since $\mathcal{S}_2 = \{\mathrm{Id}, (1,2)\}$. Thus we can further rewrite

$$g(x_1, x_2) = \sum_{\sigma \in \mathcal{S}_2} \lambda_{\sigma(1),1} \lambda_{\sigma(2),2} g(b_{\sigma(1)}, b_{\sigma(2)}).$$

Looking back at definition 4.216, we have $g(b_{\sigma(1)}, b_{\sigma(2)}) = \varepsilon(\sigma) g(b_1, b_2)$. Hence we can factorize $g(b_1, b_2)$ out to obtain

$$g(x_1, x_2) = g(b_1, b_2) \sum_{\sigma \in \mathcal{S}_2} \varepsilon(\sigma) \lambda_{\sigma(1),1} \lambda_{\sigma(2),2}.$$

Since the λ_{ij} were defined as $\langle x_j, f_i \rangle$, it becomes

$$g(x_1, x_2) = g(b_1, b_2) \sum_{\sigma \in \mathcal{S}_2} \varepsilon(\sigma) \prod_{j=1}^{2} \langle x_j, f_j \rangle = g(b_1, b_2) \det_{\mathcal{B}}(x_1, x_2).$$

Hence $g = \det_{\mathcal{B}}$ if and only if $g(b_1, b_2) = 1$. In other words, for $n = 2$ the determinant is the unique bi-linear form meeting equation (4.1).

Following the same strategy we consider the general case of $g \in \mathcal{A}_n(V)$ with V an n-dimensional vector space. Generalizing λ_{ij} to $\langle x_j, f_i \rangle$, with $1 \leq i, j \leq n$, we quickly obtain

$$g(x_1, \cdots, x_n) = \sum_{\sigma \in \mathcal{S}_n} \lambda_{\sigma(1),1} \cdots \lambda_{\sigma(n),n} g(b_{\sigma(1)}, \cdots, b_{\sigma(n)}). \tag{4.2}$$

Similar to the $n = 2$ case, we use definition 4.216 to factorize $g(b_1, \cdots, b_n)$ out and make $\varepsilon(\sigma)$ appear:

$$\begin{aligned} g(x_1, \cdots, x_n) &= g(b_1, \cdots, b_n) \sum_{\sigma \in \mathcal{S}_m} \varepsilon(\sigma) \lambda_{\sigma(1),1} \cdots \lambda_{\sigma(n),n} \\ &= g(b_1, \cdots, b_n) \sum_{\sigma \in \mathcal{S}_m} \varepsilon(\sigma) \prod_{j=1}^{n} \langle x_j, f_j \rangle \\ &= g(b_1, \cdots, b_n) \det_{\mathcal{B}}(x_1, \cdots, x_n). \end{aligned}$$

Hence $g = \det_{\mathcal{B}}$ is the only n-linear form such that $g(b_1, \cdots, b_n) = 1$.

Remark. In the previous discussion we considered an n-linear map over an n-dimensional space. So a simple question is to know what happens for a p-linear map over an n-dimensional space when $n \neq p$. First note that for $g \in \mathcal{A}_p(V)$, equation (4.2) becomes

$$g(x_1, \cdots, x_p) = \sum_{\sigma \in \mathcal{S}_p} \lambda_{\sigma(1),1} \cdots \lambda_{\sigma(p),p} g(b_{\sigma(1)}, \cdots, b_{\sigma(p)}). \tag{4.3}$$

As each of the vectors x_i, $1 \leq i \leq p$, splits over the basis of the b_j, $1 \leq j \leq n$, two cases can arise:

- $p > n$. Equation (4.3) does not have "much sense" as some of the σ will map "beyond" n, i.e. to non-existing b_j. We could "patch it up" by allowing σ not to be a permutation, i.e. it could map two elements from $[\![1, p]\!]$ to a same element in $[\![1, n]\!]$; However in that case there would exist k and l such that $b_{\sigma(k)} = b_{\sigma(l)}$, and as a result g would be 0, since it is an alternating map.

- $p < n$. In that case a problem occurs when we want to factorize out $g(b_1, \cdots, b_n)$. Indeed σ will map $[\![1, p]\!]$ to a subset of $[\![1, n]\!]$, or said otherwise some b_j will be "missing", so we will not be able to "extract" a common term. Adding null b_j to "patch it up" would yield $g = 0$.

Theorem

Let V be a finite n-dimensional vector space with basis $\mathcal{B} = \{b_1, \cdots, b_n\}$ and $\mathcal{B}^* = \{f_1, \cdots, f_n\}$ be a dual basis of V^*.

1. The unique alternating n-linear form g such that $g(b_1, \cdots, b_n) = 1$ is $\det_{\mathcal{B}}$. And for any $g \in \mathcal{A}_n$, $g = g(b_1, \cdots, b_n) \det_{\mathcal{B}}$.
2. Let $x = (x_1, \cdots, x_n) \in V^n$. The determinant of x in the basis $\{b_1, \cdots, b_n\}$, noted $\det_{\mathcal{B}}(x) = \det_{\mathcal{B}}(x_1, \cdots, x_n)$ is given by

$$\det_{\mathcal{B}}(x) = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} \langle x_{\sigma(j)}, f_j \rangle = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} \langle x_j, f_{\sigma(j)} \rangle.$$

Proof. Results have been proven by construction starting on slide 4.221. □

Remark. Saying that for any $g \in \mathcal{A}_n$, $g = g(b_1, \cdots, b_n) \det_{\mathcal{B}}$, means that $\det_{\mathcal{B}}$ is a basis for $\mathcal{A}_n(V)$. As $\det_{\mathcal{B}}$ is a linear form it means that $\dim \mathcal{A}_n(V) = 1$.

Proposition

Let V be a finite dimensional $\mathbb{K}$-vector space, and $\mathcal{B}$ be a basis on V. A subset $S \subset V$ is a basis for V if and only if $\det_{\mathcal{B}} S$ is invertible.

Proof. ($\Rightarrow$) From theorem 4.226 (1) we have both $\det_S = \lambda \det_{\mathcal{B}}$ and $\det_{\mathcal{B}} = \mu \det_S$, for $\lambda = \det_S \mathcal{B}$ and $\mu = \det_{\mathcal{B}} S$. Thus $1 = \det_{\mathcal{B}} \mathcal{B} = \lambda\mu \det_{\mathcal{B}} \mathcal{B} = \lambda\mu$. Hence μ is invertible.

($\Leftarrow$) If S is a linearly dependent then at least one vector can be written as a linear combination of others. Then by theorem 4.218, $\det_{\mathcal{B}} S = 0$. □

Remark. As highlighted in chapter 1, $\mathbb{K}$ can be any field for most results. Proposition 4.227 is phrased such that it remains true even over finite fields. However the more common statement found in most textbooks says that "S is a basis if and only if $\det_{\mathcal{B}} S \neq 0$". In that case the result is only valid on fields of characteristic 0 such as $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Q}$.

Theorem

Let V be a finite n-dimensional vector space and $u \in \mathcal{L}(V)$. Then there exists a unique constant λ such that for any $f \in \mathcal{A}_n(V)$ and any $x_1, \cdots, x_n \in V$,

$$f(u(x_1), \cdots, u(x_n)) = \lambda \cdot f(x_1, \cdots, x_n).$$

Proof. We define the endomorphism Φ on $\mathcal{A}_n(V)$ by setting for any $f \in \mathcal{A}_n(V)$, $\Phi(f) = f(u(x_1), \cdots u(x_n))$. From remark 4.227, $\dim \mathcal{A}_n(V) = 1$ and there exists $\lambda \in \mathbb{K}$ such that for any $f \in \mathcal{A}_n(V)$, $\Phi(f) = \lambda f$.

For a basis $\mathcal{B}$ of V, we have $\lambda = \lambda \det_{\mathcal{B}} \mathcal{B} = \det_{\mathcal{B}} u(\mathcal{B})$. In particular this proves the unicity of λ. □

Definition

Keeping the notations from the previous theorem, the constant λ is called the *determinant* of u and denoted $\det u$.

Remark. For any basis $\mathcal{B} = \{b_1, \cdots, b_n\}$ of V,

$$\det u = \det_{\mathcal{B}}(u(b_1,), \cdots, u(b_n)).$$

Proposition

Let V be a finite n-dimensional vector space.

1. We have $\det \mathrm{Id} = 1$.
2. For $\lambda \in \mathbb{K}$ and $u \in \mathcal{L}(V)$, $\det(\lambda u) = \lambda^n \det u$.
3. For $u, v \in \mathcal{L}(V)$, $\det(v \circ u) = \det v \det u$.

Proof. (1) and (2) directly follow from remark 4.229.

(3) Let $\mathcal{B}$ be a basis of V and $x = (x_1, \cdots, x_n)$, with $x_j = u(b_j)$. From remark 4.229 we see that $\det(v \circ u) = \det_{\mathcal{B}} v(u(\mathcal{B})) = \det_{\mathcal{B}} v(x) = \det v \det_{\mathcal{B}} x = \det v \det u$. □

Proposition

Let V be a finite dimensional $\mathbb{K}$-vector space and $u \in \mathcal{L}(V)$. Then $u \in \mathrm{GL}(V)$ if and only if $\det u$ is invertible. If this is the case, then $\det u^{-1} = (\det u)^{-1}$.

Proof. ($\Rightarrow$) From proposition 4.229, we have $1 = \det \mathrm{Id} = \det(u \circ u^{-1}) = \det u \det u^{-1}$, i.e. $\det u$ is invertible.

($\Leftarrow$) Let $\mathcal{B}$ be a basis for V. Based on remark 4.229, we have $\det u = \det_{\mathcal{B}}(u(b_1) \cdots, u(b_n))$. As $\det u$ is invertible, proposition 4.227 tells us that $\{u(b_1), \cdots, u(b_n)\}$ is a basis. Hence $u \in \mathrm{GL}(V)$. □

Definition

Let $A \in \mathcal{M}_n(\mathbb{K})$. The *determinant* of A, denoted $\det A$, is the determinant of the column vectors of A in the canonical basis.

Remark. For $A = (a_{i,j})_{1 \leq i,j \leq n}$, $\det A$ is often denoted

$$\det A = \begin{vmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,n} \\ \vdots & & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j} & \cdots & a_{n,n} \end{vmatrix}.$$

Proposition

Let $A \in \mathcal{M}_n(\mathbb{K})$. Then $\det A = \det A^\top$.

Proof. Since $a_{i,j}$ is the ith coordinate of the jth column vector in the canonical basis of V we have from theorem 4.226 (2),

$$\det A = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} a_{j,\sigma(j)} = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}. \tag{4.4}$$

□

Theorem

Let V be a finite n-dimensional vector space and $\mathcal{B}$ be a basis for V. Let $u \in \mathcal{L}(V)$ and $A \in \mathcal{M}_n(\mathbb{K})$ be the matrix of u in $\mathcal{B}$. Then $\det u = \det A$.

Proof. This is straight forward since $u(b_j) = \sum_{i=1}^{n} a_{i,j} b_i$ and $\det u = \det_{\mathcal{B}} u(\mathcal{B})$.

