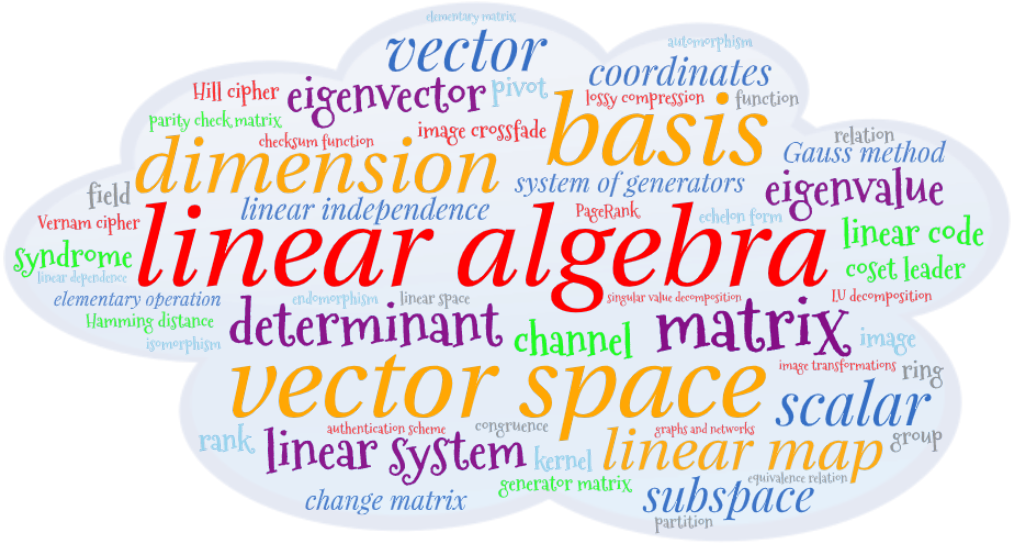
Septimiu Crivei

BASIC LINEAR ALGEBRA

Presa Universitară Clujeană

Septimiu Crivei

BASIC LINEAR ALGEBRA



Presa Universitară Clujeană

2022

Referenţi ştiinţifici:

Prof. univ. dr. Andrei Marcus

Conf. univ. dr. habil. Szántó Csaba

ISBN 978-606-37-1582-2

Universitatea Babeş-Bolyai
Presa Universitară Clujeană
Director: Codruţa Săcelean
Str. Hasdeu nr. 51
400371 Cluj-Napoca, România
Tel./fax: (+40)-264-597.401
E-mail: editura@ubbcluj.ro
http://www.editura.ubbcluj.ro/

To my family

Contents

Foreword

This textbook has grown up in the last twenty years, based on my courses of Algebra taught for first year students in Computer Science at the "Babeş-Bolyai" University of Cluj-Napoca. Besides the rather classical mathematical content, we also present some illustrations of the interplay between Algebra and Computer Science, which offers bidirectional applications. If we only restrict to Linear Algebra for Computer Science, then we should mention applications to several important research topics in Computer Science, such as Networks, Theory of Computation, Mathematics of Computing, Information Systems, Security and Privacy, Computing Methodologies and Applied Computing (see the corresponding addendum for some further details). The material is directed towards first year students in Computer Science, but it may be useful to anyone interested in an introduction to Linear Algebra and its applications to Computer Science.

The textbook is structured in four chapters, namely: *1. Preliminaries*, *2. Vector spaces*, *3. Matrices and linear systems* and *4. Introduction to coding theory*. Each chapter contains some extra material with related applications, and ends with a quiz with true or false questions and some projects to implement, which may help the reader to deepen the new concepts.

The first chapter is of a preliminary nature, setting the scene with some auxiliary needed notions, such as relations, functions, equivalence relations and partitions, algebraic structures with one and two operations as well as determinants. The core of the book consists of Chapters 2 and 3, which present the basics of Linear Algebra. In Chapter 2 we introduce and study vector spaces as algebraic generalizations of the vectors met in Physics. We also investigate subspaces, linear maps, quotient vector spaces, linear independence of vectors, bases and dimension of vector spaces. Chapter 3 reveals the strong connection between the abstract theory of vector spaces and the more practical concepts of matrices and linear systems. We study elementary operations and their applications, matrices of lists of vectors and linear maps, change of bases, linear systems and the Gauss method as well as eigenvectors and eigenvalues and some of their applications. The final chapter gives an introduction to coding theory, as a concrete application of the previously studied topics of Linear Algebra. We discuss the coding theory problem, Hamming distance, code representations, generator and parity check matrices as well as error-correcting and decoding.

The main content of the textbook is complemented by a series of addenda, namely *Historical notes* (on the concepts of matrix, determinant, vector space, basis and dimension, linear map, linear system, eigenvalue and eigenvector as well as on some scientists), *Computer Science topics using Linear Algebra* (according to the 2012 *Association for Computing Machinery* Classification System), *English-Romanian selected notions dictionary*, *Index of extra material* and *Index*.

Acknowledgements

I am grateful to Professors Iuliu Crivei and Gabriela Olteanu for inspiring conversations on the topics of this book and for their generous support.

I also thank my students throughout all these years, whose questions and interest have continuously motivated me to refine the taught material into its current form.

The author

Cluj–Napoca,
September 2022

Chapter 1

Preliminaries

In this chapter we briefly present some preliminary concepts and results that will be used in the main chapters of the book. We first discuss relations, and in particular, functions and equivalence relations. Then we study algebraic structures with one or two operations as well as substructures and special mappings between structures. We also briefly present some properties of determinants.

1.1 Relations

In this section we introduce relations, that may be seen, as far as a mathematician is concerned, as generalizations of functions. But numerous examples of relations are present in the daily life, even if we have not perceived them in the present form. Our goal is to formulate their algebraic definition and to see some relevant examples.

Definition 1.1.1 A triple $r = (A, B, R)$, where A, B are sets and

$$R \subseteq A \times B = \{(a, b) \mid a \in A, b \in B\},$$

is called a *(binary) relation*.
The set A is called the *domain*, the set B is called the *codomain* and the set R is called the *graph* of the relation r.
If $A = B$, then the relation r is called *homogeneous*.
If $(a, b) \in R$, then we sometimes write $a\, r\, b$ and we say that *a has the relation r to b* or *a and b are related with respect to the relation r*.

Definition 1.1.2 Let $r = (A, B, R)$ be a relation and let $X \subseteq A$. Then the set

$$r(X) = \{b \in B \mid \exists x \in X : x\, r\, b\}$$

is called the *relation class of X with respect to r*. If $x \in X$, then we denote

$$r < x >= r(\{x\}) = \{b \in B \mid x\, r\, b\}.$$

Remark 1.1.3 (1) Let $r = (A, B, R)$ be a relation and let $X \subseteq A$. Notice that

$$r(X) = \bigcup_{x \in X} r < x > .$$

(2) As in the case of functions, if $A, B \subseteq \mathbb{R}$, then the graph of a relation $r = (A, B, R)$ may be represented as a subset of points of the real plane $\mathbb{R} \times \mathbb{R}$, whereas if A, B are any finite sets, then $r = (A, B, R)$ may be represented by a diagram consisting of two sets with elements and connecting arrows. For instance, let $r = (A, B, R)$, where $A = \{1, 2, 3\}$, $B = \{1, 2\}$ and

$$R = \{(1, 1), (1, 2), (3, 1)\}.$$

One may draw the two sets A and B, and arrows between the elements related by R, namely arrows from 1 to 1, from 1 to 2 and from 3 to 1.

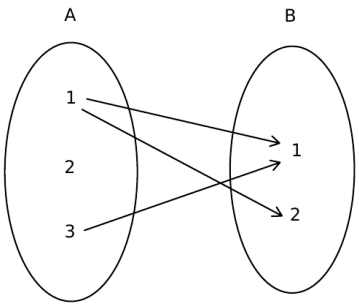


Figure 1.1: Diagram of a relation.

Also note that $r < 1 >= \{1, 2\} = r(A)$.

Example 1.1.4 (a) Let C be the set of all children and let P be the set of all parents. Then we may define the relation $r = (C, P, R)$, where

$$R = \{(c, p) \in C \times P \mid c \text{ is a child of } p\}.$$

(b) The triple $r = (\mathbb{R}, \mathbb{R}, R)$, where

$$R = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \leq y\}$$

is a homogeneous relation, called the *inequality relation* on $\mathbb{R}$. We have

$$r < 1 >= [1, \infty) = r([1, 2]).$$

(c) There are several examples from Number Theory, such as divisibility on $\mathbb{N}$ or on $\mathbb{Z}$, and Geometry, such as parallelism of lines, perpendicularity of lines, congruence of triangles, similarity of triangles.

(d) Let A and B be two sets. Then the triples

$$o = (A, B, \emptyset), \quad u = (A, B, A \times B)$$

are relations, called the *void relation* and the *universal relation* respectively.

(e) Let A be a set. Then the triple $\delta_A = (A, A, \Delta_A)$, where

$$\Delta_A = \{(a, a) \mid a \in A\}$$

is a relation called the *equality relation* on A.

(f) Every function is a relation. Indeed, a function $f : A \to B$ is determined by its domain A, its codomain B and its graph

$$G_f = \{(x, y) \in A \times B \mid y = f(x)\}.$$

Then the triple (A, B, G_f) is a relation.

(g) Every directed graph is a relation. Indeed, a directed graph (V, E) consists of a set V of vertices and a set E of directed edges ("arrows") between vertices. We may identify each directed edge with a pair in $V \times V$, where the first and the second component are respectively the starting and the ending vertex of that directed edge. Denote by P the set of those pairs. Then the triple (V, V, P) is a relation. For instance, the directed graph

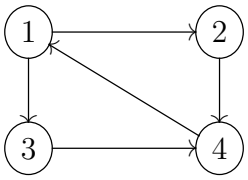


Figure 1.2: Directed graph.

may be seen as a relation (A, A, R), where $A = \{1, 2, 3, 4\}$ and

$$R = \{(1, 2), (1, 3), (2, 4), (3, 4), (4, 1)\}.$$

Extra: Relational database

Binary relations may be naturally generalized as follows.

Definition 1.1.5 A (finite) tuple

$$r = (A_1, \ldots, A_n, R),$$

where $A_1, \ldots, A_n$ are sets and

$$R \subseteq A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) \mid a_1 \in A_1, \ldots, a_n \in A_n\},$$

is called an *(n-ary) relation*. The sets $A_1, \ldots, A_n$ are called the *domains* of r, and the set R is called the *graph* of r. The number n is called the *degree (arity)* of r. A *relational database* is a (finite) set of relations.

Example 1.1.6 Consider the relation

$$student = (Integer, String, String, Integer, Student),$$

where

$$Student \subseteq Integer \times String \times String \times Integer$$

is given by the following table:

ID (Integer)	**Surname** (String)	**Name** (String)	**Grade** (Integer)
7	Ionescu	Alina	9
11	Ardelean	Cristina	10
23	Ionescu	Dan	7

Remark 1.1.7 Some known relational database management systems are:

- Oracle and RDB – Oracle
- SQL Server and Access - Microsoft

1.2 Functions

Definition 1.2.1 A relation $r = (A, B, R)$ is called a *function* if

$$\forall a \in A, \quad |r < a > | = 1,$$

that is, the relation class with respect to r of every $a \in A$ consists of exactly one element.

In other words, a relation r is a function if and only if every element of the domain has the relation r to exactly one element of the codomain.

In what follows, if $f = (A, B, F)$ is a function, we will mainly use the classical notation for a function, namely $f : A \to B$ or sometimes $A \xrightarrow{f} B$. The unique element of the set $f < a >$ will be denoted by $f(a)$. Then we have

$$(a, b) \in F \iff f(a) = b.$$

In particular, from Definition 1.1.1 for a relation, we get the following corresponding notions for a function.

Definition 1.2.2 Let $f : A \to B$ be a function. Then A is called the *domain*, B is called the *codomain* and

$$F = \{(a, f(a)) \mid a \in A\}$$

is called the *graph* of the function f.

Example 1.2.3 (a) Let A be a set. Then the equality relation (A, A, Δ_A) is a function called the *identity function (map) on* A, that is denoted by $1_A : A \to A$ and is defined by $1_A(a) = a$, $\forall a \in A$.

(b) Let B be a set and let $A \subseteq B$. Then the relation (A, B, Δ_A) is a function called the *inclusion function of* A *into* B, that is denoted by $i : A \to B$ and is defined by $i(a) = a$, $\forall a \in A$.

(c) Let $A = \{1, 2, 3\}$, $B = \{1, 2\}$ and let $r = (A, B, R)$, $s = (A, B, S)$, $t = (A, B, T)$ be the relations having the graphs

$$R = \{(1, 1), (2, 1), (3, 2)\},$$
$$S = \{(1, 2), (3, 1)\},$$
$$T = \{(1, 1), (1, 2), (2, 1), (3, 2)\}.$$

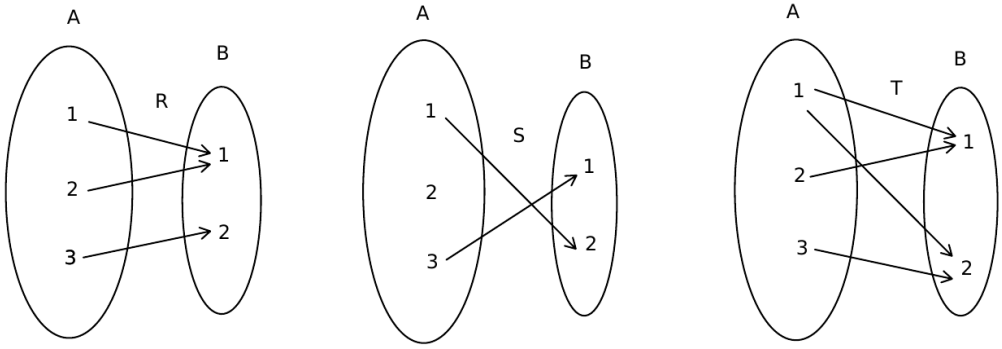


Figure 1.3: Diagrams of functions or relations.

Since $|r < a > | = 1$ for every $a \in A$, the relation r is a function. But s and t are not functions, because, for instance, we have $|s < 2 > | = 0$ and $|t < 1 > | = 2$.

Now we introduce a classical notation. Let A and B be two sets. Then we denote

$$B^A = \{f \mid f : A \to B \text{ is a function}\}.$$

If $|A| = n \in \mathbb{N}^*$, then the set B^A can be identified with the set $B^n = \underbrace{B \times \cdots \times B}_{n \text{ times}}$.

The notation is justified by the following nice result.

Theorem 1.2.4 *Let A and B be finite sets, say $|A| = n$ and $|B| = m$ $(m, n \in \mathbb{N}^*)$. Then*

$$|B^A| = m^n = |B|^{|A|}.$$

Proof. By induction on n. □

Definition 1.2.5 Let $f : A \to B$ be a function and let $X \subseteq A$.
We call the *image of X by f* the relation class of X with respect to f, that is,

$$f(X) = \{b \in B \mid \exists x \in X : x\, f\, b\} = \{f(x) \mid x \in X\}.$$

We denote $\mathrm{Im} f = f(A)$ and call it the *image of f*.

Definition 1.2.6 A function $f : A \to B$ is said to be:
(1) *injective* (or an *injection*) if

$$x_1, x_2 \in A\,,\ x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2)\,.$$

(2) *surjective* (or a *surjection*) if

$$\forall y \in B\,,\ \exists x \in A : f(x) = y\,.$$

(3) *bijective* (or a *bijection*) if f is both injective and surjective.

We recall the following well known characterizations of injective, surjective and bijective functions.

Lemma 1.2.7 *Let $f : A \to B$ be a function. The following conditions are equivalent to the injectivity of f:*
(i) $x_1, x_2 \in A$, $f(x_1) = f(x_2) \Longrightarrow x_1 = x_2$.
(ii) $\forall y \in B$, the equation $f(x) = y$ has at most one solution in A.
If $A, B \subseteq \mathbb{R}$, then we may add:
(iii) Every parallel to the Ox axis passing through a point of B intersects the graph of f in at most one point.

Lemma 1.2.8 *Let $f : A \to B$ be a function. The following conditions are equivalent to the surjectivity of f:*
(i) $f(A) = B$.
(ii) $\forall y \in B$, the equation $f(x) = y$ has at least one solution in A.
If $A, B \subseteq \mathbb{R}$, then we may add:
(iii) Every parallel to the Ox axis passing through a point of B intersects the graph of f in at least one point.

The reader may combine the above characterizations of injective and surjective functions in order to obtain ones of bijective functions.

Definition 1.2.9 Let $f : A \to B$ be a function. Then a function $g : B \to A$ is called an *inverse* of f if $g \circ f = 1_A$ and $f \circ g = 1_B$.

Lemma 1.2.10 *Let $f : A \to B$ be a function. Then f has an inverse if and only if f is bijective. In this case the inverse of f is unique and we denote it by f^{-1}.*

Example 1.2.11 (a) Let B be a set and let $A \subseteq B$. Then the inclusion function $i : A \to B$ is injective (see Example 1.2.3).

(b) The function $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = 2x$ is injective.

(c) Let $f : A \to B$ be a function. Then the function $f' : A \to \operatorname{Im} f$ defined by $f'(a) = f(a)$ is surjective.

(*d*) Let $f : \mathbb{R} \to \mathbb{R}_+$ be defined by $f(x) = x^2$, $\forall x \in \mathbb{R}$. Then f is surjective.

(*e*) Let A be a set. Then the identity function $1_A : A \to A$ is bijective.

(*f*) The function $f : \mathbb{R} \to (0, \infty)$ defined by $f(x) = e^x$ is bijective, and its inverse is the function $g : (0, \infty) \to \mathbb{R}$ defined by $g(x) = \ln(x)$.

1.3 Equivalence relations and partitions

Recall that a relation $r = (A, B, R)$ is called *homogeneous* if $A = B$. Some special type of such relations is the subject of the present section.

Definition 1.3.1 A homogeneous relation $r = (A, A, R)$ on A is called:
(1) *reflexive* (r) if: $\forall x \in A,\ x\, r\, x$.
(2) *transitive* (t) if: $x, y, z \in A,\ x\, r\, y$ and $y\, r\, z \Longrightarrow x\, r\, z$.
(3) *symmetric* (s) if: $x, y \in A,\ x\, r\, y \Longrightarrow y\, r\, x$.
A homogeneous relation $r = (A, A, R)$ is called an *equivalence relation* if r has the properties (r), (t) and (s).

Example 1.3.2 (*a*) The equality relation δ_A on a set A has all 3 properties, hence δ_A is an equivalence relation on A.

(*b*) The similarity of triangles is an equivalence relations on the set of all triangles.

(*c*) The inequality relation " $\leq$ " on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{R}$ has (r) and (t), but not (s). Hence it is not an equivalence relation on the corresponding set.

(*d*) Let $n \in \mathbb{N}$ and let ρ_n be the relation defined on $\mathbb{Z}$ by

$$x\, \rho_n\, y \Longleftrightarrow x \equiv y \pmod{n},$$

that is, $n|(x-y)$ or equivalently for $n \neq 0$, x and y give the same remainder when divided by n. Then ρ_n is called the *congruence modulo* n and it has the properties (r), (t) and (s), hence it is an equivalence relation.

For $n = 0$, we have $x\, \rho_0\, y \Longleftrightarrow 0|x - y \Longleftrightarrow x = y$, hence $\rho_0 = \delta_{\mathbb{Z}} = (\mathbb{Z}, \mathbb{Z}, \Delta_{\mathbb{Z}})$.

For $n = 1$, we have $x\, \rho_1\, y \Longleftrightarrow 1|x - y$, which is always true, and thus $\rho_1 = u = (\mathbb{Z}, \mathbb{Z}, \mathbb{Z} \times \mathbb{Z})$.

Definition 1.3.3 Let A be a non-empty set. Then a family $(A_i)_{i \in I}$ of non-empty subsets of A is called a *partition* of A if:
(i) The family $(A_i)_{i \in I}$ covers A, that is,

$$\bigcup_{i \in I} A_i = A.$$

(ii) The A_i's are pairwise disjoint, that is,

$$i, j \in I, i \neq j \Longrightarrow A_i \cap A_j = \emptyset.$$

Example 1.3.4 (*a*) Let $A = \{1, 2, 3, 4, 5\}$ and $A_1 = \{1, 2, 3\}$, $A_2 = \{4\}$, $A_3 = \{5\}$. Then $\{A_1, A_2, A_3\}$ is a partition of A.

(*b*) Let A be a set. Then $\{\{a\} \mid a \in A\}$ and $\{A\}$ are partitions of A.

(*c*) Let A_1 be the set of even integers and A_2 the set of odd integers. Then $\{A_1, A_2\}$ is a partition of $\mathbb{Z}$.

(*d*) Consider the intervals

$$A_n = [n, n + 1)$$

for every $n \in \mathbb{Z}$. Then the family $(A_n)_{n\in\mathbb{Z}}$ is a partition of $\mathbb{R}$.

Denote by $E(A)$ the set of all equivalence relations and by $P(A)$ the set of all partitions on a set A.

Definition 1.3.5 Let $r \in E(A)$.
The relation class $r < x >$ of an element $x \in A$ with respect to r is called the *equivalence class of x with respect to r*, while the element x is called a *representative* of $r < x >$.
The set

$$A/r = \{r < x > | \ x \in A\},$$

which is the set of all equivalence classes of elements of A with respect to r, is called the *quotient set of A by r*.

Definition 1.3.6 Let $\pi = (A_i)_{i\in I} \in P(A)$ and define the relation r_π on A by

$$x \, r_\pi \, y \iff \exists i \in I : x, y \in A_i \, .$$

Then r_π is called the *relation associated to the partition π*.

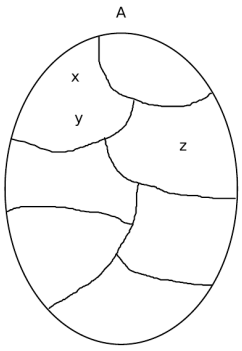


Figure 1.4: Relation associated to a partition.

The following theorem establishes the fundamental connection between equivalence relations and partitions.

Theorem 1.3.7 *(i) Let $r \in E(A)$. Then $A/r \in P(A)$.*
(ii) Let $\pi = (A_i)_{i\in I} \in P(A)$. Then $r_\pi \in E(A)$.
(iii) Let $F : E(A) \to P(A)$ be defined by

$$F(r) = A/r, \quad \forall r \in E(A).$$

Then F is a bijection, whose inverse is $G : P(A) \to E(A)$, defined by

$$G(\pi) = r_\pi, \quad \forall \pi \in P(A).$$

Proof. (i) Since r is reflexive, we have $x \in r < x >$, $\forall x \in A$, hence $A \subseteq \bigcup_{x\in A} r < x >$. The converse inclusion is obvious. Therefore, $A = \bigcup_{x\in A} r < x >$.

Now let $x, y \in A$ and $a \in r < x > \cap r < y >$. Then $x\, r\, a$ and $y\, r\, a$, hence $x\, r\, y$ and $y\, r\, x$ by the symmetry and the transitivity of r. Then $y \in r < x >$ and $x \in r < y >$, whence we get $r < x >= r < y >$. Hence the equivalence classes in A/r are pairwise disjoint. Therefore, $A/r \in P(A)$.

(ii) It is easy to see that the relation r_π is reflexive and symmetric. Now let $x, y, z \in A$ be such that $x\, r_\pi\, y$ and $y\, r_\pi z$. Then $\exists i, j \in I$ such that $x, y \in A_i$ and $y, z \in A_j$, hence $y \in A_i \cap A_j$, which implies that $A_i = A_j$. It follows that $x, z \in A_i$, hence $x\, r_\pi\, z$. Therefore, r_π is transitive and consequently $r_\pi \in E(A)$.

(iii) We show that $G \circ F = 1_{E(A)}$ and $F \circ G = 1_{P(A)}$.
For every $r \in E(A)$, we have

$$(G \circ F)(r) = G(A/r) = r\,,$$

because for every $x, y \in A$,

$$x\, G(A/r)\, y \iff \exists a \in A : x, y \in r < a > \iff x\, r\, y\,.$$

For every $\pi = (A_i)_{i\in I} \in P(A)$, we have

$$(F \circ G)(\pi) = A/G(\pi) = \{r_\pi < x > |\ x \in A\} = \pi\,,$$

because for every $x \in A$, the class of the partition $A/G(\pi)$ containing x is the same as the class of the partition π containing x. Indeed, let $x \in A_i$. Then

$$r_\pi < x >= \{y \in A \mid x\, r_\pi\, y\} = \{y \in A \mid y \in A_i\} = A_i\,.$$

Thus, $G \circ F = 1_{E(A)}$ and $F \circ G = 1_{P(A)}$, that is, F and G are bijections. □

Example 1.3.8 (a) Consider the set A of all first-year students in Computer Science, and its partition, say $\pi = \{A_1, \ldots, A_7\}$, where A_i denotes the set of all students in Group i for $i \in \{1, \ldots, 7\}$. Then the equivalence relation r_π on A corresponding to the partition π is defined as follows: student $x \in A$ has relation r_π to student $y \in A$ if and only if students x and y are in the same group i.

(b) Let $A = \{1, 2, 3\}$ and let r and s be the homogeneous relations defined on A with the graphs

$$R = \{(1,1), (2,2), (3,3), (1,2), (2,1)\}\,,$$
$$S = \{(1,1), (2,2), (3,3), (1,2), (2,3)\}\,.$$

Then r is an equivalence relation, but s is not. The partition corresponding to r is

$$A/r = \{\{1,2\},\{3\}\}.$$

(c) Consider the following families of sets:

$$\pi = \{\{1\},\{2,3\},\{4\}\},$$

$$\pi' = \{\{1,2\},\{2,3\},\{4\}\}.$$

Then π is a partition of $A = \{1,2,3,4\}$, but π' is not. The equivalence relation corresponding to π has the graph

$$R_\pi = \{(1,1),(2,2),(2,3),(3,2),(3,3),(4,4)\}.$$

(d) The congruence relation modulo n is an equivalence relation on $\mathbb{Z}$ and its corresponding partition is

$$\begin{aligned}\mathbb{Z}/\rho_n &= \{\rho_n < x > |\ x \in \mathbb{Z}\}\\ &= \{x + n\mathbb{Z} \mid x \in \mathbb{Z}\}\\ &= \{\widehat{x} \mid x \in \mathbb{Z}\},\end{aligned}$$

where an equivalence class is denoted by $\widehat{x}$. For $n \geq 2$, we denote

$$\mathbb{Z}_n = \mathbb{Z}/\rho_n = \{\widehat{0},\widehat{1},\ldots,\widehat{n-1}\}.$$

For $n = 0$ and $n = 1$, we have seen in Example 1.3.2 that $\rho_0 = \delta_{\mathbb{Z}}$ and $\rho_1 = u$, and we get

$$\begin{aligned}\mathbb{Z}/\rho_0 &= \{\{x\} \mid x \in \mathbb{Z}\},\\ \mathbb{Z}/\rho_1 &= \{\mathbb{Z}\},\end{aligned}$$

that are the two extreme partitions of $\mathbb{Z}$.

1.4 Operations

Definition 1.4.1 By an *operation* (or *composition law*) on a set A we understand a function

$$\varphi : A \times A \to A.$$

Usually, we denote operations by symbols like $\cdot$, $+$, $*$, so that $\varphi(x,y)$ is denoted by $x \cdot y$, $x + y$, $x * y$, $\forall (x,y) \in A \times A$. We denote by $(A,\cdot)$ the fact that "$\cdot$" is an operation on a set A.

Example 1.4.2 The usual addition and multiplication are operations on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and the usual subtraction is an operation on $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, but not on $\mathbb{N}$. The usual division is not an operation on either of the five numerical sets, because of the element zero.

Definition 1.4.3 Let " $\cdot$ " be an operation on an arbitrary set A. Define the following laws:
(1) *Associative law*:

$$(x \cdot y) \cdot z = x \cdot (y \cdot z)\,, \quad \forall x, y, z \in A.$$

(2) *Commutative law*:

$$x \cdot y = y \cdot x\,, \quad \forall x, y \in A.$$

(3) *Identity law*:

$$\exists e \in A \text{ such that } \forall a \in A,\ a \cdot e = e \cdot a = a\,.$$

In this case, e is called an *identity element*.
(4) *Inverse law*:

$$\forall a \in A, \exists a' \in A \text{ such that } a \cdot a' = a' \cdot a = e,$$

where e is the identity element. In this case, a' is called an *inverse element for* a.

Lemma 1.4.4 *Let " $\cdot$ " be an operation on a set A.*
(i) If there exists an identity element in A, then it is unique.
(ii) Assume further that the operation " $\cdot$ " is associative and has identity element e and let $a \in A$. If an inverse element for a does exist, then it is unique.

Proof. (i) Assume that $e_1, e_2 \in A$ are identity elements in A. Then by computing their product in two ways, we have $e_1 \cdot e_2 = e_1 = e_2$.

(ii) Suppose that a has $a_1, a_2 \in A$ as inverses. Then by the associative law, we may compute the product $a_1 \cdot a \cdot a_2$ in two ways as

$$a_1 \cdot a \cdot a_2 = a_1 \cdot (a \cdot a_2) = a_1 \cdot e = a_1,$$
$$a_1 \cdot a \cdot a_2 = (a_1 \cdot a) \cdot a_2 = e \cdot a_2 = a_2,$$

and we obtain $a_1 = a_2$. □

Let us now discuss some special subsets of sets endowed with an operation.

Definition 1.4.5 Consider an operation $\varphi : A \times A \to A$ on a set A and let $B \subseteq A$. Then B is called a *stable subset of A with respect to φ* (or *closed subset of A under the operation φ*) if

$$\forall x, y \in B\,, \quad \varphi(x, y) \in B\,.$$

In this case, we may consider the operation $\varphi' : B \times B \to B$ on B defined by

$$\varphi'(x, y) = \varphi(x, y)\,, \quad \forall (x, y) \in B \times B\,,$$

that is called the *operation induced by φ in the stable subset B of A*.
When using a symbol "$\cdot$" for φ, we simply say that *B is a stable subset of $(A, \cdot)$*.

Example 1.4.6 (a) The set

$$2\mathbb{Z} = \{2k \mid k \in \mathbb{Z}\}$$

of even integers is stable in $(\mathbb{Z}, +)$, but the set

$$2\mathbb{Z} + 1 = \{2k + 1 \mid k \in \mathbb{Z}\}$$

of odd integers is not stable in $(\mathbb{Z}, +)$.

(b) The interval $[0, 1]$ is stable in $(\mathbb{R}, \cdot)$, but the interval $[-1, 0]$ is not stable in $(\mathbb{R}, \cdot)$.

Remark 1.4.7 Notice that the associative, the commutative (and later on, the distributive laws) still hold in a stable subset (endowed with the induced operation), since they are true for every element in the initial set (only the universal quantifier $\forall$ appears in their definition). But the identity element and the inverse element do not transfer (their definition uses the existential quantifier $\exists$ as well).

1.5 Groups and rings

Definition 1.5.1 Let " $\cdot$ " be an operation on a set A. Then $(A, \cdot)$ is called a:
(1) *semigroup* if the associative law holds.
(2) *monoid* if it is a semigroup with identity element.
(3) *group* if it is a monoid in which every element has an inverse.

If the operation is commutative as well, then the structure is called *commutative*. A commutative group is also called an *abelian group* (after the name of N. H. Abel).

Remark 1.5.2 We denote by 1 the identity element of a group $(G, \cdot)$ and by x^{-1} the inverse of an element $x \in G$. In case of an additive group $(G, +)$, the identity element is denoted by 0, while the inverse of an element $x \in G$ is called the *symmetric* of x and is denoted by $-x$.

Definition 1.5.3 Let $(G, \cdot)$ be a semigroup, let $x \in G$ and let $n \in \mathbb{N}^*$. Then we may use the associative law and define

$$x^n = \underbrace{x \cdot x \cdot \ldots \cdot x}_{n \text{ times}}.$$

If $(G, \cdot)$ is a monoid, then we may also define

$$x^0 = 1\,.$$

If $(G, \cdot)$ is a group, then we may also define

$$x^{-n} = (x^{-1})^n\,.$$

Remark 1.5.4 If the operation is denoted by "+", then we replace the notation x^n by nx.

We may now give some standard properties of group computation.

Lemma 1.5.5 *Let $(G, \cdot)$ be a group, let $x \in G$ and let $m, n \in \mathbb{Z}$. Then:*
(i) $x^m \cdot x^n = x^{m+n}$.
(ii) $(x^m)^n = x^{mn}$.

Proof. This follows by induction on n for positive values, and then by using the definition. □

Lemma 1.5.6 *Let $(G, \cdot)$ be a group and let $a, x, y \in G$. Then:*
(i) $a \cdot x = a \cdot y \Longrightarrow x = y$,
$x \cdot a = y \cdot a \Longrightarrow x = y$ *(cancellation laws).*
(ii) $(x^{-1})^{-1} = x$.
(iii) $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$.

Proof. They are immediate by definitions. □

Remark 1.5.7 A finite group may be defined by its operation table, that specifies the result of any multiplication of two elements of the group. Using the cancellation laws, it is easy to see that the operation table of a group has the property that every element appears exactly once on each row and each column.

Example 1.5.8 (a) The operation "$-$" defined on $\mathbb{Z}$ is not associative.

(b) $(\mathbb{N}^*, +)$ is a semigroup, but not a monoid.

(c) $(\mathbb{N}, +)$, $(\mathbb{N}, \cdot)$, $(\mathbb{Z}, \cdot)$, $(\mathbb{Q}, \cdot)$, $(\mathbb{R}, \cdot)$, $(\mathbb{C}, \cdot)$ are monoids, but not groups.

(d) $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, $(\mathbb{Q}^*, \cdot)$, $(\mathbb{R}^*, \cdot)$ and $(\mathbb{C}^*, \cdot)$ are groups.

(e) Let X be a non-empty set. By a *word on* X of length n we understand a string of n elements from X for some $n \in \mathbb{N}$. The word of length 0 is called the *void word* and is denoted by e. On the set X^* of words on X consider the operation "$\cdot$" given by concatenation. Then $(X^*, \cdot)$ is a monoid with identity element e, called the *free monoid* on the set X.

(f) Let $\{e\}$ be a single element set and let "$\cdot$" be the only operation on $\{e\}$, defined by $e \cdot e = e$. Then $(\{e\}, \cdot)$ is an abelian group, called the *trivial group*.

(g) Let $n \in \mathbb{N}$, $n \geq 2$. Then $(\mathbb{Z}_n, +)$ is an abelian group, called the *group of residue classes modulo* n. The addition is defined by

$$\widehat{x} + \widehat{y} = \widehat{x + y}\,, \quad \forall \widehat{x}, \widehat{y} \in \mathbb{Z}_n\,.$$

(h) Let $n \in \mathbb{N}$ with $n \geq 2$. Denote by $M_{m,n}(\mathbb{R})$ the set of $m \times n$-matrices with entries in $\mathbb{R}$ and by $M_n(\mathbb{R})$ the set of $n \times n$-matrices with entries in $\mathbb{R}$. Then $(M_{m,n}(\mathbb{R}), +)$ is an abelian group and $(M_n(\mathbb{R}), \cdot)$ is a monoid.

Denote by

$$GL_n(\mathbb{R}) = \{A \in M_n(\mathbb{R}) \mid \det(A) \neq 0\}$$

the set of invertible $n \times n$-matrices with real entries. Then $(GL_n(\mathbb{R}), \cdot)$ is a group, called the *general linear group of rank* n.

(i) Let M be a set and let

$$S_M = \{f : M \to M \mid f \text{ is bijective}\}.$$

Then $(S_M, \circ)$ is a group, called the *symmetric group of* M. The identity element is the identity map 1_M and the inverse of an element f (which is a bijection) is the inverse function f^{-1}.

If $|M| = n$, then S_M is denoted by S_n, and the group $(S_n, \circ)$ is in fact the *permutation group* of n elements.

(j) Let $K = \{e, a, b, c\}$ and define an operation " $\cdot$ " on K by the following table:

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Then $(K, \cdot)$ is an abelian group, called *Klein's group*. It comes from Geometry, and it may be viewed as the group of geometric transformations of a rectangle:

- e is the identical transformation,
- a is the symmetry with respect to the horizontal symmetry axis of the rectangle,
- b is the symmetry with respect to the vertical symmetry axis of the rectangle,
- c is the symmetry with respect to the center of the circumscribed circle of the rectangle.

The product $x \cdot y$ of two transformations x and y of K is defined by performing first y and then x.

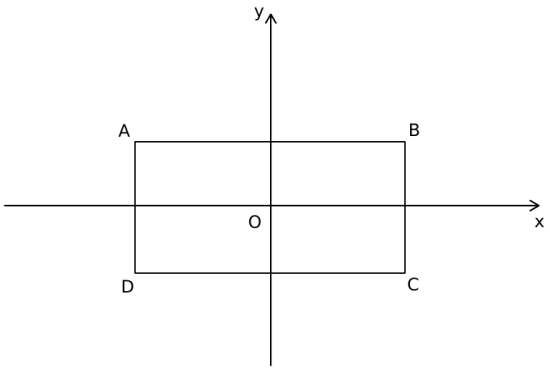


Figure 1.5: Klein's group.