□

Remark. Let $A, B \in \mathcal{M}_n(\mathbb{K})$ and $\lambda \in \mathbb{K}$. From the previous discussion and results we have indirectly obtained the following extra results:

- If A is transformed into a matrix C using a permutation σ, then $\det C = \varepsilon(\sigma) \det A$;
- The determinant of A linearly depends on each of its columns;
- The determinant of a matrix does not change if a linear combination of other columns is added to a column;
- The determinant is equal to 0 if a column is a linear combination of other columns;
- Since $\det A = \det A^{\perp}$, all the determinant properties phrased in term of columns are also true for the rows;
- We have $\det AB = \det A \det B$, and $\det \lambda A = \lambda^n \det A$;
- If A is invertible then $\det A^{-1} = (\det A)^{-1}$;
- If $n = 1$, $A = (\lambda)$ and $\det A = \lambda$;

Example. Given an orthonormal basis we want to classify conic sections. We know that a conic section is given by the equation

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0 \tag{4.5}$$

If we define $M = \left(\begin{smallmatrix} A & B \\ B & C \end{smallmatrix}\right)$, $L = \left(\begin{smallmatrix} D & E \end{smallmatrix}\right)$, and $X = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$, then equation (4.5) can be rewritten as

$$X^\top MX + 2LX + F = 0. \tag{4.6}$$

We now apply the change of basis

$$X = TX_1 + K, \text{ with } T = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \text{ and } K = \begin{pmatrix} \lambda \\ \mu \end{pmatrix}.$$

When plugging it into equation (4.6), it yields

$$X_1^\top M_1 X_1 + L_1 X_1 + F_1, \text{ with } M_1 = T^\top MT \text{ and } L_1, F_1 \in \mathcal{M}_2(\mathbb{K}).$$

We thus have $\det M = AC - B^2$ and $\det M_1 = \det M(\det T)^2$. This means that the sign of $\det M$, also called the *discriminant* of a conic section, is independent of the chosen orthonormal basis. In other words, regardless of the basis the type of a conic section can be simply deducted from its equation. More precisely

- if $\det M > 0$ then it is an ellipsis;
- if $\det M < 0$ then it is a hyperbola;
- if $\det M = 0$ then it is a parabola;

Example. Following the formula given in equation (4.4) we want to determine $\det M$ for

$$M = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}.$$

Calculation of a determinant

First since this is a 3×3 matrix we consider the permutations of $\mathcal{S}_3 = \{\mathrm{Id}, (1,2), (1,3), (2,3), (1,2,3), (3,2,1)\}$ (example 4.214). Then applying formula (4.4) we get

$$\begin{aligned}\det M = {} & a_{1,1}a_{2,2}a_{3,3} - a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,2}a_{3,1} \\ & - a_{1,1}a_{2,3}a_{3,2} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2}.\end{aligned}$$

We can factorize the determinant into

$$\begin{aligned}\det M = {} & a_{1,1}(a_{2,2}a_{3,3} - a_{2,3}a_{3,2}) - a_{1,2}(a_{2,1}a_{3,3} - a_{2,3}a_{3,1}) \\ & + a_{1,3}(a_{2,1}a_{3,2} - a_{2,2}a_{3,1}).\end{aligned}$$

We now want to generalize this formula such that we do not need to always list all the permutations of $\mathcal{S}_n$ when we desire to calculate the determinant of a matrix.

Proposition

Let $A \in \mathcal{M}_n(\mathbb{K})$ be such that

$$A = \left(\begin{array}{c|c} A' & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline * \cdots \cdots * & a_{n,n} \end{array}\right).$$

Then $\det A = a_{n,n} \det A'$.

Proof. From formula (4.4), $\det A = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}$. But as the last column is all zeros but $a_{n,n}$ it means that for any $\sigma \in \mathcal{S}_n$ with $\sigma(n) \neq n$, we have $a_{\sigma(n),n} = 0$. Hence formula (4.4) becomes

$$\det A = a_{n,n} \sum_{\substack{\sigma \in \mathcal{S}_n \\ \sigma(n)=n}} \varepsilon(\sigma) \prod_{j=1}^{n-1} a_{\sigma(j),j}.$$

In order to complete the proof we need to verify that the sum can be written over the permutations of $\mathcal{S}_{n-1}$. Clearly $\{\sigma \in \mathcal{S}_n, \sigma(n) = n\} \subset \{\sigma \in \mathcal{S}_{n-1}\}$. For the other inclusion observe that if we extent all permutations $\sigma \in \mathcal{S}_{n-1}$ by setting $\sigma(n) = n$, then we obtain a permutation of $\sigma' \in \mathcal{S}_n$. Besides $\varepsilon(\sigma') = \varepsilon(\sigma)$ since they match up to $n-1$, and for n we do not have any inversion. This shows the second inclusion. Hence we have shown that

$$\det A = a_{n,n} \sum_{\sigma \in \mathcal{S}_{n-1}} \varepsilon(\sigma) \prod_{j=1}^{n-1} a_{\sigma(j),j} = a_{n,n} \det A'.$$

□

Corollary

The determinant of a triangular matrix is given by the product of the terms on its diagonal.

Proof. From proposition 4.231 we only need to consider the lower triangular case which is a trivial extension of proposition 4.237. □

Definitions

Let $A = (a_{i,j})_{1 \leq 1, j \leq m} \in \mathcal{M}_n(\mathbb{K})$.

1. The determinant $\Delta_{i,j}$ of the matrix formed by deleting the ith row and jth column from A is called the *minor* of $a_{i,j}$.
2. The *cofactor* of $a_{i,j}$ is equal to $(-1)^{i+j}\Delta_{i,j}$.

Theorem

If $A = (a_{i,j})_{1 \leq 1, j \leq n} \in \mathcal{M}_n(\mathbb{K})$, then for any $j \in [\![1, n]\!]$,

$$\det A = \sum_{i=1}^{n} a_{i,j}(-1)^{i+j}\Delta_{i,j}.$$

Proof. Let $\mathcal{B}$ be the canonical basis of V and $C_1, \cdots, C_n$ be the columns of A. Then for any $j \in [\![1, n]\!]$, $C_j = \sum_{i=1}^{n} a_{i,j} b_i$, and by using the multi-linearity of the determinant we get

$$\det A = \det_{\mathcal{B}} \left(C_1, \cdots, C_{j-1}, \sum_{i=1}^{n} a_{i,j} b_i, C_{j+1}, \cdots, C_n \right)$$

$$= \sum_{i=1}^{n} a_{i,j} \underbrace{\det_{\mathcal{B}} \left(C_1, \cdots, C_{j-1}, b_i, C_{j+1}, \cdots, C_n \right)}_{D_{i,j}}$$

By construction $D_{i,j}$ is the determinant of the matrix with 0 all along the jth column but in ith row which features a 1. In order to match the setup of proposition 4.237 we need to move the 1 from position (i, j) to position (n, n). This can be easily performed using $n - i$ row and $n - j$ column transposes.

Development following a column

Knowing that each transpose has signature -1, it means that the determinant will be affected by a factor $(-1)^{2n-(i+j)} = (-1)^{i+j}$.
Hence $D_{i,j} = (-1)^{i+j}\Delta_{i,j}$ and $\det A = \sum_{i=1}^{n} a_{i,j}(-1)^{i+j}\Delta_{i,j}$. □

Corollary

If $A = (a_{i,j})_{1\leq 1, j\leq n} \in \mathcal{M}_n(\mathbb{K})$, then for any $i \in [\![1, n]\!]$,

$$\det A = \sum_{j=1}^{n} a_{i,j}(-1)^{i+j}\Delta_{i,j}.$$

Remark. This last result simply translates the fact that the determinant can be developed not only with respect to the columns, but also with respect to the rows.

Determinant of a bloc triangular matrix

Proposition

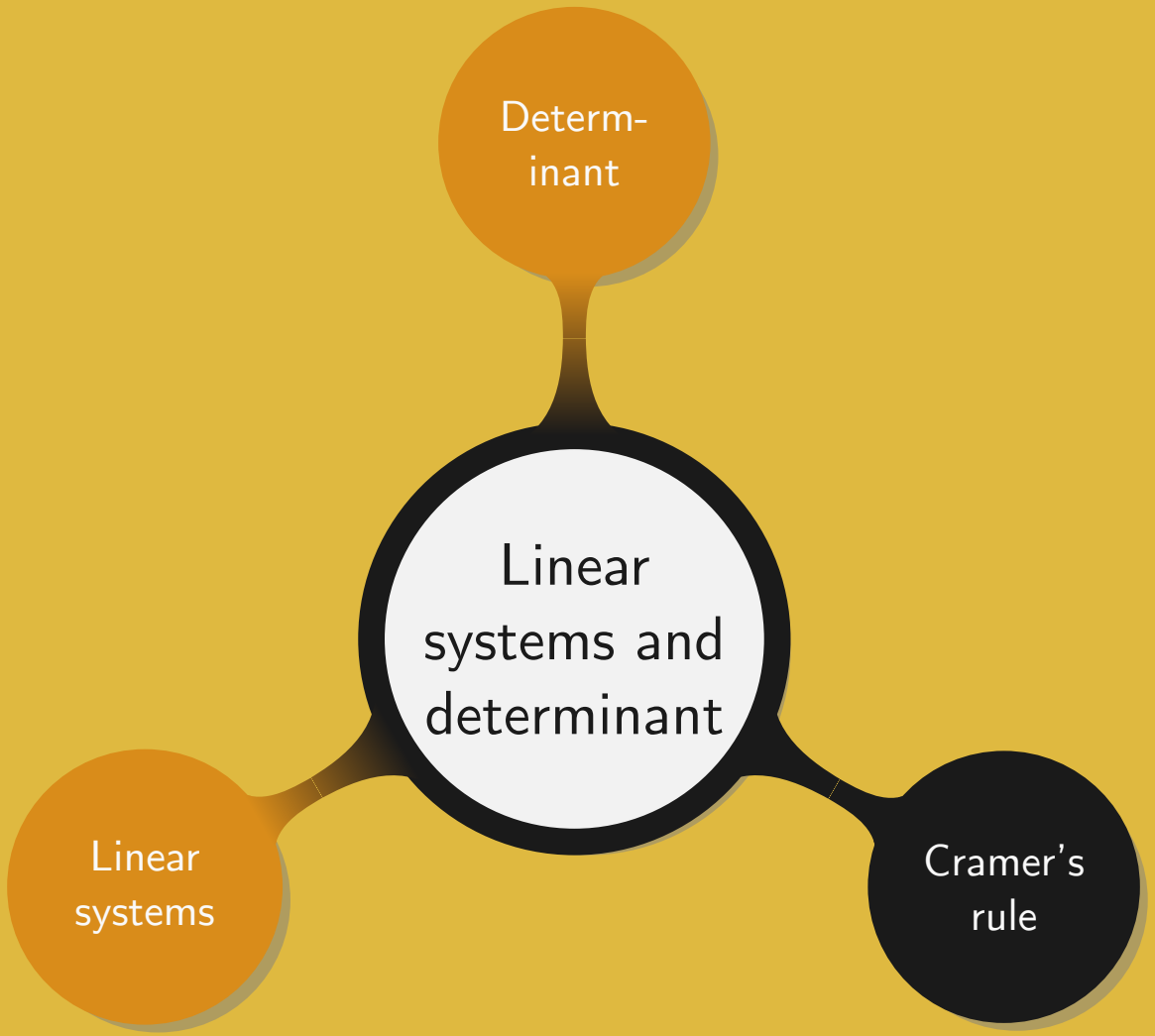
Let $A \in \mathcal{M}_n(\mathbb{K})$, $B \in \mathcal{M}_{n,p}(\mathbb{K})$, and $C \in \mathcal{M}_p(\mathbb{K})$. Then $\det\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det A \det C$.

Proof. First note that $\begin{vmatrix} A & B \\ 0 & C \end{vmatrix} = \begin{vmatrix} I_n & 0 \\ 0 & C \end{vmatrix} \begin{vmatrix} A & B \\ 0 & I_p \end{vmatrix}$ and then observe that when expanding $\begin{vmatrix} I_n & 0 \\ 0 & C \end{vmatrix}$ from the first row we find $\det C$.

Similarly expanding $\begin{vmatrix} A & B \\ 0 & I_p \end{vmatrix}$ from the last row we get $\begin{vmatrix} A & B \\ 0 & I_p \end{vmatrix} = \det A$.

$\det\begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det A \det C$. □

Chapter organisation



The history of the determinant

The determinant was discovered by the French mathematician Cramer. In his book *Introduction à l'analyse des lignes courbes algébriques*, published in 1750, he addresses the problem of finding the coefficients of a curve of order v passing through a given set of points. In particular he proves that a curve of order v is uniquely determined by $\frac{v^2}{2} + \frac{3v}{2}$ points and takes as example a curve of order two defined by

$$A + By + Cx + Dy^2 + Exy + x^2 = 0,$$

where A to E are five points through which the curve passes. He then constructs a system of five equations with five unknowns, A to F. While solving this system of equations he realised the existence of a "convenient and general rule, when there is an arbitrary number of equations and unknowns, all with degree at most one."

The details of this rule are presented in the appendix on his book. In particular he lists all the permutation of $\mathcal{S}_3$ and discusses the sign to be used depending on the order of the indices.

Definition

A *Cramer system* is a system with n equations and n unknowns which admits a unique solution.

Proposition

Let S be a linear system with n equations and n unknowns. Then the following properties are equivalent.

i. S is a Cramer system;

ii. The homogeneous system S_0 associated to S has a unique trivial solution $(0, \cdots, 0)$;

iii. The matrix of S is invertible;

Proof. Let A be the matrix of S and u be its corresponding endomorphism in the canoical basis.
(i⇒ii) Let X be the unique solution of S. If X_0 is a solution of S_0, then $X + X_0$ is a solution of X, and $X_0 = 0$.

(ii⇒iii) Since $(0, \cdots, 0)$ is the unique solution to S_0, it means that $\ker u = \{0\}$. Thus u is injective and bijective, i.e. A is invertible.

(iii⇒i) Since A is invertible, S admits a unique solution $X = A^{-1}B$. □

Remarks.

- To show that a matrix $A \in \mathcal{M}_n(\mathbb{K})$ is invertible it suffices to prove that the only solution to the system represented by $AX = 0$ is $(0, \cdots, 0)$.
- If a system with n equations and n unknowns can be transformed, using elementary row and column operations, into a triangular Cramer system, then it not only proves it is a Cramer system but also provides its solution.

Definition (Comatrix)

Let $A = (a_{i,j})_{1 \leq i,j \leq n} \in \mathcal{M}_n(\mathbb{K})$. The *comatrix* of A, denoted $\operatorname{com} A$, is the matrix composed of the cofactors of A.

Proposition

Let B be the comatrix of $A \in \mathcal{M}_n(\mathbb{K})$. Then

$$AB^\top = B^\top A = (\det A)\, \mathrm{I}_n\,.$$

Proof. Let $AB^\top = (c_{i,j})_{1 \leq i,j \leq n}$. For $i, k \in [\![1, n]\!]$,

$$c_{i,k} = \sum_{j=1}^{n} a_{i,j} b_{k,j} = \sum_{j=1}^{n} (-1)^{j+k} \Delta_{k,j} a_{i,j}.$$

If $k = i$, then the second sum corresponds to the development of the determinant of A with respect to the ith row. Hence $c_{i,i} = \det A$.

If $k \neq i$, then we define $A' = (a'_{i,j})_{1 \leq i,j \leq n}$ the matrix obtained from A by overwriting its kth row with its ith row. This gives

$$0 = \det A' = \sum_{j=1}^{n} (-1)^{j+k} \Delta'_{k,j} a'_{k,j}.$$

Recall that by construction we have $a'_{k,j} = a_{i,j}$, and since the rows of A and A', others than the kth one, are identical we have

$$0 = \det A' = \sum_{j=1}^{n} a_{i,j} (-1)^{j+k} \Delta_{k,j} = c_{i,k}.$$

Hence we have $AB^\top = (\det A)\, \mathrm{I}_n$. For second equality $B^\top A = (\det A)\, \mathrm{I}_n$, it suffices to consider the columns instead of the rows. □

Corollary

If B is the comatrix of a matrix $A \in \mathrm{GL}_n(\mathbb{K})$, then $A^{-1} = \frac{1}{\det A} B^\top$.

Example. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Its comatrix $B = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$. Then if A is invertible

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & b \\ -c & a \end{pmatrix}.$$

Theorem

Let S be a Cramer system expressed as $AX = B$. The unique solution of S is the n-tuple $(x_1, \cdots, x_n)$ such that for any $i \in [\![1, n]\!]$,

$$x_i = \frac{\det A_i}{\det A},$$

where A_i is the matrix A with ith column replaced by B.

Proof. As a Cramer system A has a unique solution and $\sum_{j=1}^{n} x_j C_j = B$, where the C_j are the columns of A.

So for any $i \in [\![1, n]\!]$,

$$\det(C_1, \cdots, C_{i-1}, B, C_{i+1}, \cdots, C_n) = \sum_{j=1}^{n} x_j \det(C_1, \cdots, C_{i-1}, C_j, C_{i+1}, \cdots, C_n).$$

Since all the terms in the sum are 0, except when $j = i$ we have shown that for all $i \in [\![1, n]\!]$, $\det A_i = x_i \det A$. □

Remarks.

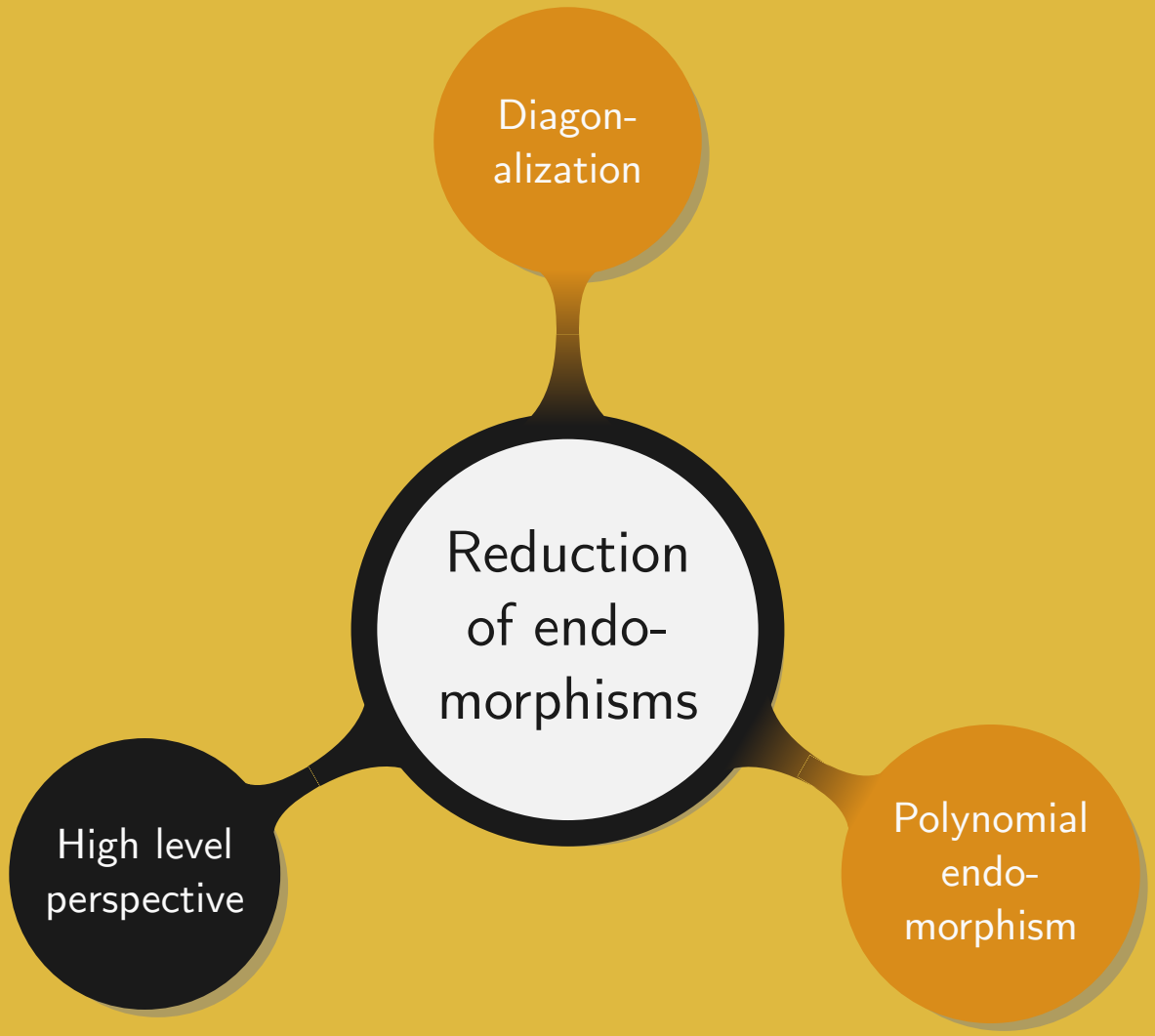
- Cramer's rule is seldom used to invert matrices. It is often preferable to directly solve the system of equations using Gauss elimination.
- When the matrix A depends upon a parameter Cramer's rule can be effectively used to determine various properties of the matrix A^{-1}. For instance information regarding the differentiability or continuity can be inferred.

Key points

- Describe how a linear system can be interpreted
- What is a multi-linear map?
- What is the connection between determinant and dual space?
- List all the most important properties of the determinant
- When to use Cramer's rule?

5. Reduction of endomorphisms

Chapter organisation



Looking back

As already emphasized, one of the main goal of mathematics is the classification of objects with respect to how they "look" or "behave". In the case of algebra, invariance is defined with respect to "transformations" applied when solving linear systems of equations. In particular in chapter 4 we observed that elementary operations on the rows and columns do not affect a linear system.

Stepping back a bit we noticed that a linear system can be represented as a matrix, which in turn can also be viewed as the expression of a linear map in some basis. In particular to a linear system with a unique solution corresponds an invertible matrix representing an isomorphism.

Those three perspectives, namely linear systems, matrices, and endomorphisms, offer much flexibility when solving a problem. Indeed, an alternative view on a problem can render it easier to solve or provide a nicer representation for instance when dealing with algorithms.