(k) Let ABC be an equilateral triangle and consider the following geometric transformations, that transform the vertices A, B and C into themselves:

- e is the identical transformation (or the rotation counterclockwise of 0^0),
- α is the rotation counterclockwise of 120^0,
- β is the rotation counterclockwise through 240^0,
- a is the symmetry with respect to the axis d_1, passing through A and perpendicular to BC,
- b is the symmetry with respect to the axis d_2, passing through B and perpendicular to AC,

• c is the symmetry with respect to the axis d_3, passing through C and perpendicular to AB.

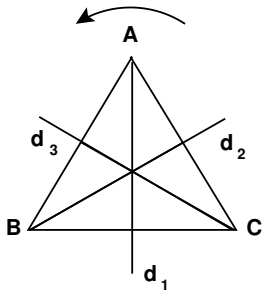


Figure 1.6: The 3rd dihedral group.

Denote $D_3 = \{e, \alpha, \beta, a, b, c\}$. Define the product $x \cdot y$ of two transformations x and y of D_3 by performing first y and then x. Then$(D_3, \cdot)$ is a group, called the 3^{rd} *dihedral group*.

Generalizing, for every $n \in \mathbb{N}$, $n \geq 3$, we can define the n^{th} *dihedral group* D_n of rotations and symmetries of a regular n-gon, consisting of n rotations and n symmetries. For instance, D_4 is the group of rotations and symmetries of a square, and it has 4 rotations (of angles 0^0, 90^0, 180^0 and 270^0) and 4 symmetries (with respect to the two diagonals of the square and the two perpendicular lines passing through the middle of the edges of the square).

Definition 1.5.9 Let R be a set. A structure with two operations $(R, +, \cdot)$ is called a:
(1) *ring* if $(R, +)$ is an abelian group, $(R, \cdot)$ is a semigroup and the *distributive laws* hold:

$$x \cdot (y + z) = x \cdot y + x \cdot z, \quad \forall x, y, z \in R,$$
$$(y + z) \cdot x = y \cdot x + z \cdot x, \quad \forall x, y, z \in R.$$

(2) *unitary ring* if $(R, +, \cdot)$ is a ring and there is an identity element with respect to " $\cdot$ ".
(3) *division ring* (or *skew field*) if $(R, +)$ is an abelian group, $(R^*, \cdot)$ is a group and the distributive laws hold.
(4) *field* if it is a commutative division ring.
The ring $(R, +, \cdot)$ is called *commutative* if the operation " $\cdot$ " is commutative.

If $(R, +, \cdot)$ is a ring, then we denote the identity elements with respect to " $+$ " and " $\cdot$ " by 0 and 1 respectively. We also use the notation $R^* = R \setminus \{0\}$.

Remark 1.5.10 (1) A ring $(R, +, \cdot)$ is a division ring if and only if $|R| \geq 2$ and any $x \in R^*$ has an inverse $x^{-1} \in R^*$.

(2) If $(R, +, \cdot)$ is a ring, then $(R, +)$ is a group and $(R, \cdot)$ is a semigroup, so that we may talk about multiples and positive powers of elements of R.

Definition 1.5.11 Let $(R,+,\cdot)$ be a ring, let $x \in R$ and let $n \in \mathbb{N}^*$. Then we define

$$\begin{aligned} n \cdot x &= \underbrace{x + x + \cdots + x}_{n \text{ times}}, \\ 0 \cdot x &= 0\,, \\ (-n) \cdot x &= -n \cdot x\,, \\ x^n &= \underbrace{x \cdot x \cdot \ldots \cdot x}_{n \text{ times}}. \end{aligned}$$

If R is a unitary ring, then we may also consider

$$x^0 = 1\,.$$

If R is a division ring, then we may also define negative powers of x, namely

$$x^{-n} = (x^{-1})^n\,.$$

Remark 1.5.12 Notice that in the definition $0 \cdot x = 0$, the first 0 is the integer zero and the second 0 is the zero element of the ring R, that is, the identity element of the group $(R,+)$.

Clearly, the first computational properties of a ring $(R,+,\cdot)$ are the properties of the group $(R,+)$ and of the semigroup $(R,\cdot)$. Some relationship properties between the two operations are given in the following result, in which all zeros are the zero element of the ring R.

Lemma 1.5.13 *Let $(R,+,\cdot)$ be a ring and let $x,y,z \in R$. Then:*
(i) $x \cdot (y-z) = x \cdot y - x \cdot z$.
$(y-z) \cdot x = y \cdot x - z \cdot x$.
(ii) $x \cdot 0 = 0 \cdot x = 0$.
(iii) $x \cdot (-y) = (-x) \cdot y = -x \cdot y$.

Proof. (i) We have

$$x \cdot (y-z) = x \cdot y - x \cdot z \iff x \cdot (y-z) + x \cdot z = x \cdot y \iff x \cdot (y - z + z) = x \cdot y\,,$$

the last equality being obviously true. Similarly, $(y-z) \cdot x = y \cdot x - z \cdot x$.

(ii) We have

$$x \cdot 0 = x \cdot (y-y) = x \cdot y - x \cdot y = 0\,.$$

Similarly, $0 \cdot x = 0$.

(iii) We have

$$x \cdot (-y) = -x \cdot y \iff x \cdot (-y) + x \cdot y = 0 \iff x \cdot (-y+y) = 0 \iff x \cdot 0 = 0\,,$$

the last equality being true by (ii). □

Example 1.5.14 (*a*) $(\mathbb{Z}, +, \cdot)$ is a unitary ring, but not a field.

(*b*) $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$ are fields.

(*c*) Let $\{e\}$ be a single element set and let both " $+$ " and " $\cdot$ " be the only operation on $\{e\}$, defined by $e + e = e$ and $e \cdot e = e$. Then $(\{e\}, +, \cdot)$ is a commutative unitary ring, called the *trivial ring*.

(*d*) Let $n \in \mathbb{N}$, $n \geq 2$. Then $(\mathbb{Z}_n, +, \cdot)$ is a commutative unitary ring, called the *ring of residue classes modulo* n. The addition and the multiplication are defined by

$$\widehat{x} + \widehat{y} = \widehat{x + y}\,, \quad \widehat{x} \cdot \widehat{y} = \widehat{x \cdot y}\,, \quad \forall \widehat{x}, \widehat{y} \in \mathbb{Z}_n.$$

Note that $(\mathbb{Z}_n, +, \cdot)$ is a field if and only if n is prime.

(*e*) Let $(R, +, \cdot)$ be a commutative unitary ring. Then $(R[X], +, \cdot)$ is a commutative unitary ring, called the *polynomial ring over* R *in the indeterminate* X, where the operations are the usual addition and multiplication of polynomials.

(*f*) Let $n \in \mathbb{N}$, $n \geq 2$ and let $(R, +, \cdot)$ be a ring. Then $(M_n(R), +, \cdot)$ is a ring, called the *ring of matrices* $n \times n$ *with entries in* R, where the operations are the usual addition and multiplication of matrices.

(*g*) Let M be a non-empty set and let $(R, +, \cdot)$ be a ring. Define on the set

$$R^M = \{f \mid f : M \to R\}$$

two operations by: $\forall f, g \in R^M$, we have $f + g : M \to R$, $f \cdot g : M \to R$, where

$$(f + g)(x) = f(x) + g(x)\,, \quad \forall x \in M\,,$$

$$(f \cdot g)(x) = f(x) \cdot g(x)\,, \quad \forall x \in M\,.$$

Then $(R^M, +, \cdot)$ is a ring, called the *ring of functions with a set as domain and a ring as codomain*. The zero element is

$$\theta : M \to R, \quad \theta(x) = 0, \quad \forall x \in M.$$

The symmetric of any $f : M \to R$ is

$$-f : M \to R, \quad (-f)(x) = -f(x), \quad \forall x \in M.$$

(*h*) A ring $(R, +, \cdot)$ is called *Boolean* (after the name of G. Boole) if $a^2 = a$ for every $a \in R$. If M is a set and $\mathcal{P}(M)$ is the power set of M (that is, the set of all subsets of M), then $(\mathcal{P}(M), \Delta, \cap)$ is a Boolean ring, where Δ is the *symmetric difference* operation defined by

$$A \Delta B = (A \setminus B) \cup (B \setminus A)$$

for every $A, B \in \mathcal{P}(M)$.

Extra: Fast adding

We describe a method for fast adding large natural numbers, following [14].

Remark 1.5.15 If a and b are two natural numbers, then it makes no difference if we add them as natural numbers or as elements (that is, residue classes) of some group $(\mathbb{Z}_n, +)$ for some $n > a + b$.

Theorem 1.5.16 (Chinese Remainder Theorem) *If $n = p_1^{r_1} \cdots p_k^{r_k}$ for some distinct primes $p_1, \ldots, p_k$, then there is an isomorphism of additive groups:*

$$\mathbb{Z}_n \simeq \mathbb{Z}_{p_1^{r_1}} \times \cdots \times \mathbb{Z}_{p_k^{r_k}}$$

given by

$$\varphi : \mathbb{Z}_n \to \mathbb{Z}_{p_1^{r_1}} \times \cdots \times \mathbb{Z}_{p_k^{r_k}}, \quad \varphi([x]_n) = ([x]_{p_1^{r_1}}, \ldots, [x]_{p_k^{r_k}}),$$

where $[x]_m$ denotes the residue class of x modulo $m \in \mathbb{N}$.
If we denote $n_i = p_i^{r_i}$, $N_i = \frac{n}{n_i}$ and $K_i = [N_i^{-1}]_{n_i}$ for every $i \in \{1, \ldots, k\}$, then the inverse of φ is given by

$$\psi : \mathbb{Z}_{p_1^{r_1}} \times \cdots \times \mathbb{Z}_{p_k^{r_k}} \to \mathbb{Z}_n, \quad \psi(a_1, \ldots, a_k) = \left[\sum_{i=1}^{k} a_i N_i K_i \right]_n .$$

This allows one (the computer) to replace the addition of large natural numbers by parallel "small" simultaneous additions. This technique is used in the design of computer software in order to speed up calculations.

Example 1.5.17 Let $a = 37$, $b = 56$, and choose $n = 140 = 2^2 \cdot 5 \cdot 7$.

$$\begin{aligned} a = 37 \to [37]_{140} \to ([37]_4, [37]_5, [37]_7) &= ([1]_4, [2]_5, [2]_7) \quad + \\ b = 56 \to [56]_{140} \to ([56]_4, [56]_5, [56]_7) &= ([0]_4, [1]_5, [0]_7) \\ a + b \qquad\qquad\qquad\qquad\quad &= ([1]_4, [3]_5, [2]_7) \end{aligned}$$

Now one solves the following system by the *Chinese Remainder Theorem*:

$$\begin{cases} x = 1 \pmod 4 \\ x = 3 \pmod 5 \\ x = 2 \pmod 7 \end{cases} .$$

We have:

$$\begin{aligned} n_1 = 4, n_2 = 5, n_3 = 7, n = n_1 \cdot n_2 \cdot n_3 = 140, \\ N_1 = \frac{n}{n_1} = 35, N_2 = \frac{n}{n_2} = 28, N_3 = \frac{n}{n_3} = 20. \end{aligned}$$

Note that $K_i = [N_i^{-1}]_{n_i}$ means that $N_i K_i = 1 \pmod{n_i}$. Hence we have:

$$\begin{aligned} K_1 = N_1^{-1} \bmod n_1 = 35^{-1} \bmod 4 = 3^{-1} \bmod 4 = 3, \\ K_2 = N_2^{-1} \bmod n_2 = 28^{-1} \bmod 5 = 3^{-1} \bmod 5 = 7, \\ K_3 = N_3^{-1} \bmod n_3 = 20^{-1} \bmod 7 = 6^{-1} \bmod 7 = 6. \end{aligned}$$

Finally, we get the solution

$$x = a_1 N_1 K_1 + a_2 N_2 K_2 + a_3 N_3 K_3 = 93$$

(unique solution modulo n). Hence $a + b = 93$.

1.6 Subgroups and subrings

We turn now our attention to the study of a group or ring inside another group or ring respectively. Recall that the associative and the commutative laws transfer in a stable subset, while the identity element and an inverse element do not transfer in general.

Definition 1.6.1 Let $(G, \cdot)$ be a group and let $H \subseteq G$. Then H is called a *subgroup* of G if:
(i) H is a stable subset of $(G, \cdot)$.
(ii) $(H, \cdot)$ is a group.

We denote by $H \leq G$ the fact that H is a subgroup of a group G.

The next two characterization theorems give more efficient ways to check that a subset of a group is a subgroup.

Theorem 1.6.2 *Let $(G, \cdot)$ be a group and let $H \subseteq G$. Then*

$$H \leq G \iff \begin{cases} H \neq \emptyset \ (1 \in H) \\ \forall x, y \in H, \ x \cdot y \in H \\ \forall x \in H, \ x^{-1} \in H. \end{cases}$$

Proof. $\boxed{\Longrightarrow}$ Suppose that $H \leq G$. Since $(H, \cdot)$ is a group, there exists an identity element $1' \in H$, hence $H \neq \emptyset$. Moreover, we have

$$x \cdot 1' = 1' \cdot x = x, \quad \forall x \in H.$$

By multiplying by x^{-1}, we get $1' = 1 \in H$. Therefore, a subgroup must contain the identity element of the group.

Since H is a stable subset of $(G, \cdot)$, for every $x, y \in H$, we have $x \cdot y \in H$.

Now let $x \in H$ and denote by x' its inverse in the group $(H, \cdot)$. Then

$$x \cdot x' = x' \cdot x = 1.$$

But $x \in H \subseteq G$ has an inverse $x^{-1} \in G$. Then by multiplying by x^{-1}, we get $x' = x^{-1} \in H$.

$\boxed{\Longleftarrow}$ Suppose that the three conditions hold. Then H is a stable subset of $(G, \cdot)$. Clearly, the associative law holds also in H. Take $x \in H \neq \emptyset$. Then $x^{-1} \in H$ and $1 = x \cdot x^{-1} \in H$. Hence 1 is the identity element in H, and every element of H has an inverse in H. Hence $(H, \cdot)$ is a group. □

Theorem 1.6.3 *Let $(G, \cdot)$ be a group and let $H \subseteq G$. Then*

$$H \leq G \iff \begin{cases} H \neq \emptyset \ (1 \in H) \\ \forall x, y \in H, \ x \cdot y^{-1} \in H. \end{cases}$$

Proof. This is immediate by definition and Theorem 1.6.2. □

Remark 1.6.4 (1) In case of an additive group $(G,+)$, the last two conditions in Theorem 1.6.2 become:
- $\forall x,y\in H,\ x+y\in H$.
- $\forall x\in H,\ -x\in H$.

(2) In case of an additive group $(G,+)$, the last condition in Theorem 1.6.3 becomes:
- $\forall x,y\in H,\ x-y\in H$.

Let us now see some examples of subgroups.

Example 1.6.5 (a) Every non-trivial group $(G,\cdot)$ has two subgroups, namely $\{1\}$ and G, called the *trivial subgroups*.

(b) $\mathbb{Z}$ is a subgroup of $(\mathbb{Q},+)$, $(\mathbb{R},+)$ and $(\mathbb{C},+)$, $\mathbb{Q}$ is a subgroup of $(\mathbb{R},+)$ and $(\mathbb{C},+)$, $\mathbb{R}$ is a subgroup of $(\mathbb{C},+)$.

(c) The set

$$n\mathbb{Z}=\{nx\mid x\in\mathbb{Z}\}$$

is a subgroup of $(\mathbb{Z},+)$ for every $n\in\mathbb{N}$.

(d) The set

$$H=\{z\in\mathbb{C}\mid |z|=1\}$$

is a subgroup of the group $(\mathbb{C}^*,\cdot)$, called the *circle group*. But it is not a subgroup of the group $(\mathbb{C},+)$.

(e) The set

$$U_n=\{z\in\mathbb{C}\mid z^n=1\}\quad (n\in\mathbb{N}^*)$$

is a subgroup of the group $(\mathbb{C}^*,\cdot)$, called the *group of* n^{th} *roots of unity*. Its elements are the following:

$$\varepsilon_k=\cos\frac{2k\pi}{n}+i\sin\frac{2k\pi}{n},\quad k\in\{0,\dots,n-1\}.$$

(f) Consider the general linear group $(GL_n(\mathbb{R}),\cdot)$ of rank n, where

$$GL_n(\mathbb{R})=\{A\in M_n(\mathbb{R})\mid \det(A)\neq 0\}$$

$(n\in\mathbb{N},\ n\geq 2)$ and denote

$$SL_n(\mathbb{R})=\{A\in M_n(\mathbb{R})\mid \det(A)=1\}.$$

Then $SL_n(\mathbb{R})$ is a subgroup of $(GL_n(\mathbb{R}),\cdot)$, called the *special linear group of rank* n.

Definition 1.6.6 Let $(R,+,\cdot)$ be a ring and let $A\subseteq R$. Then A is called a *subring* of R if:
(i) A is a stable subset of $(R,+,\cdot)$.
(ii) $(A,+,\cdot)$ is a ring.

Definition 1.6.7 Let $(K,+,\cdot)$ be a field and let $A\subseteq K$. Then A is called a *subfield* of K if:
(i) A is a stable subset of $(K,+,\cdot)$.
(ii) $(A,+,\cdot)$ is a field.

We denote by $A \leq R$ ($A \leq K$) the fact that A is a subring (subfield) of a ring R (field K).

In practice, one checks that a subset of a ring (field) is a subring (subfield) by using one of the next two characterization theorems.

Theorem 1.6.8 *Let $(R, +, \cdot)$ be a ring and let $A \subseteq R$. Then*

$$A \text{ is a subring of } R \iff \begin{cases} A \neq \emptyset \ (0 \in A) \\ \forall x, y \in A, \ x - y \in A \\ \forall x, y \in A, \ x \cdot y \in A. \end{cases}$$

Proof. $\boxed{\Longrightarrow}$ Assume that A is a subring of $(R, +, \cdot)$. Since $(A, +)$ is a group, we have $0 \in A \neq \emptyset$. But A is a stable subset of $(R, +)$ and $(R, +)$ is group, hence A is a subgroup of $(R, +)$. By Theorem 1.6.3, we have $x - y \in A$, $\forall x, y \in A$. Since A is a stable subset of $(R, \cdot)$, $\forall x, y \in A$, we have $x \cdot y \in A$.

$\boxed{\Longleftarrow}$ Assume that the three conditions hold. By the first two of them and Theorem 1.6.3, $(A, +)$ is a subgroup of $(R, +)$, and consequently a stable subset of $(R, +)$. The last condition tells us that A is a stable subset of $(R, \cdot)$. Now all needed properties for $(A, +, \cdot)$ to be a ring follow easily. □

Theorem 1.6.9 *Let $(K, +, \cdot)$ be a field and let $A \subseteq K$. Then*

$$A \text{ is a subfield of } K \iff \begin{cases} |A| \geq 2 \ (0, 1 \in A) \\ \forall x, y \in A, \ x - y \in A \\ \forall x, y \in A \text{ with } y \neq 0, \ x \cdot y^{-1} \in A. \end{cases}$$

Proof. $\boxed{\Longrightarrow}$ Assume that A is a subfield of $(K, +, \cdot)$. Since $(A, +)$ is a group, we have $0 \in A \neq \emptyset$. Since $(A, +, \cdot)$ is a field, $(A^*, \cdot)$ is a group, and thus $1 \in A$, and consequently $|A| \geq 2$. But A is a stable subset of $(K, +)$ and $(A, +)$ is group, hence A is a subgroup of $(K, +)$. By Theorem 1.6.3 we have $x - y \in A$, $\forall x, y \in A$. Since $(A^*, \cdot)$ is a subgroup of the group $(K^*, \cdot)$, by Theorem 1.6.2 we have $y^{-1} \in A$, $\forall y \in A^*$. But A is also a stable subset of $(K, \cdot)$, hence $\forall x, y \in A$ with $y \neq 0$, we have $x \cdot y^{-1} \in A$.

$\boxed{\Longleftarrow}$ Assume that the three conditions hold. By the first two of them and Theorem 1.6.3, $(A, +)$ is a subgroup of $(K, +)$ and consequently a stable subset of $(K, +)$. Now let $x, y \in A$. If $y = 0$, then $x \cdot y = 0 \in A$ by the first condition. If $y \neq 0$, then $x \cdot y = x \cdot (y^{-1})^{-1} \in A$ by the last condition. Hence A is a stable subset of $(K, \cdot)$. Now all needed properties for $(A, +, \cdot)$ to be a field follow easily. □

Let us now see some examples of subrings and subfields.

Example 1.6.10 (a) Every non-trivial ring $(R, +, \cdot)$ has two subrings, namely $\{0\}$ and R, called the *trivial subrings*.

(b) $\mathbb{Z}$ is a subring of $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$.

(c) $\mathbb{Q}$ is a subfield of $(\mathbb{R}, +, \cdot)$ and $(\mathbb{C}, +, \cdot)$, while $\mathbb{R}$ is a subfield of $(\mathbb{C}, +, \cdot)$.

(*d*) The set

$$n\mathbb{Z} = \{nx \mid x \in \mathbb{Z}\}$$

is a subring of $(\mathbb{Z}, +, \cdot)$ for every $n \in \mathbb{N}$. Note that $n\mathbb{Z}$ does not have identity for $n \geq 2$.

(*e*) The set

$$\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$$

is a subring of the field $(\mathbb{C}, +, \cdot)$, but not a subfield. It is called the *ring of Gauss integers*.

1.7 Group and ring homomorphisms

Let us now define some special maps between groups or rings. For the sake of simplicity, we denote by the same symbol operations in different arbitrary structures.

Definition 1.7.1 Let $(G, \cdot)$ and $(G', \cdot)$ be groups and let $f : G \to G'$. Then f is called a *group homomorphism* if

$$f(x \cdot y) = f(x) \cdot f(y)\,, \quad \forall x, y \in G\,.$$

Also, f is called a *group isomorphism* if it is a bijective group homomorphism.

We denote by $G \simeq G'$ the fact that two groups G and G' are isomorphic.
Usually, we denote by 1 and $1'$ the identity elements in G and G' respectively.

Example 1.7.2 (*a*) Let $(G, \cdot)$ and $(G', \cdot)$ be groups and let $f : G \to G'$ be defined by $f(x) = 1'$, $\forall x \in G$. Then f is a homomorphism, called the *trivial group homomorphism*.

(*b*) Let $(G, \cdot)$ be a group. Then the identity map $1_G : G \to G$ is a group isomorphism.

(*c*) Let $n \in \mathbb{N}$ and let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by $f(x) = nx$. Then f is a group homomorphism from the group $(\mathbb{Z}, +)$ to itself.

(*d*) Let $n \in \mathbb{N}$ with $n \geq 2$. The map $f : \mathbb{Z} \to \mathbb{Z}_n$ defined by $f(x) = \widehat{x}$ is a group homomorphism between the groups $(\mathbb{Z}, +)$ and $(\mathbb{Z}_n, +)$.

(*e*) Let $f : \mathbb{C}^* \to \mathbb{R}^*$ be defined by $f(z) = |z|$. Then f is a group homomorphism between $(\mathbb{C}^*, \cdot)$ and $(\mathbb{R}^*, \cdot)$. But $f : \mathbb{C} \to \mathbb{R}$ defined by $f(z) = |z|$ is not a group homomorphism between the groups $(\mathbb{C}, +)$ and $(\mathbb{R}, +)$.

(*f*) Let $n \in \mathbb{N}$, $n \geq 2$ and let $f : GL_n(\mathbb{R}) \to \mathbb{R}^*$ be defined by $f(A) = \det(A)$. Then f is a group homomorphism between the groups $(GL_n(\mathbb{R}), \cdot)$ and $(\mathbb{R}^*, \cdot)$.

Theorem 1.7.3 *Let $f : G \to G'$ be a group homomorphism. Then:*
(i) $f(1) = 1'$.
(ii) $(f(x))^{-1} = f(x^{-1})$, $\forall x \in G$.

Proof. (*i*) For every $x \in G$, we have $1 \cdot x = x \cdot 1 = x$, so that

$$f(1 \cdot x) = f(x \cdot 1) = f(x).$$

Since f is a group homomorphism, it follows that

$$f(1) \cdot f(x) = f(x) \cdot f(1) = f(x)\,,$$

whence we get $f(1) = 1'$ by multiplying by $(f(x))^{-1}$.

(ii) Let $x \in G$. Since $x \cdot x^{-1} = x^{-1} \cdot x = 1$, f is a group homomorphism and $f(1) = 1'$, it follows that

$$f(x) \cdot f(x^{-1}) = f(x^{-1}) \cdot f(x) = 1'\,.$$

Hence $(f(x))^{-1} = f(x^{-1})$. □

Definition 1.7.4 Let $(R, +, \cdot)$ and $(R', +, \cdot)$ be rings and $f : R \to R'$. Then f is called a *ring homomorphism* if $\forall x, y \in R$ we have

$$f(x + y) = f(x) + f(y)\,,$$
$$f(x \cdot y) = f(x) \cdot f(y)\,.$$

Also, f is called a *ring isomorphism* if it is a bijective ring homomorphism.

We denote by $R \simeq R'$ the fact that two rings R and R' are isomorphic.

Example 1.7.5 (a) Let $(R, +, \cdot)$ and $(R', +, \cdot)$ be rings and let $f : R \to R'$ be defined by $f(x) = 0'$, $\forall x \in R$. Then f is a ring homomorphism, called the *trivial ring homomorphism.*

(b) Let $(R, +, \cdot)$ be a ring. Then the identity map $1_R : R \to R$ is a ring isomorphism.

(c) Let $n \in \mathbb{N}$ with $n \geq 2$. The map $f : \mathbb{Z} \to \mathbb{Z}_n$ defined by $f(x) = \widehat{x}$ is a ring homomorphism between the rings $(\mathbb{Z}, +, \cdot)$ and $(\mathbb{Z}_n, +, \cdot)$.

(d) The map $f : \mathbb{C} \to \mathbb{R}$ defined by $f(z) = |z|$ is not a ring (field) homomorphism between the fields $(\mathbb{C}, +, \cdot)$ and $(\mathbb{R}, +, \cdot)$.

(e) The map $f : \mathbb{R} \to M_2(\mathbb{R})$ defined by $f(x) = \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}$, $\forall x \in \mathbb{R}$, is a ring homomorphim between the rings $(\mathbb{R}, +, \cdot)$ and $(M_2(\mathbb{R}), +, \cdot)$.

(f) Let $n \in \mathbb{N}$, $n \geq 2$ and let $f : M_n(\mathbb{R}) \to \mathbb{R}$ be defined by $f(A) = \det(A)$. Then f is not a ring homomorphism between the rings $(M_n(\mathbb{R}), +, \cdot)$ and $(\mathbb{R}, +, \cdot)$.

Remark 1.7.6 If $f : R \to R'$ is a ring homomorphism, then the first condition from its definition tells us that f is a group homomorphism between $(R, +)$ and $(R', +)$. Then f takes the identity element of $(R, +)$ to the identity element of $(R', +)$, that is, $f(0) = 0'$ and we also have $f(-x) = -f(x)$, $\forall x \in R$. But in general, even if R and R' have identities, denoted by 1 and $1'$ respectively, in general it does not follow that a ring homomorphism $f : R \to R'$ has the property that $f(1) = 1'$.

Definition 1.7.7 Let $(R, +, \cdot)$ and $(R', +, \cdot)$ be rings with identity elements 1 and $1'$ respectively, and let $f : R \to R'$ be a ring homomorphism. Then f is called *unitary* if $f(1) = 1'$.

Theorem 1.7.8 *Let $(R,+,\cdot)$ and $(R',+,\cdot)$ be unitary rings with identity elements 1 and $1'$ respectively, and let $f:R\to R'$ be a ring homomorphism.*
(i) If f is surjective, then f is unitary.
(ii) If f is a ring isomorphism, then f is unitary.
(iii) If f is unitary and $x\in R$ has an inverse element $x^{-1}\in R$, then $f(x)$ has an inverse and

$$(f(x))^{-1}=f(x^{-1})\,.$$

Proof. (i) Let $x'\in R'$. Then $\exists x\in R$ such that $f(x)=x'$, because f is surjective. Then we have

$$x'\cdot f(1)=f(x)\cdot f(1)=f(x\cdot 1)=f(x)=x'\,,$$
$$f(1)\cdot x'=f(1)\cdot f(x)=f(1\cdot x)=f(x)=x'\,,$$

hence $f(1)=1'$ is the identity element of R'.

(ii) This follows by (i).

(iii) We have

$$x\cdot x^{-1}=x^{-1}\cdot x=1\Longrightarrow f(x\cdot x^{-1})=f(x^{-1}\cdot x)=f(1)\Longrightarrow$$
$$\Longrightarrow f(x)\cdot f(x^{-1})=f(x^{-1})\cdot f(x)=1'\,,$$

whence it follows that $(f(x))^{-1}=f(x^{-1})$. □

1.8 Determinants

Throughout this section K will be a field and $m,n\in\mathbb{N}$ with $m,n\geq 2$. Denote by $M_{m,n}(K)$ the set of $m\times n$-matrices with entries in K, and by $M_n(K)$ the set of $n\times n$-matrices with entries in K.

Let us first give the following definition.

Definition 1.8.1 Let $\sigma\in S_n$ and let $i,j\in\{1,\dots,n\}$ with $i\neq j$. We say that (i,j) is an *inversion* of the permutation σ if $i<j$ and $\sigma(i)>\sigma(j)$. We denote by $\mathrm{inv}(\sigma)$ the number of inversions of σ, and we define the *signature* of σ by

$$\mathrm{sgn}(\sigma)=(-1)^{\mathrm{inv}(\sigma)}.$$

The concept of determinant with entries in K is defined similarly as the determinant of a matrix with real or complex entries.

Definition 1.8.2 The *determinant of order n* is defined as the function $\det: M_n(K)\to K$ given by

$$\det(A)=\sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)a_{1\sigma(1)}\cdots a_{n\sigma(n)},$$

where $A=(a_{ij})\in M_n(K)$.

Example 1.8.3 (*a*) The determinant of order 2 is the function $\det : M_2(K) \to K$ given by

$$\det(A) = a_{11}a_{22} - a_{21}a_{21},$$

where $A = (a_{ij}) \in M_2(K)$.

(*b*) The determinant of order 3 is the function $\det : M_3(K) \to K$ given by

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33},$$

where $A = (a_{ij}) \in M_3(K)$.

We will use the following concepts.

Definition 1.8.4 Let $A = (a_{ij}) \in M_{m,n}(K)$. Denote $A^T = (a_{ji}^T) \in M_{n,m}(K)$, where $a_{ji}^T = a_{ij}$ for every $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$. Then A^T is called the *transpose* of A.

Definition 1.8.5 Let $A = (a_{ij}) \in M_n(K)$. For $i, j \in \{1, \ldots, n\}$, let $A_{ij} \in M_{n-1}(K)$ be the matrix obtained from A by deleting the row i and the column j, and denote

$$c_{ij} = (-1)^{i+j} \det(A_{ij}).$$

Then c_{ij} is called the *cofactor* of a_{ij}, and the matrix $C = (c_{ij}) \in M_n(K)$ is called the *cofactor matrix* of A.

We list some basic properties of determinants, whose proofs are similar to those for determinants of matrices with real or complex entries.

Theorem 1.8.6 (Laplace Theorem) *Let $A = (a_{ij}) \in M_n(K)$, and let $C = (c_{ij}) \in M_n(K)$ be the cofactor matrix of A.*
(i) For every $i \in \{1, \ldots, n\}$, we have the following cofactor expansion along the row i of A:

$$\det(A) = \sum_{j=1}^{n} a_{ij}c_{ij}.$$

(ii) For every $j \in \{1, \ldots, n\}$, we have the following cofactor expansion along the column j of A:

$$\det(A) = \sum_{i=1}^{n} a_{ij}c_{ij}.$$

Example 1.8.7 Consider the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \in M_3(\mathbb{R}).$$

By using Laplace's Theorem with the cofactor expansion along the first row of A we have:

$$\det(A) = 0 \cdot (-1)^{1+1} \cdot \begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} + 1 \cdot (-1)^{1+2} \cdot \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} + 1 \cdot (-1)^{1+3} \cdot \begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1.$$

Remark 1.8.8 In practice, determinants of large matrices may be needed. Their computation by using definition or the cofactor expansion by using Laplace's Theorem is computationally expensive, or even infeasible. In general, the cofactor expansion requires more than $n!$ multiplications. For instance, for a 25×25-matrix, it would need more than 25! multiplications, which is approximately $1.5 \cdot 10^{25}$. If a computer performs 10^{12} multiplications per second, then it would need more than 500000 years to compute that determinant.

The following properties make the computation of determinants much easier.

Theorem 1.8.9 *Let $A, B \in M_n(K)$.*
(i) If A has two equal rows (columns), then $\det(A) = 0$.
(ii) If two rows (columns) of A are interchanged, then the determinant of the resulting matrix is equal to $-\det(A)$.
(iii) If one multiplies a row (column) of A by $\alpha \in K^*$, then the determinant of the resulting matrix is equal to $\alpha \det(A)$.
(iv) If one multiplies a row (column) by an element of K and adds the result to another row (column), then the determinant of the matrix does not change.
(v) The determinant of an upper (lower) triangular matrix*, that is, a matrix all of whose elements under (above) its principal diagonal are zero, is the product of the elements of its principal diagonal.*
(vi) $\det(A) = \det(A^T)$.
(vii) $\det(A \cdot B) = \det(A) \cdot \det(B)$.
(viii) $\det(\alpha A) = \alpha^n \det(A)$ *for every* $\alpha \in K$.

Example 1.8.10 Consider the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \in M_3(\mathbb{R}) .$$

By using Theorem 1.8.9 we may successively interchange the first two rows, multiply the first row by -1 and add it to the third row, multiply the second row by -1 and add it to the third row, and we obtain:

$$\det(A) = -\begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{vmatrix} = -\begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{vmatrix} = -\begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{vmatrix} = 1.$$

Now consider the same matrix with entries in $\mathbb{Z}_2$, that is,

$$A = \begin{pmatrix} \widehat{0} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{0} & \widehat{1} \end{pmatrix} \in M_3(\mathbb{Z}_2) .$$

The above computations, considered now in $\mathbb{Z}_2$, give $\det(A) = \widehat{1}$. Note that in $\mathbb{Z}_2$ we have $\widehat{1} + \widehat{1} = \widehat{0}$, hence $\widehat{-1} = \widehat{1}$.

Remark 1.8.11 The computation of a determinant of an $n \times n$-matrix by using row operations needs about $\frac{2}{3}n^3$ arithmetic operations. For instance, for a 25×25-matrix it would need about 10000 operations, which may be calculated by a modern computer in a fraction of a second.

Next we present some properties of invertible square matrices and of the rank of a matrix, which are introduced as for matrices with real or complex entries, and have similar proofs.

Definition 1.8.12 Let $A \in M_n(K)$ and let $C = (c_{ij}) \in M_n(K)$ be the cofactor matrix of A. Then the matrix $A^* = C^T$ is called the *adjugate* (or *adjoint*) matrix of A.

Theorem 1.8.13 *Let $A \in M_n(K)$. Then:*
(i) $A \cdot A^* = A^* \cdot A = \det(A) \cdot I_n$.
(ii) A is invertible if and only if $\det(A) \neq 0$.
(iii) If A is invertible, then $A^{-1} = (\det(A))^{-1} \cdot A^*$ *and* $\det(A^{-1}) = (\det(A))^{-1}$.

Example 1.8.14 Consider the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \in M_3(\mathbb{R}) .$$

Then $\det(A) = 1 \neq 0$, hence A is invertible.

The cofactor matrix of A is

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

It follows that

$$A^{-1} = (\det(A))^{-1} \cdot A^* = A^* = C^T = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} .$$

Definition 1.8.15 Let $A = (a_{ij}) \in M_{m,n}(K)$.
A *minor of order r of the matrix A* is the determinant of any $r \times r$-square matrix obtained by crossing out rows and columns of A.
The *rank of the matrix A* is denoted by $\operatorname{rank}(A)$ and is defined as the maximum order of the non-zero minors of A.

Theorem 1.8.16 *Let $A \in M_{m,n}(K)$. Then:*
(i) $\operatorname{rank}(A) \leq \min(m, n)$.
(ii) $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.
(iii) For $m = n$, A is invertible if and only if $\operatorname{rank}(A) = n$.

Example 1.8.17 Consider the matrix

$$A = \begin{pmatrix} -3 & 5 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -3 \end{pmatrix} \in M_{3,4}(\mathbb{R}) .$$

All minors of A order 3 are zero, and there is a non-zero minor of A of order 2, say $\begin{vmatrix} -1 & 1 \\ 1 & 1 \end{vmatrix}$, hence $\operatorname{rank}(A) = 2$.