For instance we can solve a linear system in terms of matrices. But in fact even before solving it, we can apply the determinant to provide us with much insight on it. In chapter 4 we proved that the determinant of a matrix is invertible if and only if its associated linear system has a unique solution.

Moreover if A is a square matrix and $B = P^{-1}AP$ is similar to A, for some invertible matrix P (definition 3.140), then $\det A = \det B$. This is clear since

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

The matrices A and B being similar simply means that they represent a same endomorphism in two different bases. Hence the determinant is an invariant, i.e. regardless of the basis its value remains unchanged, so what matters is its value and not "where it is defined".

Jacobian determinant

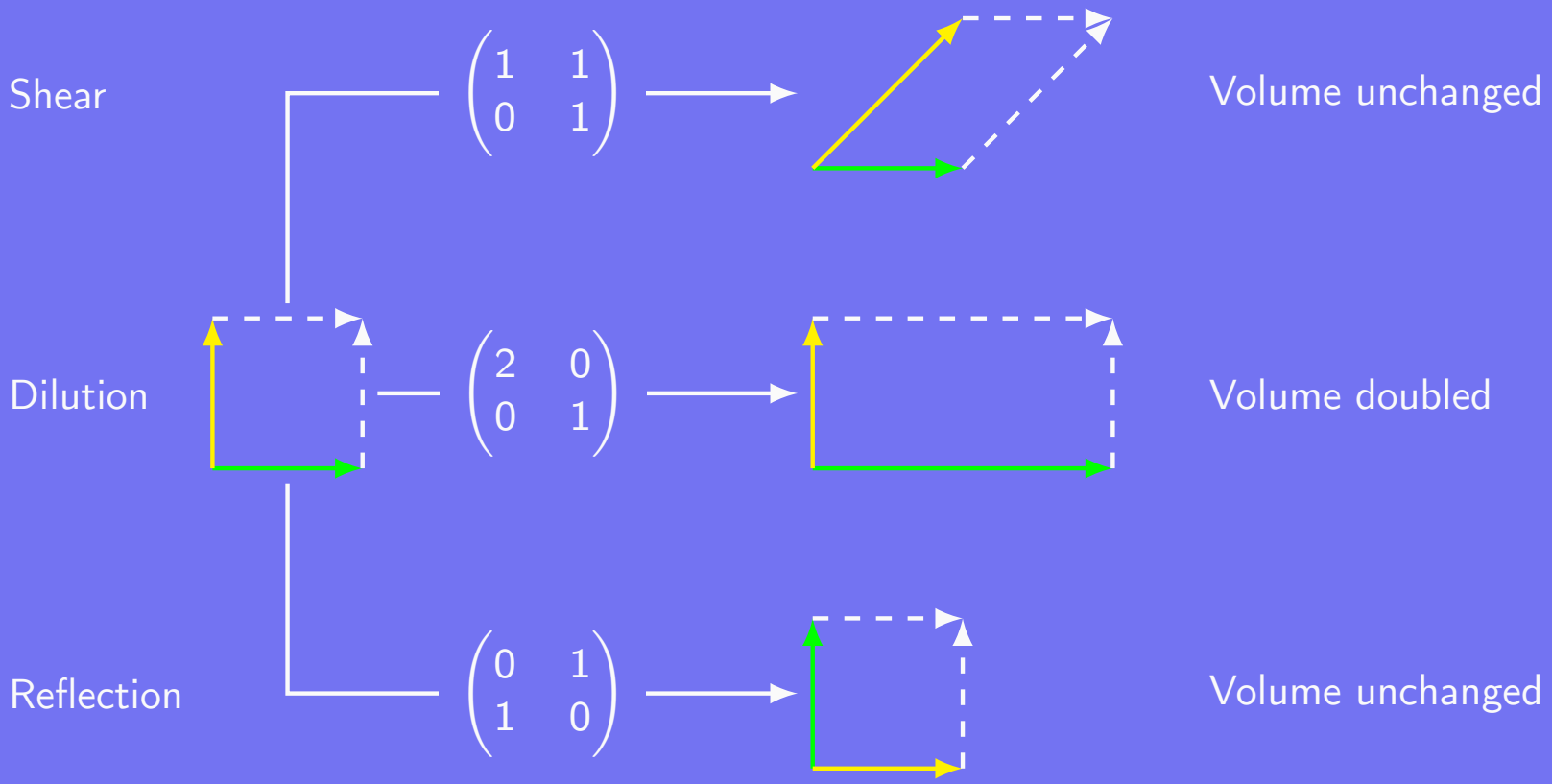
If we ponder on that last comment we can further relate our work on the determinant to other parts of mathematics. For instance in multivariate calculus, the *Jacobian determinant* is the determinant of the first oder approximation of a function at a given point.

For a function $f : \mathbb{R}^n \to \mathbb{R}^n$, the Jacobian determinant of f at a point $p \in \mathbb{R}^n$ provides information on the behaviour of f in its neighborhood. For instance if it is invertible then it means that the continuously differentiable function f is invertible in a neighborhood of p.

The Jacobian determinant also conveys information on how the function f shrinks or expands the space around p. This is especially important when applying a change of variable during an integration to determine a volume V. In order not to alter V, we need to account for the discrepancy introduced by the change of variable. To see it, note that in an n-dimensional space, $\mathrm{d}V$ is a parallelepiped and its volume is given by the determinant of its edge vectors.

Geometrical interpretation of the determinant

The above discussion highlights a new perspective on the determinant: its has a geometrical meaning. To illustrate it we consider three simple transformations from $\mathbb{R}^2$.



Geometrical interpretation of the determinant

The third transformation highlights that the determinant bears more than just the concept of volume, in fact it is a "signed volume". In that example the determinant is -1 so its absolute value is 1, meaning no change in volume. But it also records the "*chirality*" of the transformation, i.e. whether the resulting object can be superimposed on its own mirror image.

This, in fact hints for a topological interpretation of the determinant. In that field, object are classified with respect to their behaviour under smooth continuous transformations. Looking at our two first transformations it is clear that they are continuous, however for the reflection we notice that to go from the initial vector setup to the final one, using a deformation, we will have to consider the case where the two vectors are linearly dependent, i.e. their determinant or "signed volume" is 0. Hence a transformation with negative determinant does not preserve the orientation of the object it is applied to.

Definition

Let $A \in \mathcal{M}_n(\mathbb{K})$ representing an endomorphism $f \in \mathcal{L}(V)$.

- The *trace* of A is the sum of all the diagonal elements of A.
- The *trace* of f is the trace of any matrix representing f.

Proposition

1. The trace $\operatorname{tr} : \mathcal{M}_n(\mathbb{K}) \to \mathbb{K}$ is a linear form.
2. For any $A \in \mathcal{M}_{n,p}(\mathbb{K})$ and $B \in \mathcal{M}_{p,n}(\mathbb{K})$, $\operatorname{tr} AB = \operatorname{tr} BA$.
3. For any $A \in \mathcal{M}_n(\mathbb{K})$ and $P \in \mathrm{GL}_n(\mathbb{K})$, $\operatorname{tr} A = \operatorname{tr}(P^{-1}AP)$.

Proof. (1) Let $A = (a_{i,j})_{1 \leq i,j \leq n}$, $B = (b_{i,j})_{1 \leq i,j \leq n}$, and $\alpha \in \mathbb{K}$. Then

$$\operatorname{tr}(\alpha A + B) = \sum_{i=1}^{n} (\alpha a_{i,i} + b_{i,i}) = \alpha \sum_{i=1}^{n} a_{i,i} + \sum_{i=1}^{n} b_{i,i} = \alpha \operatorname{tr} A + \operatorname{tr} B.$$

(2) First observe that both AB and BA are in $\mathcal{M}_n(\mathbb{K})$. Then we have

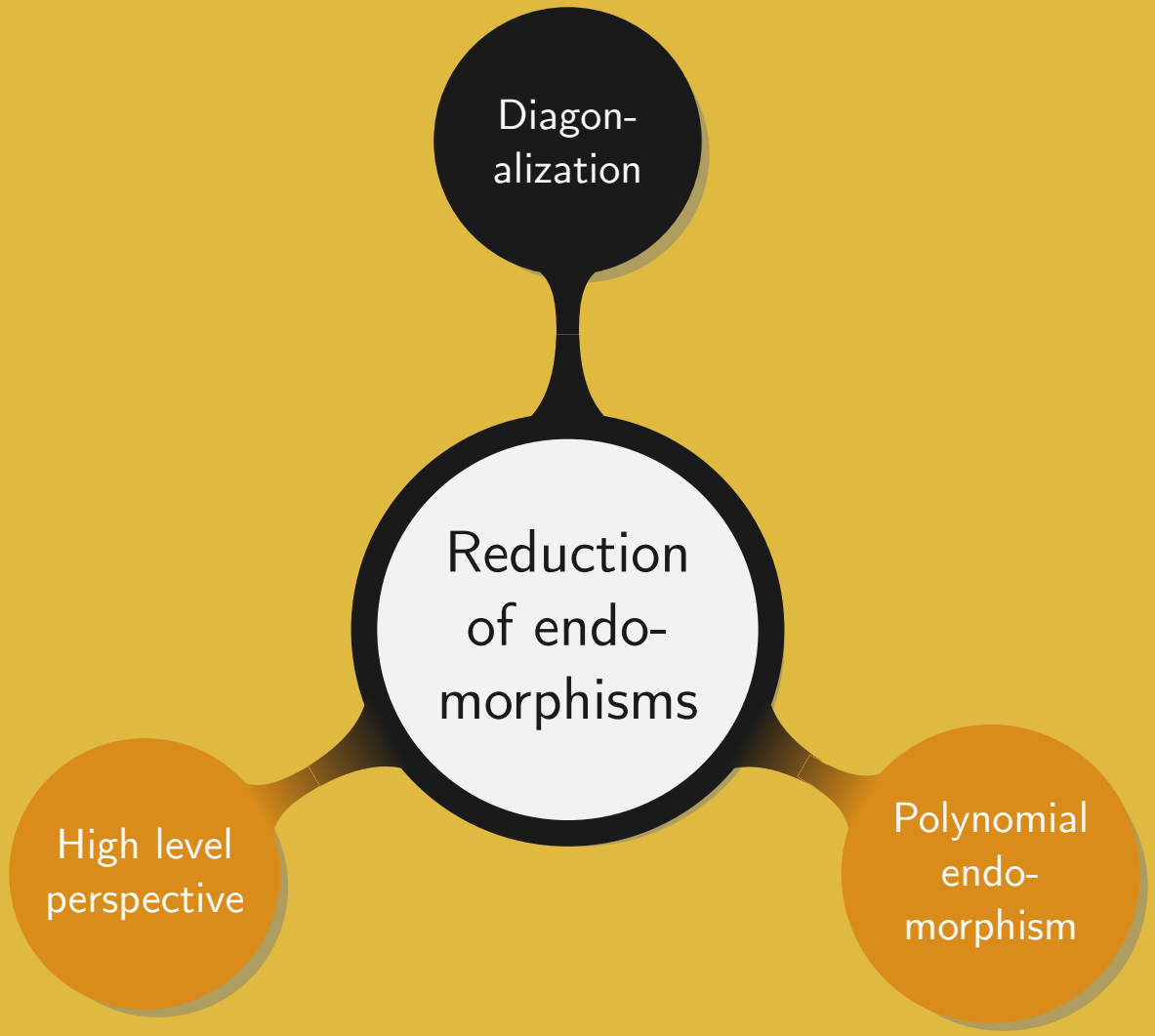
$$\operatorname{tr} AB = \sum_{i=1}^{n} \left(\sum_{j=1}^{p} a_{i,j} b_{j,i} \right) = \sum_{j=1}^{p} \left(\sum_{i=1}^{n} b_{j,i} a_{i,j} \right) = \operatorname{tr} BA.$$

(3) Using (2) we write $\operatorname{tr}(P^{-1}AP) = \operatorname{tr}(PP^{-1}A) = \operatorname{tr} A$. □

Remark. Comments on proposition 5.259.