Remark 1.8.18 In case of matrices of large size, which may appear in practice, the computations of the inverse of an invertible square matrix and of the rank of a matrix by using determinants is computationally expensive. In Chapter 3 we will see more efficient methods to compute them by using elementary operations.

Chapter 1 quiz

Decide whether the following statements are **true** or **false**.

1. Every symmetric homogeneous relation is reflexive.

2. The intersection of all subsets of a partition of a set is equal to the empty set.

3. A partition of a set A is a subset of A.

4. The number of partitions of a finite set A is the same as the number of equivalence relations on A.

5. Every group is a semigroup.

6. Let $(G, \cdot)$ be a group that is not abelian. Then $x \cdot y \neq y \cdot x$ for all $x, y \in G$.

7. For every elements x, y of a group $(G, \cdot)$ we have

$$(x \cdot y)^{-1} = x^{-1} \cdot y^{-1}.$$

8. The identity element in a subgroup H of a group G must be the same as the identity element of G.

9. An element x of a subgroup H of a group $(G, \cdot)$ has an inverse x^{-1} in H that may be different than its inverse in G.

10. Any subgroup of an abelian group is abelian.

11. If H is a stable subset of a group $(G, \cdot)$, then H is a subgroup of $(G, \cdot)$.

12. There exists at least one group homomorphism from any group G to any group G'.

13. Any group homomorphism is a group isomorphism.

14. Every ring is a group with respect to each of the operations of addition and multiplication.

15. If $(R, +, \cdot)$ is a non-commutative ring, then $(R, +)$ is a non-commutative group.

16. In any commutative ring $(R, +, \cdot)$, for every $x, y \in R$ we have

$$x^2 - y^2 = (x + y)(x - y).$$

17. For every field $(K, +, \cdot)$, K is an abelian group with respect to each of the operations of addition and multiplication.

18. Any field is a ring with identity.

19. Any field is a division ring.

20. If A is a subring of a ring R with identity, then A is a ring with identity.

21. Every subfield of a field has an identity element with respect to multiplication.

22. A ring homomorphism between rings R and R' is a group homomorphism from the additive group R to the additive group R'.

23. A ring homomorphism between two rings with identity takes the identity element of the domain into the identity element of the codomain.

24. Let $n \in \mathbb{N}$ with $n \geq 2$ and let $A, B \in M_n(K)$. Then

$$\det(A + B) = \det(A) + \det(B).$$

25. Let $m, n \in \mathbb{N}$ with $m, n \geq 2$ and let $A \in M_{m,n}(K)$. Then $\operatorname{rank}(A) = \max(m, n)$.

Chapter 1 projects

Use a programming language of your choice and implement the following projects.

Project 1.1

- *Input*: non-zero natural number n
- *Output*:
  1. the number of partitions on a set $A = \{a_1, \ldots, a_n\}$
  2. the partitions on a set $A = \{a_1, \ldots, a_n\}$ and the graphs of their corresponding equivalence relations (for $n \leq 8$)

Example:

- *Input*: $n = 3$
- *Output*:

1. the number of partitions on a set $A = \{a_1, a_2, a_3\}$ is 5
2. using the notation $\Delta_A = \{(a_1, a_1), (a_2, a_2), (a_3, a_3)\}$, the partitions on a set $A = \{a_1, a_2, a_3\}$ and the graphs of their corresponding equivalence relations are:

$$\begin{aligned}
\{a_1\}, \{a_2\}, \{a_3\}\} &\rightsquigarrow \Delta_A \\
\{a_1, a_2\}, \{a_3\}\} &\rightsquigarrow \Delta_A \cup \{(a_1, a_2), (a_2, a_1)\} \\
\{a_1, a_3\}, \{a_2\}\} &\rightsquigarrow \Delta_A \cup \{(a_1, a_3), (a_3, a_1)\} \\
\{a_2, a_3\}, \{a_1\}\} &\rightsquigarrow \Delta_A \cup \{(a_2, a_3), (a_3, a_2)\} \\
\{\{a_1, a_2, a_3\}\} &\rightsquigarrow A \times A
\end{aligned}$$

Project 1.2

- *Input*: non-zero natural number n
- *Output*:
  1. the number of transitive relations on a set $A = \{a_1, \ldots, a_n\}$
  2. the graphs of the transitive relations on a set $A = \{a_1, \ldots, a_n\}$ (for $n \leq 4$)

Example:

- *Input*: $n = 2$
- *Output*:
  1. the number of transitive relations on a set $A = \{a_1, a_2\}$ is 13
  2. the graphs of the transitive relations on a set $A = \{a_1, a_2\}$ are:

$$\begin{aligned}
R_1 &= \emptyset \\
R_2 &= \{(a_1, a_1)\} \\
R_3 &= \{(a_1, a_2)\} \\
R_4 &= \{(a_2, a_1)\} \\
R_5 &= \{(a_2, a_2)\} \\
R_6 &= \{(a_1, a_1), (a_1, a_2)\} \\
R_7 &= \{(a_1, a_1), (a_2, a_1)\} \\
R_8 &= \{(a_1, a_1), (a_2, a_2)\} \\
R_9 &= \{(a_1, a_2), (a_2, a_2)\} \\
R_{10} &= \{(a_2, a_1), (a_2, a_2)\} \\
R_{11} &= \{(a_1, a_1), (a_2, a_2), (a_1, a_2)\} \\
R_{12} &= \{(a_1, a_1), (a_2, a_2), (a_2, a_1)\} \\
R_{13} &= \{(a_1, a_1), (a_1, a_2), (a_2, a_1), (a_2, a_2)\}
\end{aligned}$$

Project 1.3

- *Input*: non-zero natural number n
- *Output*:
  1. the number of associative operations on a set $A = \{a_1, \ldots, a_n\}$
  2. the operation table of each associative operation (for $n \leq 4$)

Example:

- *Input*: $n = 2$
- *Output*:
  1. the number of associative operations on a set $A = \{a_1, a_2\}$ is 8
  2. identifying an operation table

	a_1	a_2
a_1	x	y
a_2	z	t

by the matrix $\begin{pmatrix} x & y \\ z & t \end{pmatrix} \in M_2(A)$, the operation tables of the associative operations on $A = \{a_1, a_2\}$ are given by the matrices:

$$\begin{pmatrix} a_1 & a_1 \\ a_1 & a_1 \end{pmatrix}, \begin{pmatrix} a_1 & a_1 \\ a_1 & a_2 \end{pmatrix}, \begin{pmatrix} a_1 & a_1 \\ a_2 & a_2 \end{pmatrix}, \begin{pmatrix} a_1 & a_2 \\ a_1 & a_2 \end{pmatrix},$$

$$\begin{pmatrix} a_1 & a_2 \\ a_2 & a_1 \end{pmatrix}, \begin{pmatrix} a_1 & a_2 \\ a_2 & a_2 \end{pmatrix}, \begin{pmatrix} a_2 & a_1 \\ a_1 & a_2 \end{pmatrix}, \begin{pmatrix} a_2 & a_2 \\ a_2 & a_2 \end{pmatrix}.$$

Project 1.4

- *Input*: non-zero natural number n
- *Output*:
  1. the number of abelian group structures which can be defined on a set $A = \{a_1, \ldots, a_n\}$
  2. the operation table of each such abelian group (for $n \leq 7$)

Example: The operation table of a group G has the property that each element of G appears exactly once on each row and on each column. The operation table of an abelian group is symmetric with respect to the main diagonal. We may identify an operation table by a matrix. Make sure that the operations are associative and have identity element.

- *Input*: $n = 4$

- *Output*:

  1. the number of abelian group structures on a set $G = \{a_1, a_2, a_3, a_4\}$ is 16
  2. the abelian group structures on G with identity element a_1 are given by the matrices:

$$\begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} \\ \mathbf{a_2} & \boxed{a_1} & a_4 & a_3 \\ \mathbf{a_3} & a_4 & a_1 & a_2 \\ \mathbf{a_4} & a_3 & a_2 & a_1 \end{pmatrix}, \begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} \\ \mathbf{a_2} & \boxed{a_1} & a_4 & a_3 \\ \mathbf{a_3} & a_4 & a_2 & a_1 \\ \mathbf{a_4} & a_3 & a_1 & a_2 \end{pmatrix},$$

$$\begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} \\ \mathbf{a_2} & \boxed{a_3} & a_4 & a_1 \\ \mathbf{a_3} & a_4 & a_1 & a_2 \\ \mathbf{a_4} & a_1 & a_2 & a_3 \end{pmatrix}, \begin{pmatrix} \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} \\ \mathbf{a_2} & \boxed{a_4} & a_1 & a_3 \\ \mathbf{a_3} & a_1 & a_4 & a_2 \\ \mathbf{a_4} & a_3 & a_2 & a_1 \end{pmatrix}.$$

     There are 4 similar abelian group structures for each possible identity element.

Chapter 2

Vector spaces

This chapter deals with vector spaces (also called linear spaces), that are algebraic structures having also an "external operation" beside a usual operation, as it was the case in the previous chapter. They are the bricks of Linear Algebra and have numerous applications in different branches of Mathematics, but also in Physics, Computer Science and in other fields of Natural Sciences. Some of their algebraic applications will be studied in the next chapter, dedicated to the study of matrices and linear systems of equations.

Throughout the present chapter K will always denote a field.

2.1 Basic properties

The reader surely remembers the notion of a vector, as an object met in elementary Physics, characterized by an origin, a direction, a sense and a length. We might wonder if the notion of a vector space yet to be defined is somehow connected to or generalizes the classical notion of a vector used in Physics, namely an object described by direction, sense, starting point and length. The answer is positive and we are going to see that in a forthcoming example.

But let us begin with the definition of the key notion of a vector space.

Definition 2.1.1 A *vector space over* K (or a K*-vector space*) is an abelian group $(V, +)$ together with a so-called *external operation* or *scalar multiplication*

$$\cdot : K \times V \to V\,, \quad (k, v) \mapsto k \cdot v \quad (\text{or simply } kv)\,,$$

satisfying the following axioms:
(L_1) $k \cdot (v_1 + v_2) = k \cdot v_1 + k \cdot v_2$;
(L_2) $(k_1 + k_2) \cdot v = k_1 \cdot v + k_2 \cdot v$;
(L_3) $(k_1 \cdot k_2) \cdot v = k_1 \cdot (k_2 \cdot v)$;
(L_4) $1 \cdot v = v$,
for every $k, k_1, k_2 \in K$ and every $v, v_1, v_2 \in V$.
In this context, the elements of K are called *scalars* and the elements of V are called *vectors*.
Sometimes a vector space is also called a *linear space*.

We usually denote a vector space V over K by ${}_KV$, which emphasizes the fact that vectors are multiplied by scalars on the left hand side. Sometimes, we also use the notation $(V, K, +, \cdot)$.

Remark 2.1.2 (1) Notice that in the definition of a vector space there are present four operations, two denoted by the same symbol "+" and two denoted by the same symbol "$\cdot$". Of course, they are not the same, but as we have already done it several times before, we use the convention to denote them identically for the sake of simplicity of writing. There are 3 operations in the classical sense, namely the addition and the multiplication in the field K and the addition in the group V and, on the other hand, there is also an external operation of multiplication by scalars.

(2) The axioms (L_1) and (L_2) look like some distributive laws and the axiom (L_3) looks like an associative law, but they are not, since the involved elements are not taken from the same set.

(3) The definition we have just given is that of a *left vector space*. It is also possible to give the definition of a *right vector space* by considering an external operation

$$\cdot : V \times K \to V\,, \quad (v, k) \mapsto v \cdot k,$$

satisfying some similar axioms, but on the right hand side.

Since one can show that there is a bijection between the left and the right vector spaces of the field K, we are going to study only the left vector spaces and omit the adjective "left".

Let us now see several important examples of vector spaces.

Example 2.1.3 (a) Let V_2 be the set of all vectors (in the classical sense) in the plane with a fixed origin O. Then V_2 is a vector space over $\mathbb{R}$ (or a *real vector space*), where the addition is the usual addition of two vectors by the parallelogram rule and the external operation is the usual scalar multiplication of vectors by real scalars.

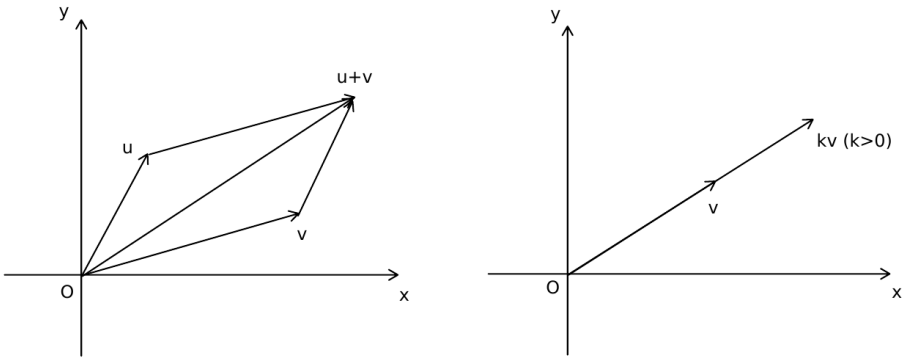


Figure 2.1: Vector addition and scalar multiplication.

If we consider two coordinate axes Ox and Oy in the plane, each vector in V_2 is perfectly determined by the coordinates of its ending point. Therefore, the addition of vectors and the scalar multiplication of vectors by real numbers become:

$$(x, y) + (x', y') = (x + x', y + y')\,,$$
$$k \cdot (x, y) = (k \cdot x, k \cdot y)\,,$$

$\forall k \in \mathbb{R}$ and $\forall (x,y),(x',y') \in \mathbb{R}\times\mathbb{R}$. Thus, $(\mathbb{R}^2,\mathbb{R},+,\cdot)$ is a vector space.

Similarly, one can consider the real vector space V_3 of all vectors in the space with a fixed origin. Moreover, a further, but more algebraical, generalization is possible, as we may see in the following example.

(*b*) Let $n \in \mathbb{N}^*$. Define

$$(x_1,\ldots,x_n)+(y_1,\ldots,y_n)=(x_1+y_1,\ldots,x_n+y_n)\,,$$
$$k\cdot(x_1,\ldots,x_n)=(kx_1,\ldots,kx_n)\,,$$

$\forall (x_1,\ldots,x_n),(y_1,\ldots,y_n)\in K^n$ and $\forall k\in K$. Then $(K^n,K,+,\cdot)$ is a vector space, called the *canonical vector space* (or *standard vector space*) over K.

Let us discuss some particular cases. For $K=\mathbb{Z}_2$, $\mathbb{Z}_2^n$ is a vector space over $\mathbb{Z}_2$. For $n=1$, we get that ${}_KK$ is a vector space. Hence, as far as the classical numerical fields are concerned, ${}_{\mathbb{Q}}\mathbb{Q}$, ${}_{\mathbb{R}}\mathbb{R}$ and ${}_{\mathbb{C}}\mathbb{C}$ are vector spaces.

(*c*) If $V=\{e\}$ is a single element set, then we know that there is a unique structure of an abelian group for V, namely that one defined by $e+e=e$. Then we can define a unique scalar multiplication, namely $k\cdot e=e$, $\forall k\in K$. Thus, V is a vector space, called the *zero (null) vector space* and denoted by $\{0\}$.

(*d*) If A is a subfield of the field K, then K is a vector space over A, where the addition and the scalar multiplication are just the addition and the multiplication of elements in the field K.

In particular, ${}_{\mathbb{Q}}\mathbb{R}$, ${}_{\mathbb{Q}}\mathbb{C}$ and ${}_{\mathbb{R}}\mathbb{C}$ are vector spaces. Note that $\mathbb{R}$ may be viewed as a vector space over $\mathbb{Q}$ or $\mathbb{R}$, while $\mathbb{C}$ may be viewed as a vector space over any of the fields $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$.

(*e*) $(K[X],K,+,\cdot)$ is a vector space, where the addition is the usual addition of polynomials and the scalar multiplication is defined as follows: $\forall f=a_0+a_1X+\cdots+a_nX^n\in K[X]$, $\forall k\in K$,

$$kf=(ka_0)+(ka_1)X+\cdots+(ka_n)X^n\,.$$

(*f*) Let $m,n\in\mathbb{N}$, $m,n\geq 2$. Then $(M_{m,n}(K),K,+,\cdot)$ is a vector space, where the operations are the usual addition and scalar multiplication of matrices.

(*g*) Let A be a non-empty set. Denote

$$K^A=\{f\mid f:A\to K\}\,.$$

Then $(K^A,K,+,\cdot)$ is a vector space, where the addition and the scalar multiplication are defined as follows: $\forall f,g\in K^A$, $\forall k\in K$, we have $f+g\in K^A$, $kf\in K^A$, where

$$(f+g)(x)=f(x)+g(x)\,,$$
$$(kf)(x)=kf(x)$$

$\forall x\in A$. As a particular case, we obtain the vector space $(\mathbb{R}^{\mathbb{R}},\mathbb{R},+,\cdot)$ of real functions of a real variable.

(*h*) Let V and V' be K-vector spaces. Then the cartesian product $V\times V'$ is a K-vector space, called the *direct product* of V and V', where the addition and the scalar

multiplication are defined as follows:

$$(v_1, v_1') + (v_2, v_2') = (v_1 + v_1', v_2 + v_2'),$$
$$k(v_1, v_1') = (kv_1, kv_1')$$

$\forall (v_1, v_1'), (v_2, v_2') \in V \times V'$ and $\forall k \in K$.

(i) We have seen that $V = K \times K$ has a canonical structure of vector space over K. Let us now see what happens if we change the addition or the scalar multiplication.

Let us first define them as follows:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + 2y_2),$$
$$k \cdot (x_1, y_1) = (kx_1, ky_1)$$

$\forall (x_1, y_1), (x_2, y_2) \in V$ and $\forall k \in K$. Then V is still a vector space over K, with a different structure of vector space than the canonical one.

Now let us define them as follows:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2),$$
$$k \cdot (x_1, y_1) = (kx_1, y_1)$$

$\forall (x_1, y_1), (x_2, y_2) \in V$ and $\forall k \in K$. In general, they do not define a structure of vector space for V over K, because the axiom (L_2) does not hold. For instance, for $K = \mathbb{R}$, we have

$$(1+2) \cdot (3,4) = 3 \cdot (3,4) = (9,4) \neq (9,8) = (3,4) + (6,4) = 1 \cdot (3,4) + 2 \cdot (3,4).$$

Let us now state some computation rules in a vector space. Notice that we denote by 0 both the zero scalar and the zero vector.