- Translated in term of endomorphism it gives
  1. $\operatorname{tr} : \mathcal{L}(V) \to \mathbb{K}$ is a linear form;
  2. for any $f \in \mathcal{L}(V, W)$ and $g \in \mathcal{L}(W, V)$, $\operatorname{tr}(f \circ g) = \operatorname{tr}(g \circ f)$;
  3. for any $f \in \mathcal{L}(V)$ and $h \in \mathrm{GL}(V)$, $\operatorname{tr}(h^{-1} \circ f \circ h) = \operatorname{tr} f$;
- The third point states that the trace is independent of the basis, or in other words the trace is an invariant.

Chapter organisation



Definitions

Let V be a vector space, $f \in \mathcal{L}(V)$, $M \in \mathcal{M}_n(\mathbb{K})$, and $\lambda \in \mathbb{K}$.

1. We say that λ is an *eigenvalue* of f, if and only if there exists a non-zero $x \in V$, such that $f(x) = \lambda x$. The vector x is called the *eigenvector* corresponding to λ. The *spectrum* of f, denoted σ_f is the set of the eigenvalues of f.
2. We say that λ is an *eigenvalue* of M, if and only if there exists a non-zero $X \in \mathcal{M}_{n,1}(\mathbb{K})$, such that $MX = \lambda X$. The vector X is called the *eigenvector* corresponding to λ. The *spectrum* of M, denoted σ_M is the set of the eigenvalues of f.

Remark. If V has finite dimension and $\mathcal{B}$ be a basis such that M is the matrix of f, then being an eigenvector or an eigenvalue for f or M is equivalent.

Proposition

Let $f \in \mathcal{L}(V)$ and $\lambda \in \mathbb{K}$. The following statements are equivalent.

i. λ is an eigenvalue;
ii. $\ker(f - \lambda \operatorname{Id}) \neq \{0\}$;
iii. $f - \lambda \operatorname{Id}$ is not injective;

Proof. Those are reformulations of the definition of an eigenvalue. □

Remark. In the finite dimensional case, for a matrix M we have

$$\lambda \in \sigma_M \Leftrightarrow \operatorname{null}(M - \lambda \mathrm{I}_n) \neq \{0\} \Leftrightarrow M - \lambda \mathrm{I}_n \notin \mathrm{GL}_n(\mathbb{K}).$$

In that case this is also equivalent to $\operatorname{rank}(M - \lambda \mathrm{I}_n) < n$. Hence M is invertible means that $0 \notin \sigma_M$.

Definition

Let V be a vector space, $f \in \mathcal{L}(V)$. For any eigenvalue λ the vector subspace $\ker(f - \lambda \operatorname{Id})$ is composed of $\{0\}$ and of the eigenvectors associated to λ. Its is called the *eigenspace* associated to λ and we denote it $E_\lambda(f)$.

Remarks.

- In the finite dimensional case let M and X be the matrix representations of f and x, in a basis $\mathcal{B}$, respectively. It is equivalent to say that x is in the eigenspace of f associated to λ, and that X is in the eigenspace of M associated to λ.
- For a square matrix M, we will denote by M_λ, the eigenspace associated to λ.

Example. We want to determine the eigenvalues and their associated eigenvectors for $M \in \mathcal{M}_n(\mathbb{K})$ defined by

$$\begin{pmatrix} 1 & \cdots & \cdots & 1 \\ \vdots & & & \vdots \\ \vdots & & 1 & \vdots \\ \vdots & & & \vdots \\ 1 & \cdots & \cdots & 1 \end{pmatrix}.$$

Let $\lambda \in \mathbb{R}$ and $X = (x_1, \cdots, x_n) \in \mathcal{M}_{n,1}(\mathbb{K}) \setminus \{0\}$. By equivalence we have, $X \in M_\lambda$ if and only if $MX = \lambda X$. This can be translated in terms of linear system

$$\begin{cases} x_1 + \cdots + x_n = \lambda x_1 \\ \qquad \vdots \\ x_1 + \cdots + x_n = \lambda x_n \end{cases},$$

which can have two "types of solution": (i) $\lambda \neq 0$, $x_1 = \cdots = x_n$, and $\lambda = n$, and (ii) $\lambda = 0$ and $x_1 + \cdots + x_n = 0$.

Therefore we can conclude $\sigma_M = \{0, n\}$ and

- $M_0 = \left\{(x_1, \cdots, x_n) \in \mathcal{M}_{n,1}(\mathbb{K})) : \sum_{i=1}^{n} x_i = 0\right\}$ is a hyperplane of $\mathcal{M}_{n,1}(\mathbb{R})$;
- $M_n = \left\{(1, \cdots, 1)\mathbb{R}\right\}$ is a vectorial line;

Proposition

Let $f \in \mathcal{L}(V)$, and $\lambda_1, \cdots, \lambda_N$ be N eigenvalues of f, distinct two by two. The eigenspaces associated to the eigenvalues of f are in direct sum.

Proof. We will prove the result by induction on N.

The case $N = 1$ being trivial we assume it is true for N and show it is true for $N + 1$. For $i \in [\![1, N + 1]\!]$, let $x_i \in V$, such that $x_i \in E_{\lambda_i}(f)$ and the sum of all the x_i is 0.

Then after applying f we obtain $0 = \sum_{i=1}^{N+1} f(x_i) = \sum_{i=1}^{N+1} \lambda_i x_i$. This yields

$$\begin{cases} x_1 + \cdots + x_N + x_{N+1} = 0 \\ \lambda_1 x_1 + \cdots + \lambda_N x_N + \lambda_{N+1} x_{N+1} = 0, \end{cases}$$

which becomes $(\lambda_{N+1} - \lambda_1)x_1 + \cdots + (\lambda_{N+1} - \lambda_N)x_N = 0$, if we do λ_{N+1} times the first row minus the second.

For any $i \in [\![1, N]\!]$, $(\lambda_{N+1} - \lambda_i)x_i \in E_{\lambda_i}(f)$. Moreover by applying our induction hypothesis all those eigenspaces are in direct sum, meaning that $(\lambda_{N+1} - \lambda_i)x_i = 0$. Recalling that all the λ_i are distinct two by two we see that $x_i = 0$, for all $i \in [\![1, N]\!]$.

Therefore, $x_{N+1} = \sum_{i=1}^{N} x_i = 0$, and by applying the induction principle this shows that all the $E_{\lambda_i}(f)$ are in direct sum. $\square$

Remark. Proposition 5.267 only states that the eigenspaces associated to the eigenvalues of f are in direct sum. It however does not mean that their direct sum equals to V. Necessary and sufficient conditions to ensure that the direct sum of all eigenspaces is V will be investigated in theorem 5.279.

Although the following discussion can be adjusted to suit any field, we restrict our attention to the case where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$.

Let $A \in \mathcal{M}_n(\mathbb{K})$. By construction of the determinant the map

$$\begin{aligned} \chi_A : \mathbb{K} &\longrightarrow \mathbb{K} \\ \lambda &\longmapsto \det(A - \lambda \, \mathrm{I}_n) \end{aligned}$$

defines a polynomial in λ. Moreover using theorem 4.232, we see that we can also define a similar polynomial $\det(f - \lambda \, \mathrm{Id})$, where f is the linear map whose matrix representation is A in some given basis. We call χ_A and χ_f, the *characteristic polynomial* of A and f, respectively.

Proposition

Let $A \in \mathcal{M}_n(\mathbb{K})$. Then for any $\lambda \in \mathbb{K}$,

$$\chi_A(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} \operatorname{tr}(A) \lambda^{n-1} + \cdots + \det A.$$

Proof. Using formula (4.4) we have

$$\begin{aligned} \chi_A(\lambda) &= \det(A - \lambda \mathrm{I}_n) \\ &= \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{j=1}^{n} \alpha_{\sigma(j),j}, \end{aligned}$$

where $\alpha_{i,j} = a_{i,j}$, if $i \neq j$ and $a_{i,i} - \lambda$ otherwise.

Considering all the permutations we notice that only the identity can yields degrees n and $n-1$. Indeed as soon as an element i is permuted with j it means that both i and j are permuted.

Hence we directly get

$$\begin{aligned}\sum_{\sigma\in\mathcal{S}_n}\varepsilon(\sigma)\prod_{j=1}^{n}\alpha_{\sigma(j),j} &= \prod_{j=1}^{n}\alpha_{j,j}+\sum_{\sigma\in\mathcal{S}_n\setminus\{\mathrm{Id}\}}\varepsilon(\sigma)\prod_{j=1}^{n}\alpha_{\sigma(j),j}\\ &= \prod_{j=1}^{n}(a_{j,j}-\lambda)+\sum_{\sigma\in\mathcal{S}_n\setminus\{\mathrm{Id}\}}\varepsilon(\sigma)\prod_{j=1}^{n}\alpha_{\sigma(j),j}\\ &= (-\lambda)^n+\sum_{j=1}^{n}a_{j,j}(-\lambda)^{n-1}+Q(\lambda)\\ &= (-\lambda)^n+(-1)^{n-1}\operatorname{tr}(A)\lambda^{n-1}+Q(\lambda),\end{aligned}$$

where $Q(\lambda)$ is a polynomial over $\mathbb{K}$, such that $Q(0)=\det A=\chi_A(0)$.

□

Proposition

If A and B are two similar matrices in $\mathcal{M}_n(\mathbb{K})$, then they have same characteristic polynomial.

Proof. Since A and B are similar, there exists $P \in \mathrm{GL}_n(\mathbb{K})$ such that $B = P^{-1}AP$. Thus

$$\begin{aligned}\chi_B(\lambda) &= \det(B - \lambda \mathrm{I}_n) = \det(P^{-1}AP - \lambda \mathrm{I}_n) \\ &= \det(P^{-1}(A - \lambda \mathrm{I}_n)P) = (\det P)^{-1} \det(A - \lambda \mathrm{I}_n) \det P \\ &= \det(A - \lambda \mathrm{I}_n) = \chi_A(\lambda).\end{aligned}$$

□

Remark. The converse of this proposition is false. Indeed note that for $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$, we have $\chi_A(\lambda) = \chi_B(\lambda) = \lambda^2$ but A and B are not similar.

Proposition

For any $f \in \mathcal{L}(V)$, the spectrum of f, σ_f, is given by $\chi_f^{-1}(\{0\})$, and for $A \in \mathcal{M}_n(\mathbb{K})$, the matrix representation of f, $\sigma_A = \chi_A^{-1}(\{0\})$.

Proof. It suffices to prove the result for f. We take $\lambda \in \mathbb{K}$ and see that by definition $\lambda \in \sigma_f$ is equivalent to $\ker(f - \lambda \operatorname{Id}) \neq \{0\}$. This is the case if and only if $f - \lambda \operatorname{Id}$ is not injective, that is if and only if $\det(f - \lambda \operatorname{Id}) = 0$. This exactly means $\chi_f(\lambda) = 0$. □

Corollary

A square matrix of rank n possesses at most n eigenvalues.

Remark. The two above results are very important in practice as they provide a simple strategy to determine the eigenvalues of a matrix.

Proposition

Let $f \in \mathcal{L}(V)$, V_0 be a vector subspace of V, and g be the restriction of f to V_0. Then χ_g divides χ_f.

Proof. Let $\mathcal{B}_0 = \{e_1, \cdots, e_p\}$ be a basis of V_0. Then by the incomplete basis theorem (2.93), we can expand it into $\mathcal{B} = \mathcal{B}_0 \cup \{e_{p+1}, \cdots, e_n\}$, a basis for V.

Calling A the matrix of g in basis $\mathcal{B}_0$, there exist $B \in \mathcal{M}_{p,n-p}(\mathbb{K})$ and $C \in \mathcal{M}_{n-p}(\mathbb{K})$, such that $M_{\mathcal{B}}(f) = \left(\begin{smallmatrix} A & B \\ 0 & C \end{smallmatrix}\right)$. Hence, by proposition 4.242, for any $\lambda \in \mathbb{K}$

$$\chi_f(\lambda) = \det(A - \lambda \mathrm{I}_p) \det(C - \lambda \mathrm{I}_{n-p}) = \chi_g \chi_C(\lambda).$$

Since $\chi_C \in \mathbb{K}[X]$, we have proven that χ_g divides χ_f. □

Definitions

Let $f \in \mathcal{L}(V)$ and λ_0 be an eigenvalue for f.

1. The *algebraic multiplicity* of λ_0 is its degree as root of χ_f.
2. The dimension of the eigenspace $E_{\lambda_0}(f)$ is called the *geometric multiplicity* of λ_0.

Remark. The above definition remains the same if $f \in \mathcal{L}(V)$ is changed for $A \in \mathcal{M}_n(\mathbb{K})$.

Proposition

Let $f \in \mathcal{L}(V)$ and λ_0 an eigenvalue of f, with algebraic and geometric multiplicities ω_0 and d_0, respectively. Then $1 \leq d_0 \leq \omega_0$.

Proof. Clearly as $E_{\lambda_0}(f) = \ker(f - \lambda_0 \operatorname{Id}) \neq \{0\}$, we know that d_0 is at least 1.

Recalling that $E_{\lambda_0}(f)$ is a d_0-dimensional $\mathbb{K}$-vector space, it admits a basis $\{e_1, \cdots, e_{d_0}\}$. By the incomplete basis theorem (2.93), there exist $e_{d_0+1}, \cdots, e_n \in V$, such that $\mathcal{B} = \{e_1, \cdots, e_{d_0}, e_{d_0+1}, \cdots, e_n\}$ is a basis of V.

Thus we can write the matrix of f in $\mathcal{B}$, $M_{\mathcal{B}}(f) = \begin{pmatrix} \lambda_0 \operatorname{I}_{d_0} & C \\ 0 & B \end{pmatrix}$, where $C \in \mathcal{M}_{d_0, n-d_0}(\mathbb{K})$ and $B \in \mathcal{M}_{n-d_0}(\mathbb{K})$. This leads to

$$\chi_f(\lambda) = \begin{vmatrix} (\lambda_0 - \lambda) \operatorname{I}_{d_0} & C \\ 0 & B - \lambda \operatorname{I}_{n-d_0} \end{vmatrix},$$

which by proposition 4.242 yields

$$\chi_f(\lambda) = (\lambda_0 - \lambda)^{d_0} \det(B - \lambda \operatorname{I}_{n-d_0}) = (\lambda_0 - \lambda)^{d_0} \chi_B(\lambda).$$

Hence $(\lambda_0 - \lambda)^{d_0}$ divides $\chi_f(\lambda)$, i.e. $d_0 \leq \omega_0$. □

Definitions

1. A map $f \in \mathcal{L}(V)$ is said to be *diagonalizable* if there exists a basis in which the matrix of f is diagonal.
2. A matrix $A \in \mathcal{M}_n(\mathbb{K})$ is said to be *diagonalizable* if there exists a matrix $D \in \mathcal{M}_n(\mathbb{K})$ such that A and D are similar.

Remarks.

- If f is diagonalizable, then its matrix A is similar to a diagonal matrix D, i.e. there exists $P \in \mathrm{GL}_n(\mathbb{K})$ such that $A = PDP^{-1}$. In general D and P are not unique.
- If A represents f, it is equivalent to say that f or A is diagonalizable.
- Some endomorphisms cannot be diagonalized, but only triangularized, i.e. written as a triangular matrix, while others can neither be diagonalized not triangularized.

Lemma

For a $f \in \mathcal{L}(V)$, the following properties are equivalent.

i. f can be diagonalized;
ii. The eigenvectors of f form a basis for V;
iii. The sum of the eigenspaces of f is equal to V;
iv. The sum of the dimensions of the eigenspaces of f is equal to the dimension of V;

Proof. (i)⇒(ii) If f is diagonalizable, then there exists a basis $\mathcal{B} = \{e_1, \cdots, e_n\}$, such that $M_{\mathcal{B}}(f)$ is diagonal, i.e. with values $\lambda_1, \cdots, \lambda_n$ on the diagonal and 0 everywhere else. Since for any $1 \leq i \leq n$, $f(e_i) = \lambda_i e_i$ and $e_i \neq 0$ this shows that the e_i are eigenvectors of f.

(ii)⇒(iii) Assuming the existence of a basis $\mathcal{B}$ composed of the eigenvectors of f, then the set of the eigenvalues of f associated to the eigenvectors composing $\mathcal{B}$, $\{\lambda_i : 1 \leq i \leq n\}$, is included in the spectrum of f. But as V has dimension n it shows that

$$\sum_{i=1}^{n} \ker(f - \lambda_i \operatorname{Id}) = \sum_{i=1}^{n} \mathbb{K} e_i = V.$$

(iii)⇒(iv) By proposition 5.266, all the eigenspaces are in direct sum, so the sum of their dimensions must equal the dimension of V.

(iv)⇒(i) Let k be the cardinal of σ_f. Noting that each eigenspace $E_{\lambda_i}(f)$ has a basis $\mathcal{B}_i$, let $\mathcal{B} = \cup_{i=1}^{k} \mathcal{B}_i$.
Since each $\mathcal{B}_i$ are linearly independent and the eigenspaces of f are in direct sum, we know that $\mathcal{B}$ is linearly independent. Besides, $\operatorname{card} \mathcal{B} = \sum_{i=1}^{k} \operatorname{card} \mathcal{B}_i = \dim V = n$, $\mathcal{B}$ is a basis for V, and as it is composed of the eigenvectors of f, then the matrix of f in basis $\mathcal{B}$ is diagonal. □

Remark. From the previous proof we see that if f is diagonalizable into a matrix D, then the elements on the diagonal of D are the eigenvalues of f, each appearing as many times as its algebraic multiplicity.

Theorem

An endomorphism $f \in \mathcal{L}(V)$ is diagonalizable if and only if χ_f splits on $\mathbb{K}$ and for each eigenvalue λ of f, $\dim(E_\lambda(f))$ is equal to the algebraic multiplicity of λ.

Proof. ($\Rightarrow$) Since f is diagonalizable in a basis $\mathcal{B}$, we easily obtain $\chi_f = \det(f - \lambda \mathrm{Id}) = \prod_{i=1}^{n}(\lambda_i - \lambda)$, where the λ_i are the elements on the diagonal of $M_{\mathcal{B}}(f)$. Hence χ_f splits on $\mathbb{K}$ and $\sum_{\lambda \in \sigma_f} \omega(\lambda) = n$, where $\omega(\lambda)$ is the algebraic multiplicity of λ.

Moreover lemma 5.277 says that $n = \sum_{\lambda \in \sigma_f} d(\lambda)$, for $d(\lambda) = \dim(E_\lambda(f))$.

Recalling that for any $\lambda \in \sigma_f$, $d(\lambda) = \dim(E_\lambda(f)) \leq \omega(\lambda)$, with $\omega(\lambda)$ (proposition 5.274), this shows that $d(\lambda) = \omega(\lambda)$ for all $\lambda \in \sigma_f$.

($\Leftarrow$) Since χ_f splits on $\mathbb{K}$ and its roots are the eigenvalues of f we get $\sum_{\lambda \in \sigma_f} \omega(\lambda) = n$. But since by assumption $d(\lambda) = \omega(\lambda)$ for all λ, this shows that $\sum_{\lambda \in \sigma_f} d(\lambda) = n$. Thus by lemma 5.277 we know the f is diagonalizable. □

Corollary

Let $f \in \mathcal{L}(V)$. If f has n eigenvalues distinct two by two, then f is diagonalizable.

Remarks.

- The two previous results immediately translate for $A \in \mathcal{M}_n(\mathbb{K})$.
- A triangular matrix with its diagonal elements distinct two by two is diagonalizable since its characteristic polynomial splits.

Example. In order to diagonalize $A = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 1 & 1 \\ -2 & 0 & -1 \end{pmatrix} \in \mathcal{M}_3(\mathbb{R})$, we start by calculating χ_A. Developing the determinant with respect to the first row, we get

$$\chi_A(\lambda) = (2-\lambda)(1-\lambda)(-1-\lambda)+2(1-\lambda) = (1-\lambda)(\lambda^2-\lambda) = -\lambda(\lambda-1)^2.$$

Hence A has two eigenvalues, 0 and 1, with algebraic multiplicity 1 and 2, respectively. We now determine their associated eigenvectors.

- $\lambda = 0$: we need to find a vector $X = (x, y, z)$ such that $AX = 0$. Thus

$$\begin{cases} 2x + z = 0 \\ x + y + z = 0 \\ -2x - z = 0 \end{cases} \iff X = (1, 1, -2).$$

which yields $E_0(A)$ has dimension one and basis $\mathcal{B}_0 = \{X\}$.