Theorem 2.1.4 *Let V be a vector space over K. Then $\forall k, k' \in K$ and $\forall v, v' \in V$ we have:*
(i) $k \cdot 0 = 0 \cdot v = 0$.
(ii) $k(-v) = (-k)v = -kv$.
(iii) $k(v - v') = kv - kv'$.
(iv) $(k - k')v = kv - k'v$.

Proof. (i) We have:

$$k \cdot 0 + k \cdot v = k(0 + v) = kv \Longrightarrow k \cdot 0 = 0,$$

$$0 \cdot v + k \cdot v = (0 + k)v = kv \Longrightarrow 0 \cdot v = 0.$$

(ii) We have:

$$kv + k(-v) = k(v - v) = k \cdot 0 = 0 \Longrightarrow k(-v) = -kv,$$

$$kv + (-k)v = (k - k)v = 0 \cdot v = 0 \Longrightarrow (-k)v = -kv.$$

(iii) We have:

$$k(v-v')+kv'=k(v-v'+v')=kv \Longrightarrow k(v-v')=kv-kv'.$$

(iv) We have:

$$(k-k')v+k'v=(k-k'+k')v=kv \Longrightarrow (k-k')v=kv-k'v.$$

Hence all the above properties are true. □

Theorem 2.1.5 *Let V be a vector space over K and let $k \in K$ and $v \in V$. Then*

$$kv=0 \Longleftrightarrow k=0 \text{ or } v=0\,.$$

Proof. $\boxed{\Longrightarrow}$ Assume that $kv=0$. Suppose that $k \neq 0$. Then k is invertible in the field K and we have

$$kv=0 \Longrightarrow kv=k\cdot 0 \Longrightarrow k^{-1}(kv)=k^{-1}(k\cdot 0) \Longrightarrow (k^{-1}k)v=(k^{-1}k)\cdot 0 \Longrightarrow v=0\,.$$

$\boxed{\Longleftarrow}$ This is Theorem 2.1.4 (i). □

Extra: Vernam cipher

Following [11], we describe an easy, but secure cipher. Let $n \in \mathbb{N}^*$ and consider the canonical vector space $V=\mathbb{Z}_2^n$ over $\mathbb{Z}_2$. The vectors of V may be identified with n-bit binary strings. Suppose that Alice needs to send an n-bit plaintext $p \in \mathbb{Z}_2^n$ to Bob.

Vernam cipher:

1. (*Key establishment*) Alice and Bob randomly choose a vector $k \in \mathbb{Z}_2^n$ as a key.
2. (*Encryption*) Alice computes the ciphertext c according to the formula
$$c=p+k,$$
where the sum is a vector in $\mathbb{Z}_2^n$.
3. (*Decryption*) Bob computes the plaintext p according to the formula
$$p=c-k=c+k,$$
where the sum is a vector in $\mathbb{Z}_2^n$.

Remark 2.1.6 The system satisfies perfect secrecy, but the key k must be distributed in advance.

Example 2.1.7 Alice wants to send the message

$$p=(0,0,0,1,1,1,0,1,0,1) \in \mathbb{Z}_2^{10}$$

to Bob.
Alice and Bob agree on the vector

$$k=(0,1,1,0,1,0,0,0,0,1) \in \mathbb{Z}_2^{10}$$

as a key.
Alice encrypts the message by computing the ciphertext c as:

$$c = p+k = (0,0,0,1,1,1,0,1,0,1)+(0,1,1,0,1,0,0,0,0,1) = (0,1,1,1,0,1,0,1,0,0) \in \mathbb{Z}_2^{10}.$$

Bob decrypts the message by computing the plaintext p as:

$$p = c+k = (0,1,1,1,0,1,0,1,0,0)+(0,1,1,0,1,0,0,0,0,1) = (0,0,0,1,1,1,0,1,0,1) \in \mathbb{Z}_2^{10}.$$

2.2 Subspaces

Let us now discuss some special subsets of vector spaces, namely *subspaces*. We are going to define a subspace in the same general way as we did for subgroups or subrings.

Definition 2.2.1 Let V be a vector space over K and let $S \subseteq V$. Then S is a *subspace* of V if:
(i) $S \neq \emptyset$.
(ii) $\forall v_1, v_2 \in S$, $v_1 + v_2 \in S$.
(iii) $\forall k \in K$, $\forall v \in S$, $kv \in S$.

We usually denote by $S \leq_K V$, or simply by $S \leq V$, the fact that S is a subspace of the vector space V over K.

Remark 2.2.2 Notice that every subspace S of a vector space V over K is a subgroup of the additive group $(V, +)$, hence S must contain 0.

We have the following characterization theorem for subspaces.

Theorem 2.2.3 *Let V be a vector space over K and let $S \subseteq V$. Then*

$$S \leq V \iff \begin{cases} S \neq \emptyset \quad (0 \in S) \\ \forall k_1, k_2 \in K\,,\ \forall v_1, v_2 \in S\,,\ k_1 v_1 + k_2 v_2 \in S\,. \end{cases}$$

Proof. $\boxed{\Longrightarrow}$ Taking $k = 0$ and $v_1 \in S \neq \emptyset$, we have $0 = 0 \cdot v_1 \in S$. Now let $k_1, k_2 \in K$ and $v_1, v_2 \in S$. Then we have $k_1 v_1, k_2 v_2 \in S$, and then $k_1 v_1 + k_2 v_2 \in S$.

$\boxed{\Longleftarrow}$ Choose $k_1 = k_2 = 1$ and then $k_2 = 0$ and use Definition 2.2.1. □

Example 2.2.4 (a) Every non-zero vector space V over K has two subspaces, namely $\{0\}$ and V. They are called the *trivial subspaces*.

(b) Let

$$S = \{(x,y,z) \in \mathbb{R}^3 \mid x + y + z = 0\}\,,$$
$$T = \{(x,y,z) \in \mathbb{R}^3 \mid x = y = z\}\,.$$

We have $S \neq \emptyset$, because $(0,0,0) \in S$. Now let $k_1, k_2 \in \mathbb{R}$ and $v_1, v_2 \in S$. Then $v_1 = (x_1, y_1, z_1)$ and $v_2 = (x_2, y_2, z_2)$ for some $x_1, y_1, z_1, x_2, y_2, z_2 \in \mathbb{R}$ such that $x_1 + y_1 + z_1 = 0$ and $x_2 + y_2 + z_2 = 0$. It follows that

$$k_1 v_1 + k_2 v_2 = (k_1 x_1 + k_2 x_2, k_1 y_1 + k_2 y_2, k_1 z_1 + k_2 z_2)$$

and we have

$$(k_1 x_1 + k_2 x_2) + (k_1 y_1 + k_2 y_2) + (k_1 z_1 + k_2 z_2) = k_1(x_1 + y_1 + z_1) + k_2(x_2 + y_2 + z_2) = 0.$$

Hence $k_1 v_1 + k_2 v_2 \in S$, and thus S is a subspace of the canonical real vector space $\mathbb{R}^3$. Note that S is a plane passing through the origin. For instance, the plane

$$\{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 1\}$$

is not a subspace of $\mathbb{R}^3$ over $\mathbb{R}$.

We have $T \neq \emptyset$, because $(0,0,0) \in T$. Now let $k_1, k_2 \in \mathbb{R}$ and $v_1, v_2 \in T$. Then $v_1 = (x_1, x_1, x_1)$ and $v_2 = (x_2, x_2, x_2)$ for some $x_1, x_2 \in \mathbb{R}$. It follows that

$$k_1 v_1 + k_2 v_2 = (k_1 x_1 + k_2 x_2, k_1 x_1 + k_2 x_2, k_1 x_1 + k_2 x_2).$$

Hence $k_1 v_1 + k_2 v_2 \in T$, and thus T is a subspace of the canonical real vector space $\mathbb{R}^3$. Note that T is a line passing through the origin.

(c) More generally, the only subspaces of $\mathbb{R}^3$ are $\{(0,0,0)\}$, any line containing the origin, any plane containing the origin and $\mathbb{R}^3$.

(d) Let $n \in \mathbb{N}$ and let

$$K_n[X] = \{f \in K[X] \mid \text{degree}(f) \leq n\}.$$

Then $K_n[X]$ is a subspace of the polynomial vector space $K[X]$ over K. Note that the set $\{f \in K[X] \mid \text{degree}(f) = n\}$ is not a subspace of $K[X]$ over K.

(e) Let $n \in \mathbb{N}$ with $n \geq 2$ and let

$$UT_n(K) = \{(a_{ij}) \in M_n(K) \mid a_{ij} = 0,\ \forall i, j \in \{1, \ldots, n\} \text{ with } i > j\}$$
$$LT_n(K) = \{(a_{ij}) \in M_n(K) \mid a_{ij} = 0,\ \forall i, j \in \{1, \ldots, n\} \text{ with } i < j\}$$

be the sets of *upper-triangular* and *lower-triangular* matrices with entries in K respectively. Then $UT_n(K)$ and $LT_n(K)$ are subspaces of the vector space $M_n(K)$ over K.

(f) Let $I \subseteq \mathbb{R}$ be an interval. By Example 2.1.3,

$$\mathbb{R}^I = \{f \mid f : I \to \mathbb{R}\}$$

is a real vector space, where the addition and the scalar multiplication are defined as follows: $\forall f, g : I \to \mathbb{R}$, $\forall k \in K$, we have $f + g : I \to \mathbb{R}$, $kf : I \to \mathbb{R}$, where

$$(f + g)(x) = f(x) + g(x),$$
$$(kf)(x) = kf(x), \forall x \in I.$$

The subsets

$$C(I, \mathbb{R}) = \{f \in \mathbb{R}^I \mid f \text{ continuous on } I\},$$
$$D(I, \mathbb{R}) = \{f \in \mathbb{R}^I \mid f \text{ derivable on } I\}$$

are subspaces of $\mathbb{R}^I$, because they are non-empty and we have:

$$\forall k_1, k_2 \in \mathbb{R}, \forall f, g \in C(I, \mathbb{R}),\ k_1 f + k_2 g \in C(I, \mathbb{R}),$$
$$\forall k_1, k_2 \in \mathbb{R}, \forall f, g \in D(I, \mathbb{R}),\ k_1 f + k_2 g \in D(I, \mathbb{R}).$$

2.3 Generated subspace

For a vector space V over K, we denote by $S(V)$ the set of all subspaces of V. Sometimes, this set is denoted by $S_K(V)$ if we like to emphasize the field K.

Theorem 2.3.1 *Let V be a vector space over K and let $(S_i)_{i\in I}$ be a family of subspaces of V. Then $\bigcap_{i\in I} S_i \in S(V)$.*

Proof. For each $i \in I$, we have $S_i \in S(V)$, hence $0 \in S_i$. Then $0 \in \bigcap_{i\in I} S_i \neq \emptyset$. Now let $k_1, k_2 \in K$ and $x, y \in \bigcap_{i\in I} S_i$. Then $x, y \in S_i$, $\forall i \in I$. But $S_i \in S(V)$, $\forall i \in I$. It follows that $k_1x + k_2y \in S_i$, $\forall i \in I$, hence $k_1x + k_2y \in \bigcap_{i\in I} S_i$. Therefore, $\bigcap_{i\in I} S_i \in S(V)$. □

Remark 2.3.2 In general, the union of two subspaces of a vector space is not a subspace. For instance, $S = \{(x, 0) \mid x \in \mathbb{R}\}$ and $T = \{(0, y) \mid y \in \mathbb{R}\}$ are subspaces of the canonical real vector space $\mathbb{R}^2$, but $S \cup T$ is not a subspace of $\mathbb{R}^2$. Indeed, for instance, we have $(1, 0), (0, 1) \in S \cup T$, but $(1, 0) + (0, 1) = (1, 1) \notin S \cup T$.

Now we are interested in how to "complete" a given subset of a vector space to a subspace in a minimal way. This is the motivation for the following definition.

Definition 2.3.3 Let V be a vector space and let $X \subseteq V$. Then we denote

$$\langle X\rangle = \bigcap\{S \leq V \mid X \subseteq S\}$$

and we call it the *subspace generated by X* or the *subspace spanned by X*.
Here X is called the *generating set* of $\langle X\rangle$.
If $X = \{v_1, \ldots, v_n\}$, we denote $\langle v_1, \ldots, v_n\rangle = \langle\{v_1, \ldots, v_n\}\rangle$.

Remark 2.3.4 (1) $\langle X\rangle$ is the "smallest" (with respect to inclusion) subspace of V containing X.

(2) $\langle\emptyset\rangle = \{0\}$.

(3) If $S \leq V$, then $\langle S\rangle = S$.

Definition 2.3.5 A vector space V over K is called *finitely generated* if $\exists v_1, \ldots, v_n \in V$ $(n \in \mathbb{N})$ such that $V = \langle v_1, \ldots, v_n\rangle$. Then the set $\{v_1, \ldots, v_n\}$ is called a *system of generators for V*.

Definition 2.3.6 Let V be a vector space over K and $v_1, \ldots, v_n \in V$ $(n \in \mathbb{N})$. A finite sum of the form

$$k_1v_1 + \cdots + k_nv_n\,,$$

where $k_i \in K$, $v_i \in X$ $(i = 1, \ldots, n)$, is called a (finite) *linear combination* of the vectors $v_1, \ldots, v_n$.

Let us now determine how the elements of a generated subspace look like.

Theorem 2.3.7 *Let V be a vector space over K and let $\emptyset \neq X \subseteq V$. Then*

$$\langle X\rangle = \{k_1v_1 + \cdots + k_nv_n \mid k_i \in K\,,\ v_i \in X\,, i = 1, \ldots, n\,,\ n \in \mathbb{N}^*\}\,,$$

that is, the set of all finite linear combinations of vectors of X.

Proof. We prove the result in 3 steps, by showing that

$$L = \{k_1v_1 + \cdots + k_nv_n \mid k_i \in K\,,\ v_i \in X\,, i = 1, \ldots, n\,,\ n \in \mathbb{N}^*\}$$

is the smallest subspace of V containing X.

(i) Let $v \in X$. Then $v = 1 \cdot v \in L$, hence $L \neq \emptyset$. Now let $k, k' \in K$ and $v, v' \in L$. Then $v = \sum_{i=1}^{n} k_iv_i$ and $v' = \sum_{j=1}^{m} k'_jv'_j$ for some $k_1, \ldots, k_n, k'_1, \ldots, k'_m \in K$ and $v_1, \ldots, v_n, v'_1, \ldots, v'_m \in X$. Hence

$$kv + k'v' = k\sum_{i=1}^{n} k_iv_i + k'\sum_{j=1}^{m} k'_jv'_j = \sum_{i=1}^{n}(kk_i)v_i + \sum_{j=1}^{m}(k'k'_j)v'_j \in L\,,$$

because it is a finite linear combination of vectors of X. Hence we have $L \leq V$.

(ii) Choose $n = 1$ and $k_1 = 1$ in order to see that $X \subseteq L$.

(iii) Let $S \leq V$ be such that $X \subseteq S$. Let $k_1, \ldots, k_n \in K$ and $v_1, \ldots, v_n \in X$. Since $X \subseteq S$ and $S \leq V$, it follows that

$$k_1v_1 + \cdots + k_nv_n \in S.$$

Hence $L \subseteq S$.

Thus, we have $\langle X\rangle = L$ by the remark from the beginning of the proof. □

Corollary 2.3.8 *Let V be a vector space over K and let $x_1, \ldots, x_n \in V$. Then*

$$\langle x_1, \ldots, x_n\rangle = \{k_1x_1 + \cdots + k_nx_n \mid k_i \in K\,,\ x_i \in X\,, i = 1, \ldots, n\}\,.$$

Example 2.3.9 (a) Consider the canonical real vector space $\mathbb{R}^3$. Then

$$\begin{aligned}\langle(1,0,0),(0,1,0),(0,0,1)\rangle &= \{k_1(1,0,0) + k_2(0,1,0) + k_3(0,0,1) \mid k_1, k_2, k_3 \in \mathbb{R}\}\\ &= \{(k_1,0,0) + (0,k_2,0) + (0,0,k_3) \mid k_1, k_2, k_3 \in \mathbb{R}\}\\ &= \{(k_1,k_2,k_3) \mid k_1, k_2, k_3 \in \mathbb{R}\} = \mathbb{R}^3\,.\end{aligned}$$

Hence $\mathbb{R}^3$ is generated by the three vectors $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, and thus it is finitely generated.

(b) Consider the canonical vector space $\mathbb{Z}_2^3$ over $\mathbb{Z}_2$. Then

$$\begin{aligned}\langle(\widehat{1},\widehat{0},\widehat{0}),(\widehat{0},\widehat{1},\widehat{0})\rangle &= \{k_1(\widehat{1},\widehat{0},\widehat{0}) + k_2(\widehat{0},\widehat{1},\widehat{0}) \mid k_1, k_2 \in \mathbb{Z}_2\}\\ &= \{(k_1,\widehat{0},\widehat{0}) + (\widehat{0},k_2,\widehat{0}) \mid k_1, k_2 \in \mathbb{Z}_2\}\\ &= \{(k_1,k_2,\widehat{0}) \mid k_1, k_2 \in \mathbb{Z}_2\} \neq \mathbb{Z}_2^3\,.\end{aligned}$$

Hence $\mathbb{Z}_2^3$ is not generated by the two vectors $(\widehat{1},\widehat{0},\widehat{0})$ and $(\widehat{0},\widehat{1},\widehat{0})$. But it is generated by $(\widehat{1},\widehat{0},\widehat{0})$, $(\widehat{0},\widehat{1},\widehat{0})$ and $(\widehat{0},\widehat{0},\widehat{1})$, hence it is finitely generated.

(c) Consider the subspace

$$S = \{(x,y,z) \in \mathbb{R}^3 \mid x - y - z = 0\}$$

of the canonical real vector space $\mathbb{R}^3$. Let us write it as a generated subspace. Expressing $x = y + z$, we have:

$$\begin{aligned} S &= \{(y+z,y,z) \mid y,z \in \mathbb{R}\} \\ &= \{(y,y,0) + (z,0,z) \mid y,z \in \mathbb{R}\} \\ &= \{y(1,1,0) + z(1,0,1) \mid y,z \in \mathbb{R}\} \\ &= \langle (1,1,0),(1,0,1)\rangle. \end{aligned}$$

Alternatively, one may express y or z by using the other two components and get other writings of S as a generated subspace, namely

$$S = \langle (1,1,0),(0,-1,1)\rangle = \langle (1,0,1),(0,1,-1)\rangle.$$

We see that S is finitely generated.

In what follows we shall be interested in "decomposing" a vector space into subspaces. This allows one to study the component subspaces and then deduce properties of the whole vector space.