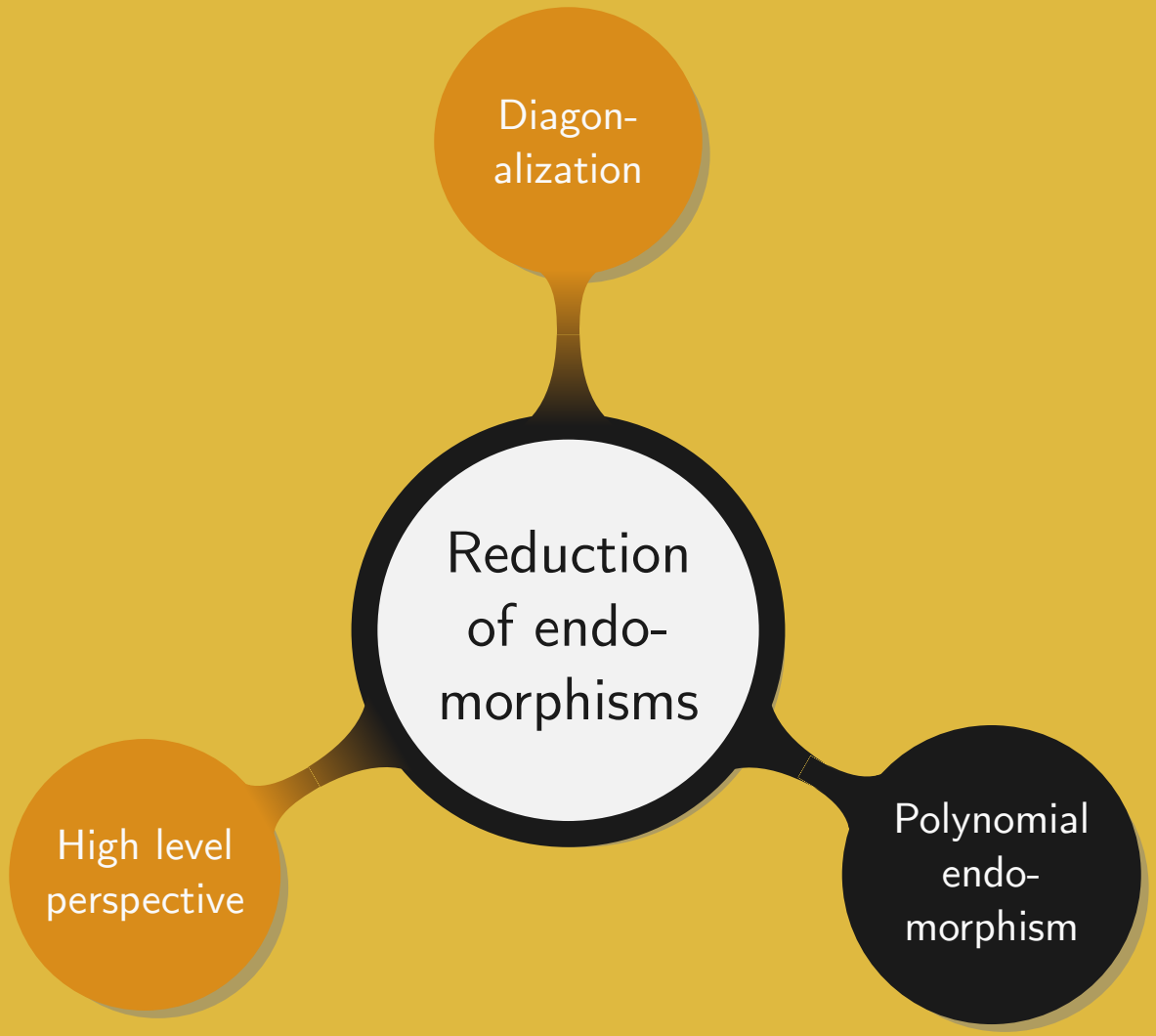
- $\lambda = 1$: we need to find up to two vectors $X = (x, y, z)$ such that $AX = X$. Thus

$$\begin{cases} 2x + z = x \\ x + y + z = y \\ -2x - z = z \end{cases} \iff x + z = 0.$$

  Hence, $E_1(A)$ has dimension two and $\mathcal{B}_1 = \{(0, 1, 0), (1, 0, -1)\}$ defines a basis for it.

This shows that χ_A splits on $\mathbb{R}$ and that the algebraic multiplicity of each eigenvalue of A is equal to the dimension of its associated eigenspace. We can therefore conclude that A is diagonizable (theorem 5.279). In particular in basis $\mathcal{B}_0 \cup \mathcal{B}_1$, we have $A = PDP^{-1}$ with $P = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -2 & 0 & -1 \end{pmatrix}$ and $D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Chapter organisation



Polynomial of an endomorphism

Definitions

Let $P(X) = \sum_{i=0}^{N} a_i X^i \in \mathbb{K}[X]$.

1. For $f \in \mathcal{L}(V)$ we define the *polynomial of the endomorphism* f as $P(f) = \sum_{i=0}^{N} a_i f^i$.
2. For $A \in \mathcal{M}_n(\mathbb{K})$ we define the *polynomial of the matrix* A as $P(A) = \sum_{i=0}^{N} a_i A^i$.

Remark. Given a basis $\mathcal{B}$ and calling A the matrix of f in $\mathcal{B}$, then $P(A) = M_{\mathcal{B}}(P(f))$.

Proposition

Let $f \in \mathcal{L}(V)$. For any $\alpha \in \mathbb{K}$ and $P, Q \in \mathbb{K}[X]$

$$(\alpha P + Q)(f) = \alpha P(f) + Q(f), \text{ and } (PQ)(f) = P(f) \circ Q(f).$$

Proof. Let $P(X) = \sum_{i=0}^{N} a_i X^i$ and $Q(X) = \sum_{i=0}^{N} b_i X^i$. Then

$$(\alpha P + Q)(f) = \left(\sum_{i=0}^{N} (\alpha a_i + b_i) X^i \right)(f) = \sum_{i=0}^{N} (\alpha a_i + b_i) f^i$$

$$= \alpha \sum_{i=0}^{N} a_i f^i + \sum_{i=0}^{N} b_i f^i = \alpha P(f) + Q(f).$$

For the second equality note that for any $i > N$, $a_i = b_i = 0$. Thus

$$(PQ)(f) = \left(\sum_{k=0}^{2N} \left(\sum_{i=0}^{k} a_i b_{k-i} \right) X^k \right)(f) = \sum_{k=0}^{2N} \sum_{i=0}^{k} a_i b_{k-i} f^k$$

$$= \left(\sum_{i=0}^{N} a_i f^i \right) \circ \left(\sum_{j=0}^{N} b_j f^j \right) = P(f) \circ Q(f).$$

□

Remarks.

- The previous result also applies to matrices of $\mathcal{M}_n(\mathbb{K})$.
- Let $A \in \mathcal{M}_n(\mathbb{K})$, $n \geq 2$. The linear map $\theta : \mathbb{K}[X] \to \mathcal{M}_n(\mathbb{K})$ such that $\theta(P) = P(A)$ might be neither injective nor surjective. For instance, take $\mathbb{K} = \mathbb{R}$ and $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Noting that $A^2 = \mathrm{I}_2$, we see that θ is not injective as $\theta(X^2 - 1) = A^2 - \mathrm{I}_2 = 0$. Moreover $\operatorname{im}\theta = \operatorname{span}\{\mathrm{I}_2, A\}$ is a 2-dimensional space while $\mathcal{M}_2(\mathbb{K})$ has dimension 4. Hence θ is not surjective.

Proposition

Let f and g be two linear maps which commute. Then any polynomial in f commutes with any polynomial in g.

Proof. Let $f, g \in \mathcal{L}(V)$ such that $f \circ g = g \circ f$. We will show by induction that for any $k \in \mathbb{N}$, $g^k \circ f = f \circ g^k$.

The property is clearly true for $k = 0$ and $k = 1$ by assumption. So we assume it to be true for k and prove it for $k + 1$. Using the induction hypothesis we write

$$g^{k+1} \circ f = g \circ (g^k \circ f) = g \circ (f \circ g^k) = (g \circ f) \circ g^k = (f \circ g) \circ g^k = f \circ g^{k+1}.$$

Based on the induction principle the property is verified for any $k \in \mathbb{N}$. Thus by linearity, for any $Q \in \mathbb{K}[X]$, we get $Q(g) \circ f = f \circ Q(g)$. We finally re-apply linearity for any $P \in \mathbb{K}[X]$ to the previous equality to finally obtain $Q(g) \circ P(f) = P(f) \circ Q(g)$. □

Remark. The previous result also applies to any two commuting matrices.

Annihilating polynomial

Definitions

1. Let $f \in \mathcal{L}(V)$. Any polynomial $P \in \mathbb{K}[X]$ such that $P(f) = 0$ is called an *annihilating polynomial* of f.
2. Let $A \in \mathcal{M}_n(\mathbb{K})$. Any polynomial $P \in \mathbb{K}[X]$ such that $P(A) = 0$ is called an *annihilating polynomial* of A.

Examples.

- f is a projector, if and only if $P(X) = X^2 - X$ is an annihilating polynomial for f.
- f is a symmetry, if and only if $P(X) = X^2 - 1$ is an annihilating polynomial for f.

Remarks.

- For any finite n-dimensional vector space V we know that $\mathcal{L}(V)$ has dimension n^2 (theorem 3.141). Therefore the set $\{\mathrm{Id}, f, \cdots, f^{n^2}\}$ composed of n^2+1 elements is linearly dependent and there exist α_i, not all zero, $1 \leq i \leq n^2+1$, such that $\alpha_0 \mathrm{Id} + \alpha_1 f + \cdots + \alpha_{n^2} f^{n^2} = 0$. Calling $P(X) = \sum_{i=0}^{n^2} \alpha_i X^i$ we see that P is not identically zero and $P(f) = 0$. This shows that in a finite dimensional space any endomorphism has a non-zero annihilating polynomial.
- If V is not a finite dimensional vector space, then the zero polynomial might be the only annihilating polynomial for some endomorphism.
- Any matrix $A \in \mathcal{M}_n(\mathbb{K})$ has a non-zero annihilating polynomial.

Proposition

Let $f \in \mathcal{L}(V)$, $\lambda \in \sigma_f$, and $x \in E_\lambda(f)$. Then for any $P \in \mathbb{K}[X]$, $\big(P(f)\big)(x) = P(\lambda)x$.

Proof. We start by proving, by induction, that for any $k \in \mathbb{N}$, $f^k(x) = \lambda^k x$. The property is clear for $k = 0$, since $f^0 = \mathrm{Id}$ and $\lambda^0 = 1$. By assumption it is also true for $k = 1$.

We proceed from k to $k+1$. Applying the induction hypothesis we write

$$f^{k+1}(x) = f\left(f^k\right)(x) = f(\lambda^k x) = \lambda^k f(x) = \lambda^{k+1} x.$$

By the induction principle this is true for any monomial, and thus for any $P(X) = a_0 + a_1 X + \cdots + a_N X^N \in \mathbb{K}[X]$,

$$\big(P(f)\big)(x) = \left(\sum_{i=0}^{N} a_i f^i\right)(x) = \sum_{i=0}^{N} a_i f^i(x) = \sum_{i=0}^{N} a_i \lambda^i x = P(\lambda)x.$$

□

Annihilating polynomial

Remarks.

- The previous result immediately translates in terms of matrices. Let $A \in \mathcal{M}_n(\mathbb{K})$, $\lambda \in \sigma_A$, and $V \in E_\lambda(A)$. Then for any $P \in \mathbb{K}[X]$, $P(A)V = P(\lambda)V$.
- This result is very useful in practice as it allows to translate a "complicated" polynomial of endomorphism into a more simple one on some eigenvalue.

Corollary

Let $f \in \mathcal{L}(V)$ and P be an annihilating polynomial for f. Then $\sigma_f \subset P^{-1}(\{0\})$.