Let us first define the sum and the direct sum of two subspaces of a vector space.

Definition 2.3.10 Let V be a vector space over K and let $S,T \leq V$.
We define the *sum* of the subspaces S and T as the set

$$S + T = \{s + t \mid s \in S\,,\ t \in T\}\,.$$

If $S \cap T = \{0\}$, then $S + T$ is denoted by $S \oplus T$ and is called the *direct sum* of the subspaces S and T.

Theorem 2.3.11 *Let V be a vector space over K and let $S,T \leq V$. Then*

$$S + T = \langle S \cup T\rangle\,.$$

Proof. We prove the equality by double inclusion.

First, let $v = s + t \in S + T$, for some $s \in S$ and $t \in T$. Then

$$v = 1 \cdot s + 1 \cdot t$$

is a linear combination of the vectors $s,t \in S \cup T$, hence $v \in \langle S \cup T\rangle$. Thus, $S+T \subseteq \langle S \cup T\rangle$.

Now let $v \in \langle S \cup T\rangle$. Then

$$v = \sum_{i=1}^{n} k_i v_i = \sum_{i \in I} k_i v_i + \sum_{j \in J} k_j v_j\,,$$

where $I = \{i \in \{1, \ldots, n\} \mid v_i \in S\}$ and $J = \{j \in \{1, \ldots, n\} \mid v_j \in T \setminus S\}$. But the first sum is a linear combination of vectors of S, hence it belongs to S, while the second sum is a linear combination of vectors of T, hence it belongs to T. Thus, $v \in S + T$ and consequently $\langle S \cup T \rangle \subseteq S + T$.

Therefore, $S + T = \langle S \cup T \rangle$. □

Corollary 2.3.12 *Let V be a vector space over K and let $S, T \leq V$. Then $S+T \leq V$.*

Proof. By Theorem 2.3.11. □

Theorem 2.3.13 *Let V be a vector space over K and let $S, T \leq V$. Then*

$$V = S \oplus T \iff \forall v \in V\,, \exists! s \in S\,, t \in T \ : v = s + t\,.$$

Proof. $\boxed{\Longrightarrow}$ Assume that $V = S \oplus T$. Let $v \in V$. Then $\exists s \in S\,, t \in T$ such that $v = s+t$. Now suppose that $\exists s' \in S\,, t' \in T$ such that $v = s' + t'$. Then $s + t = s' + t'$, whence

$$s - s' = t' - t \in S \cap T = \{0\}\,.$$

Hence $s = s'$ and $t = t'$, that show the uniqueness.

$\boxed{\Longleftarrow}$ Assume that $\forall v \in V$, $\exists! s \in S\,, t \in T$ such that $v = s + t$. Then $V \subseteq S + T$. Clearly, we have $S+T \subseteq V$ and consequently $V = S+T$. Now suppose that $0 \neq v \in S \cap T$. Then

$$v = v + 0 = 0 + v.$$

But this is a contradiction, since we have the uniqueness of writing of v as a sum of an element of S and an element of T. Therefore, $S \cap T = \{0\}$ and thus, $V = S \oplus T$. □

Example 2.3.14 Consider the canonical real vector space $\mathbb{R}^2$. Then $\mathbb{R}^2 = S \oplus T$, where $S = \{(x, 0) \mid x \in \mathbb{R}\}$ and $T = \{(0, y) \mid y \in \mathbb{R}\}$.

Extra: Image crossfade

Following [11], we describe a way to achieve an image crossfade effect.
A black-and-white image of (say) $n = 1024 \times 768$ pixels can be viewed as a vector in the real canonical vector space $\mathbb{R}^n$, where each component of the vector is the intensity of the corresponding pixel.
Let us consider two vectors representing images:

$v_1 =$  , $v_2 =$  .

Now consider the following intermediate images:

The vectors corresponding to the above images are the following linear combinations of the vectors v_1 and v_2:

$$v_1, \quad \frac{8}{9}v_1+\frac{1}{9}v_2, \quad \frac{7}{9}v_1+\frac{2}{9}v_2, \quad \frac{6}{9}v_1+\frac{3}{9}v_2, \quad \frac{5}{9}v_1+\frac{4}{9}v_2,$$
$$\frac{4}{9}v_1+\frac{5}{9}v_2, \quad \frac{3}{9}v_1+\frac{6}{9}v_2, \quad \frac{2}{9}v_1+\frac{7}{9}v_2, \quad \frac{1}{9}v_1+\frac{8}{9}v_2, \quad v_2.$$

One may use these images as frames in a video in order to get a crossfade effect.

2.4 Linear maps

Definition 2.4.1 Let V and V' be vector spaces over the same field K. A function $f : V \to V'$ is called:
(1) *(K-)linear map* (or *(vector space) homomorphism* or *linear transformation*) if

$$f(v_1+v_2) = f(v_1)+f(v_2), \quad \forall v_1, v_2 \in V,$$
$$f(kv) = kf(v), \quad \forall k \in K, \forall v \in V.$$

(2) *isomorphism* if it is a bijective K-linear map.
(3) *endomorphism* if it is a K-linear map and $V = V'$.
(4) *automorphism* if it is a bijective K-linear map and $V = V'$.

Remark 2.4.2 If $f : V \to V'$ is a K-linear map, then the first condition from its definition tells us that f is a group homomorphism between the groups $(V, +)$ and $(V', +)$. Then we have $f(0) = 0'$ and $f(-v) = -f(v)$, $\forall v \in V$.

We denote by $V \simeq V'$ the fact that two vector spaces V and V' are isomorphic. We also denote

$$\mathrm{Hom}_K(V, V') = \{f : V \to V' \mid f \text{ is } K\text{-linear}\},$$
$$\mathrm{End}_K(V) = \{f : V \to V \mid f \text{ is } K\text{-linear}\},$$
$$\mathrm{Aut}_K(V) = \{f : V \to V \mid f \text{ is bijective } K\text{-linear}\}.$$

Let us now give a characterization theorem for linear maps.

Theorem 2.4.3 *Let V and V' be vector spaces over K and $f : V \to V'$. Then*

$$f \text{ is a } K\text{-linear map} \iff f(k_1v_1+k_2v_2) = k_1f(v_1)+k_2f(v_2), \forall k_1, k_2 \in K, \forall v_1, v_2 \in V.$$

Proof. $\boxed{\Longrightarrow}$ Let $k_1, k_2 \in K$ and $v_1, v_2 \in V$. Then

$$f(k_1v_1+k_2v_2) = f(k_1v_1)+f(k_2v_2) = k_1f(v_1)+k_2f(v_2).$$

$\boxed{\Longleftarrow}$ Choose $k_1 = k_2 = 1$ and then $k_2 = 0$ to get the two conditions of a K-linear map. □

Example 2.4.4 (*a*) Let V and V' be vector spaces over K and let $f : V \to V'$ be defined by $f(v) = 0'$, $\forall v \in V$. Then f is a K-linear map, called the *trivial linear map*.

(*b*) Let V be a vector space over K. Then the identity map $1_V : V \to V$ is an automorphism of V.

(*c*) Let V be a vector space and $S \leq V$. Define $i : S \to V$ by $i(v) = v$, $\forall v \in S$. Then i is a K-linear map, called the *inclusion linear map*.

(*d*) Let V be a vector space over K and $a \in K$. Define $t_a : V \to V$ by $t_a(v) = av$, $\forall v \in V$. Then t_a is an endomorphism of V.

(*e*) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = x + y$. Then f is an $\mathbb{R}$-linear map, because we have

$$\begin{aligned} f(k_1(x_1, y_1) + k_2(x_2, y_2)) &= f(k_1x_1 + k_2x_2, k_1y_1 + k_2y_2) \\ &= (k_1x_1 + k_2x_2) + (k_1y_1 + k_2y_2) \\ &= k_1(x_1 + y_1) + k_2(x_2 + y_2) \\ &= k_1f(x_1, y_1) + k_2f(x_2, y_2) \end{aligned}$$

for every $k_1, k_2 \in K$ and for every $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$.

On the other hand, $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = xy$ is not an $\mathbb{R}$-linear map, because, for instance, we have

$$f((1, 0) + (0, 1)) = f(1, 1) = 1 \neq 0 = f(1, 0) + f(0, 1).$$

(*f*) Let $\theta \in \mathbb{R}$ and let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$f(x, y) = (x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta),$$

which is the counterclockwise rotation of angle θ about the origin in the plane. Then f is an $\mathbb{R}$-linear map. In particular, for $\theta = \frac{\pi}{2}$, we have $f(x, y) = (-y, x)$.

(*g*) For an interval $I = [a, b] \subseteq \mathbb{R}$ we considered the real vector space

$$\mathbb{R}^I = \{f \mid f : I \to \mathbb{R}\}$$

and its subspaces

$$C(I, \mathbb{R}) = \{f \in \mathbb{R}^I \mid f \text{ continuous on } I\},$$
$$D(I, \mathbb{R}) = \{f \in \mathbb{R}^I \mid f \text{ derivable on } I\}.$$

Then

$$F : D(I, \mathbb{R}) \to \mathbb{R}^I, \quad F(f) = f',$$
$$G : C(I, \mathbb{R}) \to \mathbb{R}, \quad G(f) = \int_a^b f(t)\mathrm{d}t,$$

are $\mathbb{R}$-linear maps.

Theorem 2.4.5 *(i) Let $f : V \to V'$ be an isomorphism of vector spaces over K. Then $f^{-1} : V' \to V$ is again an isomorphism of vector spaces over K.*
(ii) Let $f : V \to V'$ and $g : V' \to V''$ be K-linear maps. Then $g \circ f : V \to V''$ is a K-linear map.

Proof. (i) Since f is an isomorphism, f is bijective, hence so is f^{-1}.

Now let $k_1, k_2 \in K$ and $v_1', v_2' \in V'$. We have to prove that

$$f^{-1}(k_1 v_1' + k_2 v_2') = k_1 f^{-1}(v_1') + k_2 f^{-1}(v_2').$$

Let us denote $v_1 = f^{-1}(v_1')$ and $v_2 = f^{-1}(v_2')$. Then $f(v_1) = v_1'$ and $f(v_2) = v_2'$, hence

$$k_1 v_1' + k_2 v_2' = k_1 f(v_1) + k_2 f(v_2) = f(k_1 v_1 + k_2 v_2).$$

Thus we have

$$f^{-1}(k_1 v_1' + k_2 v_2') = k_1 v_1 + k_2 v_2 = k_1 f^{-1}(v_1') + k_2 f^{-1}(v_2').$$

Hence f^{-1} is an isomorphism of vector spaces over K.

(ii) Let $k_1, k_2 \in K$ and $v_1, v_2 \in V$. We have:

$$\begin{aligned}(g \circ f)(k_1 v_1 + k_2 v_2) &= g(f(k_1 v_1 + k_2 v_2)) \\ &= g(k_1 f(v_1) + k_2 f(v_2)) \\ &= k_1 g(f(v_1)) + k_2 g(f(v_2)) \\ &= k_1 (g \circ f)(v_1) + k_2 (g \circ f)(v_2).\end{aligned}$$

Hence $g \circ f$ is a K-linear map. □

Definition 2.4.6 Let $f : V \to V'$ be a K-linear map. Then the set

$$\operatorname{Ker} f = \{v \in V \mid f(v) = 0'\}$$

is called the *kernel* (or the *null space*) of the K-linear map f and the set

$$\operatorname{Im} f = \{f(v) \mid v \in V\}$$

is called the *image* (or the *range space*) of the K-linear map f.

Theorem 2.4.7 *Let $f : V \to V'$ be a K-linear map. Then*

$$\operatorname{Ker} f \leq V \text{ and } \operatorname{Im} f \leq V'.$$

Proof. First, note that $f(0) = 0'$, hence $0 \in \operatorname{Ker} f \neq \emptyset$. Let $k_1, k_2 \in K$ and $v_1, v_2 \in \operatorname{Ker} f$. We prove that $k_1 v_1 + k_2 v_2 \in \operatorname{Ker} f$. Indeed, we have:

$$f(k_1 v_1 + k_2 v_2) = k_1 f(v_1) + k_2 f(v_2) = 0',$$

and thus $k_1 v_1 + k_2 v_2 \in \operatorname{Ker} f$. Hence $\operatorname{Ker} f \leq V$.

Now note that $0' = f(0) \in \operatorname{Im} f \neq \emptyset$. Let $k_1, k_2 \in K$ and $v_1', v_2' \in \operatorname{Im} f$. We prove that $k_1 v_1' + k_2 v_2' \in \operatorname{Im} f$. We have $v_1' = f(v_1)$ and $v_2' = f(v_2)$ for some $v_1, v_2 \in V$. Then:

$$k_1 v_1' + k_2 v_2' = k_1 f(v_1) + k_2 f(v_2) = f(k_1 v_1 + k_2 v_2) \in \operatorname{Im} f.$$

Hence $\operatorname{Im} f \leq V'$. □

Theorem 2.4.8 *Let $f : V \to V'$ be a K-linear map. Then*

$$\operatorname{Ker} f = \{0\} \iff f \text{ is injective}.$$

Proof. $\boxed{\Longrightarrow}$ Assume that $\operatorname{Ker} f = \{0\}$. Let $v_1, v_2 \in V$ be such that $f(v_1) = f(v_2)$. It follows that $f(v_1 - v_2) = 0$, hence $v_1 - v_2 \in \operatorname{Ker} f = \{0\}$, and thus $v_1 = v_2$. Therefore, f is injective.

$\boxed{\Longleftarrow}$ Assume that f is injective. Clearly, we have $\{0\} \subseteq \operatorname{Ker} f$. Now let $v \in \operatorname{Ker} f$. Then $f(v) = 0' = f(0)$. By the injectivity of f, we deduce that $v = 0$. Thus $\operatorname{Ker} f \subseteq \{0\}$, and consequently, $\operatorname{Ker} f = \{0\}$. □

Theorem 2.4.9 *Let $f : V \to V'$ be a K-linear map and let $X \subseteq V$. Then*

$$f(\langle X \rangle) = \langle f(X) \rangle .$$

Proof. If $X = \emptyset$, then we have:

$$f(\langle \emptyset \rangle) = f(\{0\}) = \{f(0)\} = \{0'\} = \langle \emptyset \rangle = \langle f(\emptyset) \rangle.$$

Now assume that $X \neq \emptyset$. By Theorem 2.3.7 we have

$$\langle X \rangle = \{k_1 v_1 + \cdots + k_n v_n \mid k_i \in K\,,\ v_i \in X\,, i = 1, \ldots, n\,,\ n \in \mathbb{N}^*\}\,.$$

Since f is a K-linear map, it follows by Theorem 2.4.3 that

$$\begin{aligned} f(\langle X \rangle) &= \{f(k_1 v_1 + \cdots + k_n v_n) \mid k_i \in K\,,\ v_i \in X\,, i = 1, \ldots, n\,,\ n \in \mathbb{N}^*\} \\ &= \{k_1 f(v_1) + \cdots + k_n f(v_n) \mid k_i \in K\,,\ v_i \in X\,, i = 1, \ldots, n\,,\ n \in \mathbb{N}^*\} \\ &= \langle f(X) \rangle\,, \end{aligned}$$

which proves the result. □

Theorem 2.4.10 *Let V and V' be vector spaces over K. Consider on $\operatorname{Hom}_K(V, V')$ the operations: $\forall f, g \in \operatorname{Hom}_K(V, V')$ and $\forall k \in K$, $f + g, k \cdot f \in \operatorname{Hom}_K(V, V')$, where*

$$\begin{aligned} (f + g)(v) &= f(v) + g(v)\,, \\ (kf)(v) &= kf(v) \end{aligned}$$

$\forall v \in V$. Then $\operatorname{Hom}_K(V, V')$ is a vector space over K.

Proof. Let $k \in K$ and $f, g \in \operatorname{Hom}_K(V, V')$. Let us prove first that the operations are well-defined, that is, $f + g, kf \in \operatorname{Hom}_K(V, V')$. Let $k_1, k_2 \in K$ and $v_1, v_2 \in V$. Then:

$$\begin{aligned} (f + g)(k_1 x v_1 + k_2 v_2) &= f(k_1 v_1 + k_2 v_2) + g(k_1 v_1 + k_2 v_2) \\ &= k_1 f(v_1) + k_2 f(v_2) + k_1 g(v_1) + k_2 g(v_2) \\ &= k_1 (f(v_1) + g(v_1)) + k_2 (f(v_2) + g(v_2)) \\ &= k_1 (f + g)(v_1) + k_2 (f + g)(v_2)\,. \end{aligned}$$

We also have:

$$\begin{aligned}(kf)(k_1v_1 + k_2v_2) &= kf(k_1v_1 + k_2v_2)\\ &= k(k_1f(v_1)) + k(k_2f(v_2))\\ &= (kk_1)f(v_1) + (kk_2)f(v_2)\\ &= k_1(kf(v_1)) + k_2(kf(v_2))\,.\end{aligned}$$

Therefore, $f + g, kf \in \mathrm{Hom}_K(V, V')$.

It is easy to check that $(\mathrm{Hom}_K(V, V'), +)$ is an abelian group, where the identity element is the trivial linear map

$$\theta : V \to V', \quad \theta(v) = 0', \quad \forall v \in V$$

and every element $f \in \mathrm{Hom}_K(V, V')$ has a symmetric

$$-f \in \mathrm{Hom}_K(V, V'), \quad (-f)(v) = -f(v), \quad \forall v \in V.$$

Checking the axioms of the vector space for $\mathrm{Hom}_K(V, V')$ reduces, by the definitions of operations, to the axioms for the vector space V'. □

Corollary 2.4.11 *Let V be a vector space over K. Then $\mathrm{End}_K(V)$ is a vector space over K.*

Proof. Take $V = V'$ in Theorem 2.4.10. □

2.5 Quotient vector spaces

In this section we discuss the notion of quotient (factor) vector space, which will be useful in the chapter on Coding Theory.

Definition 2.5.1 Let V be a vector space over K, S a subspace of V and $v \in V$. The set denoted by

$$v + S = \{v + s \mid s \in S\}$$

is called a *coset* of S.

Theorem 2.5.2 *Let V be a vector space over K and let S be a subspace of V. Consider the relation r_S on V defined by*

$$v_1\, r_S\, v_2 \iff v_1 - v_2 \in S\,.$$

Then:
(i) r_S is an equivalence relation on V.
(ii) The quotient set

$$V/r_S = \{r_S < v >\mid v \in V\} = \{v + S \mid v \in V\}$$

is a partition of V, which will be simply denoted by V/S.

Proof. (i) We prove that r_S is an equivalence relation on V, using several times the fact that S is a subspace of V.