Proof. Let $\lambda \in \sigma_f$. Since P is an annihilating polynomial, applying proposition 5.290 to it, yields $P(\lambda)(x) = \big(P(f)\big)(x) = 0(x) = 0$. This shows that $P(\lambda) = 0$, since $x \neq 0$. □

Remarks.

- The previous corollary directly translates in terms of matrices.
- The inclusions are not necessarily equalities. For instance over $\mathbb{R}$, we can take $A = 0$ and $P(X) = X(X-1)$. In that case P is an annihilating polynomial for A, but not all its roots are eigenvalues of A. In fact, this leads to the concept of *minimal polynomial*, a monic polynomial whose roots are exactly the eigenvalues of A, or f is we look at the endomorphism whose matrix is A in some basis.

Theorem

An endomorphism $f \in \mathcal{L}(V)$ is diagonalizable if and only if there exists $P \in \mathbb{K}[X]$ such that $P(f) = 0$ and P splits on $\mathbb{K}$ with simple roots only.

Proof. ($\Rightarrow$) Let $N = \operatorname{card} \sigma_f$, i.e. there exist $\lambda_1, \cdots, \lambda_N$, distinct two by two, such that the matrix of f in a basis $\mathcal{B}$ is

$$A = \begin{pmatrix} \lambda_1 \mathrm{I}_{d_1} & & 0 \\ & \ddots & \\ 0 & & \lambda_N \mathrm{I}_{d_N} \end{pmatrix},$$

where $d_i = \dim E_\lambda(F)$, $i \in [\![1, N]\!]$. Note that $P(X) = \prod_{i=1}^{N}(X - \lambda_i)$ splits with simple roots only and $P(A) = 0$, since each term of the product cancels a $\lambda_i \mathrm{I}_{d_i}$ block. This shows that P is an annihilating polynomial for f.

($\Leftarrow$) Conversely we assume the existence of $P \in \mathbb{K}[X]$ which splits on $\mathbb{K}$ with simple roots $\lambda_1, \cdots, \lambda_p$, i.e. $P(X) = \alpha \prod_{i=1}^{p}(X - \lambda_i)$, with $\alpha \neq 0$, and such that $P(f) = 0$.

We use the Lagrange interpolation polynomial $\ell_i(X) = \prod_{\substack{j=1 \\ j\neq i}}^{p} \frac{X-\lambda_j}{\lambda_i-\lambda_j}$ introduced in example 3.176. Recalling that for $Q, R \in \mathbb{K}[X]$, $Q(f) \circ R(f) = QR(f)$, (proposition 5.284) we have for any $i \in [\![1, p]\!]$ and $x \in V$,

$$\begin{aligned}(f - \lambda_i \operatorname{Id})\left((\ell_i(f))(x)\right) &= (X - \lambda_i)(f)(\ell_i(f))(x) \\ &= ((X - \lambda_i)\ell_i(X))(f)(x).\end{aligned}$$

Note that the polynomial $(X - \lambda_i)\ell_i(X)$ has the same roots as P which is an annihilating polynomial for f. Hence $(f - \lambda_i \operatorname{Id})\left((\ell_i(f))(x)\right) = 0$, or in other words for any $1 \leq i \leq p$, $\ell_i(f) \in \ker(f - \lambda_i \operatorname{Id})$.

Now observe that $\sum_{i=1}^{p} \ell_i = 1$. Indeed, for all $1 \leq k \neq i \leq p$, $\ell_i(\lambda_k) = \prod_{\substack{j=1 \\ j\neq i}}^{p} \frac{\lambda_k-\lambda_j}{\lambda_i-\lambda_j} = 0$, but when $k = i$ we get $\ell_i(\lambda_i) = 1$. Hence $\ell_i(X) - 1$ is a degree $p - 1$ polynomial with p roots. This confirms that $\sum_{i=1}^{p} \ell_i = 1$.

In term of endomorphism we get $\sum_{i=1}^{p} \ell_i(f) = \operatorname{Id}$, which for any $x \in V$ leads to $\sum_{i=1}^{p} (\ell_i(f))(x) = x$.

This means that any $x \in V$ can be expressed in terms of $\ell_i(f) \in \ker(f - \lambda_i \operatorname{Id})$, or said otherwise, $V = \sum_{i=1}^{p} \ker(f - \lambda_i \operatorname{Id})$.

Our goal is to apply lemma 5.277 to prove that f is diagonalizable. So far we know that $V = \sum_{i=1}^{p} \ker(f - \lambda_i \operatorname{Id})$, but based on corollary 5.291, $\sigma_f \subset P^{-1}(\{0\})$. Besides from proposition 5.263, $\ker(f - \lambda \operatorname{Id}) = \{0\}$ if λ is not an eigenvalue. As a result,

$$\sum_{\lambda \in \sigma_f} E_\lambda(f) = \sum_{i=1}^{p} \ker(f - \lambda_i \operatorname{Id}) = V.$$

Applying lemma 5.277 completes the proof. □

Remarks.

- The previous result directly translates in terms of matrices.
- Based on the proof, f is diagonalizable if and only if f annihilates $\prod_{\lambda \in \sigma_f}(X - \lambda)$.

Examples.

- Any projector f is diagonalizable. This is clear since f annihilates $X(X-1)$ (example 5.288), which splits with simple roots.
- Any symmetry f is diagonalizable. This is clear since f annihilates $(X+1)(X-1)$ (example 5.288), which splits with simple roots.
- Let $A \in \mathcal{M}_{10}(\mathbb{R})$ such that $A^3 = 7A - 6\,\mathrm{I}_{10}$. Since $X^3 - 7X + 6 = (X-1)(X-2)(X+3)$, A can be diagonalized.

Remark. Let $A \in \mathcal{M}_n(\mathbb{K})$, and P be a non-zero annihilating polynomial for A. We can neither say that χ_A divides P, nor that P divides χ_A. For instance taking $A = \mathrm{I}_n$, we see that $\chi_A = (-1)^n(X-1)^n$, but for any $k \in \mathbb{N}^*$, $(X-1)^k$ is an annihilating polynomial for A.

We have the following very important result which, paraphrased, says that χ_A is an annihilating polynomial for A.

Theorem (Cayley-Hamilton)

For any $f \in \mathcal{L}(V)$, the characteristic polynomial of f, χ_f, is an annihilating polynomial for f, i.e. $\chi_f(f) = 0$.

Proof. Let $x \in V$, and $F_n = \{f^i(x) : 0 \leq i < n+1\} \subset V$. Since F_n has $n+1$ elements and V has dimension n, it means that F_n is linearly dependent. Thus there exits a maximal $p_x \in \mathbb{N}^*$, such that $F_{p_x} = \{f^i(x) : 0 \leq i < p_x\}$ is linearly independent but F_{p_x+1} is not.[2] In that case there exist $a_i \in \mathbb{K}$, $i \in [\![1, p_x - 1]\!]$, not all zero and such that

$$f^{p_x}(x) = \sum_{i=0}^{p_x-1} a_i f^i(x). \tag{5.1}$$

Let $W = \operatorname{span} F_{p_x}$. Since f^{p_x} can be written as a linear combination of f^i, $i \in [\![1, p_x - 1]\!]$, it is clear that W is closed.

[2] Note that p_x depends on x.

As f is defined over V, we consider its restriction g to $W \subset V$ and write the matrix B of g in the basis F_{p_x}. To construct it, simply observe that x is mapped to $f(x)$, $f(x)$ to $f^2(x)$, and so on, until $f(f_{p_x-1})$ is reached. This yields

$$B = \begin{pmatrix} 0 & & & a_0 \\ 1 & \ddots & 0 & \vdots \\ & \ddots & 0 & \\ 0 & & 1 & a_{p_x-1} \end{pmatrix} \text{ and } \chi_g(\lambda) = \begin{vmatrix} -\lambda & & & a_0 \\ 1 & \ddots & 0 & \vdots \\ & \ddots & -\lambda & \\ 0 & & 1 & a_{p_x-1} - \lambda \end{vmatrix}.$$

Then using proposition 5.269 and noting that $\det B = a_0$, we get

$$\chi_g(\lambda) = (-1)^{p_x} \left(\lambda^{p_x} - a_{p_x-1}\lambda^{p_x-1} - \cdots - a_0 \right) = 0,$$

based on equation (5.1).

Recalling proposition 5.269, we see that χ_g divides χ_f. Hence there exists $Q \in \mathbb{K}[X]$ such that $\chi_f = Q\chi_g$. It then suffices to apply proposition 5.284 to obtain

$$(\chi_f(f))(x) = Q(f)\left((\chi_g(f))(x)\right) = Q(f)(0) = 0.$$

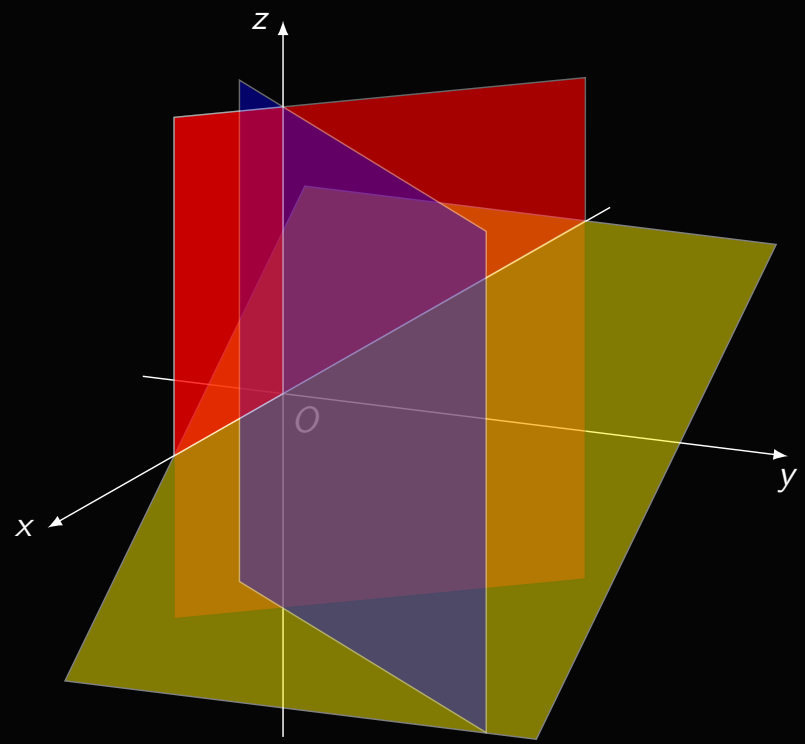
This shows that for any $x \in V$, $((\chi_f(f))(x) = 0$, or in other words $\chi_f(f) = 0$. □

Remarks.

- Cayley-Hamilton theorem easily translates in terms of matrices.
- The matrix B is often called *companion matrix*.
- Cayley-Hamilton theorem is very important as it provides a simple way to determine whether a matrix is diagonalizable or not. Once χ_f is computed it suffices to verify that it splits with simple roots.
- Knowing whether a matrix can be diagonalized and finding a "good basis" are fundamental to fields related to applied linear algebra.

Key points

- How important is the determinant? Why?
- What is the trace?
- Why are eigenvalues and eigenspaces very important in practice?
- How do diagonalizability and annihilating polynomials relate?
- What can be said about the characteristic polynomial?



Thank you, enjoy the Spring break!