The relation r_S is reflexive, since $\forall v \in V$, $v - v = 0 \in S$, that is, $v\, r_S\, v$.

Let $v_1, v_2, v_3 \in V$ be such that $v_1\, r_S\, v_2$ and $v_2\, r_S\, v_3$. Then $v_1 - v_2 \in S$ and $v_2 - v_3 \in S$, so that

$$v_1 - v_3 = (v_1 - v_2) + (v_2 - v_3) \in S.$$

Hence $v_1\, r_S\, v_3$, and consequently r_S is transitive.

Let $v_1, v_2 \in V$ be such that $v_1\, r_S\, v_2$. Hence $v_1 - v_2 \in S$. Then

$$v_2 - v_1 = -(v_1 - v_2) \in S.$$

Thus, $v_2\, r_S\, v_1$, and consequently r_S is symmetric.

Hence r_S is an equivalence relation on V.

(ii) By Theorem 1.3.7, $V/r_S = \{r_S < v > | \; v \in V\}$ is a partition of V. For every $v \in V$ we have

$$\begin{aligned} r_S < v > &= \{v' \in V \mid v\, r_S\, v'\} \\ &= \{v' \in V \mid v - v' \in S\} \\ &= \{v' \in V \mid v' = v + s \text{ for some } s \in S\} \\ &= v + S. \end{aligned}$$

Hence $V/r_S = \{v + S \mid v \in V\}$. □

Theorem 2.5.3 *Let V be a vector space over K and let S be a subspace of V. Then V/S is a vector space over K, called the* quotient (factor) vector space modulo S, *where the operations are defined by*

$$\begin{aligned} (v_1 + S) + (v_2 + S) &= (v_1 + v_2) + S\,, \\ k \cdot (v + S) &= (k \cdot v) + S\,, \end{aligned}$$

$\forall k \in K$, $\forall v_1 + S, v_2 + S, v + S \in V/S$.

Proof. All properties of the operations on the quotient set V/S rely on the corresponding operations on V. It is easy to check that $(V/S, +)$ is an abelian group. Let us just note that $S = 0 + S$ is the identity element and, for every $v \in V$, the symmetric of $v + S$ is $(-v) + S$.

Next we check the axioms of a vector space for V/S. Let $k, k_1, k_2 \in K$ and $v + S, v_1 + S, v_2 + S \in V/S$.

(L_1) We have:

$$\begin{aligned} k \cdot ((v_1 + S) + (v_2 + S)) &= k \cdot ((v_1 + v_2) + S) \\ &= (k \cdot (v_1 + v_2)) + S \\ &= (k \cdot v_1 + k \cdot v_2) + S \\ &= ((k \cdot v_1) + S) + ((k \cdot v_2) + S) \\ &= (k \cdot (v_1 + S)) + (k \cdot (v_2 + S)). \end{aligned}$$

(L_2) We have:

$$\begin{aligned}(k_1 + k_2) \cdot (v + S) &= ((k_1 + k_2) \cdot v) + S \\ &= (k_1 \cdot v + k_2 \cdot v) + S \\ &= ((k_1 \cdot v) + S) + ((k_2 \cdot v) + S) \\ &= (k_1 \cdot (v + S)) + (k_2 \cdot (v + S)).\end{aligned}$$

(L_3) We have:

$$\begin{aligned}(k_1 \cdot k_2) \cdot (v + S) &= ((k_1 \cdot k_2) \cdot v) + S \\ &= (k_1 \cdot (k_2 \cdot v)) + S \\ &= k_1 \cdot ((k_2 \cdot v) + S) \\ &= k_1 \cdot (k_2 \cdot (v + S)).\end{aligned}$$

(L_4) We have:

$$1 \cdot (v + S) = (1 \cdot v) + S = v + S.$$

Hence V/S is a vector space over K. □

Example 2.5.4 Consider the canonical vector space

$$V = \mathbb{Z}_2^3 = \{(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)\}$$

over $\mathbb{Z}_2$ and its subspace

$$S = \langle (1,0,0), (0,1,0) \rangle = \{(0,0,0), (1,0,0), (0,1,0), (1,1,0)\},$$

where we denote the elements $\widehat{0}$ and $\widehat{1}$ of $\mathbb{Z}_2$ simply by 0 and 1 respectively. The elements of the quotient vector space V/S are the cosets $v + S = \{v + s \mid s \in S\}$, where $v \in V$. We compute:

$$\begin{aligned}(0,0,0) + S &= \{(0,0,0), (1,0,0), (0,1,0), (1,1,0)\}, \\ (0,0,1) + S &= \{(0,0,1), (1,0,1), (0,1,1), (1,1,1)\}, \\ (0,1,0) + S &= \{(0,1,0), (1,1,0), (0,0,0), (1,0,0)\}, \\ (0,1,1) + S &= \{(0,1,1), (1,1,1), (0,0,1), (1,0,1)\}, \\ (1,0,0) + S &= \{(1,0,0), (0,0,0), (1,1,0), (0,1,0)\}, \\ (1,0,1) + S &= \{(1,0,1), (0,0,1), (1,1,1), (0,1,1)\}, \\ (1,1,0) + S &= \{(1,1,0), (0,1,0), (1,0,0), (0,0,0)\}, \\ (1,1,1) + S &= \{(1,1,1), (0,1,1), (1,0,1), (0,0,1)\}.\end{aligned}$$

Note that

$$\begin{aligned}(0,0,0) + S = (0,1,0) + S = (1,0,0) + S = (1,1,0) + S = S, \\ (0,0,1) + S = (0,1,1) + S = (1,0,1) + S = (1,1,1) + S \\ = \{(1,1,1), (0,1,1), (1,0,1), (0,0,1)\}.\end{aligned}$$

We may also use the fact that V/S is a partition of V. It follows that

$$\begin{aligned} V/S &= \{S, (0,0,1)+S\} \\ &= \{\{(0,0,0),(1,0,0),(0,1,0),(1,1,0)\},\{(1,1,1),(0,1,1),(1,0,1),(0,0,1)\}\}, \end{aligned}$$

and thus it has 2 elements.

Using the definition of the addition of vectors and the scalar multiplication of vectors of V/S, we obtain the following tables:

$+$	S	$(0,0,1)+S$
S	S	$(0,0,1)+S$
$(0,0,1)+S$	$(0,0,1)+S$	S

$\cdot$	S	$(0,0,1)+S$
0	S	S
1	S	$(0,0,1)+S$

We have seen that for a K-linear map $f : V \to V'$, $\operatorname{Ker} f$ is a subspace of V and $\operatorname{Im} f$ is a subspace of V'. The next important theorem shows how these subspaces are further related.

Theorem 2.5.5 (Isomorphism Theorem) *Let $f : V \to V'$ be a K-linear map. Then*

$$V/\operatorname{Ker} f \simeq \operatorname{Im} f\,.$$

Proof. Let us denote $S = \operatorname{Ker} f$. We have already seen that $S = \operatorname{Ker} f \leq V$, hence we may consider the quotient vector space $V/\operatorname{Ker} f$.

Define

$$\overline{f} : V/S \to \operatorname{Im} f \text{ by } \overline{f}(v+S) = f(v)\,,\ \forall v \in V\,.$$

Let us prove first that $\overline{f}$ is a well-defined function, that is, it does not depend on the choice of representatives. Indeed, for $v_1, v_2 \in V$, we have:

$$v_1+S = v_2+S \Longrightarrow v_1-v_2 \in S \Longrightarrow f(v_1-v_2) = 0 \Longrightarrow f(v_1)-f(v_2) = 0 \Longrightarrow f(v_1) = f(v_2)\,.$$

By the definition of the operation on the quotient vector space V/S, for every $k_1, k_2 \in K$ and for every $v_1, v_2 \in V$ we have

$$\begin{aligned} \overline{f}(k_1(v_1+S)+k_2(v_2+S)) &= \overline{f}((k_1v_1+k_2v_2)+S) \\ &= f(k_1v_1+k_2v_2) \\ &= k_1f(v_1)+k_2f(v_2) \\ &= k_1\overline{f}(v_1+S)+k_2\overline{f}(v_2+S)\,. \end{aligned}$$

Hence $\overline{f}$ is a K-linear map.

Now let $v_1, v_2 \in V$ be such that $\overline{f}(v_1+S) = \overline{f}(v_2+S)$. Then $f(v_1) = f(v_2)$, whence $f(v_1)-f(v_2) = 0$. It follows that $f(v_1-v_2) = 0$, that is, $v_1-v_2 \in S$. Then $v_1+S = v_2+S$. Therefore, $\overline{f}$ is injective.

Clearly, $\overline{f}$ is surjective and consequently, $\overline{f}$ is a K-isomorphism. □

Example 2.5.6 Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$f(x,y,z) = (x,y)\,,\ \forall (x,y,z) \in \mathbb{R}^3\,.$$

Then f is an $\mathbb{R}$-linear map and

$$\mathrm{Ker} f = \{(0,0,z) \mid z \in \mathbb{R}\} = \langle (0,0,1) \rangle .$$

Hence

$$\mathbb{R}^3/\mathrm{Ker} f = \{(x,y,z) + \langle (0,0,1) \rangle \mid (x,y,z) \in \mathbb{R}^3\} .$$

Geometrically, f is just the orthogonal projection of points in space on the plane xOy. Further, $\mathrm{Ker} f = \langle (0,0,1) \rangle$ is the set of all points in space whose projections on the plane xOy are the origin, that is, the axis Oz. An element $(x,y,z) + \langle (0,0,1) \rangle$ of the quotient vector space $\mathbb{R}^3/\mathrm{Ker} f$ is a line perpendicular to the plane xOy and containing the point (x,y,z). Hence $\mathbb{R}^3/\mathrm{Ker} f$ consists of all lines parallel to the axis Oz (which is in fact the line $\mathrm{Ker} f$), and it is isomorphic to $\mathrm{Im} f = \mathbb{R}^2$ by Theorem 2.5.5.

2.6 Linear independence

Definition 2.6.1 Let V be a vector space over K. We say that the vectors $v_1, \ldots, v_n \in V$ are (or the set of vectors $\{v_1, \ldots, v_n\}$ is):
(1) *linearly independent* in V if for every $k_1, \ldots, k_n \in K$,

$$k_1 v_1 + \cdots + k_n v_n = 0 \Longrightarrow k_1 = \cdots = k_n = 0 .$$

(2) *linearly dependent* in V if they are not linearly independent, that is, $\exists k_1, \ldots, k_n \in K$ not all zero such that

$$k_1 v_1 + \cdots + k_n v_n = 0 .$$

Remark 2.6.2 (1) A set consisting of a single vector v is linearly dependent $\Longleftrightarrow v = 0$.

(2) As an immediate consequence of the definition, we notice that if V is a vector space over K and $X, Y \subseteq V$ such that $X \subseteq Y$, then:
(i) If Y is linearly independent, then X is linearly independent.
(ii) If X is linearly dependent, then Y is linearly dependent. Thus, every set of vectors containing the zero vector is linearly dependent.

(3) More generally, an infinite set of vectors of V is called *linearly independent* if any finite subset is linearly independent, and *linearly dependent* if there exists a finite subset which is linearly dependent.

Theorem 2.6.3 *Let V be a vector space over K. Then the vectors $v_1, \ldots, v_n \in V$ are linearly dependent if and only if one of the vectors is a linear combination of the others, that is, $\exists j \in \{1, \ldots, n\}$ such that*

$$v_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha_i v_i$$

for some $\alpha_i \in K$, where $i \in \{1, \ldots, n\}$ and $i \neq j$.

Proof. $\boxed{\Longrightarrow}$ Assume that $v_1, \ldots, v_n \in V$ are linearly dependent. Then $\exists k_1, \ldots, k_n \in K$ not all zero, say $k_j \neq 0$, such that $k_1 v_1 + \cdots + k_n v_n = 0$. But this implies

$$-k_j v_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} k_i v_i$$

and further,

$$v_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} (-k_j^{-1} k_i) v_i \,.$$

Now choose $\alpha_i = -k_j^{-1} k_i$ for each $i \neq j$ to get the conclusion.

$\boxed{\Longleftarrow}$ Assume that $\exists j \in \{1, \ldots, n\}$ such that

$$v_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha_i v_i$$

for some $\alpha_i \in K$, where $i \in \{1, \ldots, n\}$ and $i \neq j$. Then

$$(-1) v_j + \sum_{\substack{i=1 \\ i \neq j}}^{n} \alpha_i v_i = 0 \,.$$

Since there exists such a linear combination equal to zero and the scalars are not all zero, the vectors $v_1, \ldots, v_n$ are linearly dependent. □

Example 2.6.4 (*a*) Let V_2 be the real vector space of all vectors (in the classical sense) in the plane with a fixed origin O. Recall that the addition is the usual addition of two vectors by the parallelogram rule and the external operation is the usual scalar multiplication of vectors by real scalars. Then:

(i) one vector v is linearly dependent in $V_2 \Longleftrightarrow v = 0$;

(ii) two vectors are linearly dependent in $V_2 \Longleftrightarrow$ they are collinear;

(iii) three vectors (or more) are always linearly dependent in V_2.

Now let V_3 be the real vector space of all vectors (in the classical sense) in the space with a fixed origin O. Then:

(i) one vector v is linearly dependent in $V_3 \Longleftrightarrow v = 0$;

(ii) two vectors are linearly dependent in $V_3 \Longleftrightarrow$ they are collinear;

(iii) three vectors are linearly dependent in $V_3 \Longleftrightarrow$ they are coplanar;

(iv) four vectors (or more) are always linearly dependent in V_3.

(*b*) If K is a field and $n \in \mathbb{N}^*$, then the vectors $e_1 = (1, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $e_n = (0, 0, 0, \ldots, 1) \in K^n$ are linearly independent in the canonical vector space K^n over K. In order to show that, let $k_1, \ldots, k_n \in K$ be such that

$$k_1 e_1 + k_2 e_2 + \cdots + k_n e_n = 0 \in K^n.$$

Then we have

$$k_1(1, 0, 0, \ldots, 0) + k_2(0, 1, 0, \ldots, 0) + \cdots + k_n(0, 0, 0, \ldots, 1) = (0, \ldots, 0),$$

and furthermore

$$(k_1, \ldots, k_n) = (0, \ldots, 0).$$

This implies that $k_1 = \cdots = k_n = 0$, and thus the vectors $e_1, \ldots, e_n$ are linearly independent in K^n.

(*c*) Let K be a field and $n \in \mathbb{N}$. Then the vectors $1, X, X^2, \ldots, X^n$ are linearly independent in the vector space $K_n[X] = \{f \in K[X] \mid \text{degree}(f) \leq n\}$ over K. More generally, the vectors

$$1, X, X^2, \ldots, X^n, \ldots \quad (n \in \mathbb{N})$$

form an infinite linearly independent set in the vector space $K[X]$ over K.

Now let us give a very useful practical result on linear dependence.

Theorem 2.6.5 *Let $n \in \mathbb{N}$, $n \geq 2$.*
(i) Two vectors in the canonical vector space K^n are linearly dependent $\iff$ their components are respectively proportional.
(ii) n vectors in the canonical vector space K^n are linearly dependent $\iff$ the determinant consisting of their components is zero.

Proof. (*i*) Let $v = (x_1, \ldots, x_n)$, $v' = (x'_1, \ldots, x'_n) \in K^n$. By Theorem 2.6.3, the vectors v and v' are linearly dependent if and only if one of them is a linear combination of the other, say $v' = kv$ for some $k \in K$. That is, $x'_i = kx_i$ for each $i \in \{1, \ldots, n\}$.

(*ii*) Let $v_1 = (x_{11}, x_{21}, \ldots, x_{n1})$, $\ldots$, $v_n = (x_{1n}, x_{2n}, \ldots, x_{nn}) \in K^n$. The vectors $v_1, \ldots, v_n$ are linearly dependent if and only if $\exists k_1, \ldots, k_n \in K$ not all zero such that

$$k_1 v_1 + \cdots + k_n v_n = 0\,.$$

But this is equivalent to

$$k_1(x_{11}, x_{21}, \ldots, x_{n1}) + \cdots + k_n(x_{1n}, x_{2n}, \ldots, x_{nn}) = (0, \ldots, 0)\,,$$

and further to

$$\begin{cases} k_1 x_{11} + k_2 x_{12} + \cdots + k_n x_{1n} = 0 \\ k_1 x_{21} + k_2 x_{22} + \cdots + k_n x_{2n} = 0 \\ \ldots\ldots\ldots\ldots\ldots \\ k_1 x_{n1} + k_2 x_{n2} + \cdots + k_n x_{nn} = 0\,. \end{cases}$$

We are interested in the existence of a non-zero solution for this homogeneous linear system. We will see later on that such a solution does exist if and only if the determinant of the system is zero. □

2.7 Bases

We are going to define a key notion related to a vector space, namely that of a *basis*, which will perfectly determine a vector space. For the sake of simplicity and because of our limited needs, til the end of the chapter, *by a vector space we will understand a finitely generated vector space.*

Definition 2.7.1 Let V be a vector space over K. A list of vectors $B = (v_1, \ldots, v_n) \in V^n$ is called a *basis* of V if:
(i) B is linearly independent in V;
(ii) B is a system of generators for V, that is, $\langle B \rangle = V$.

Theorem 2.7.2 *Every vector space has a basis.*

Proof. Let V be a vector space over K. If $V = \{0\}$, then it has the basis $\emptyset$.

Now let $V = \langle B \rangle \neq \{0\}$, where $B = (v_1, \ldots, v_n)$. If B is linearly independent, then B is a basis and we are done. Suppose that the list B is linearly dependent. Then by Theorem 2.6.3, $\exists j_1 \in \{1, \ldots, n\}$ such that

$$v_{j_1} = \sum_{\substack{i=1 \\ i \neq j_1}}^{n} k_i v_i$$

for some $k_i \in K$. It follows that $V = \langle B \setminus \{v_{j_1}\} \rangle$, because every vector of V can be written as a linear combination of the vectors of $B \setminus \{v_{j_1}\}$. If $B \setminus \{v_{j_1}\}$ is linearly independent, it is a basis and we are done. Otherwise, $\exists j_2 \in \{1, \ldots, n\} \setminus \{j_1\}$ such that

$$v_{j_2} = \sum_{\substack{i=1 \\ i \neq j_1, j_2}}^{n} k_i' v_i$$

for some $k_i' \in K$. It follows that $V = \langle B \setminus \{v_{j_1}, v_{j_2}\} \rangle$, because every vector of V can be written as a linear combination of the vectors of $B \setminus \{v_{j_1}, v_{j_2}\}$. If $B \setminus \{v_{j_1}, v_{j_2}\}$ is linearly independent, then it is a basis and we are done. Otherwise, we continue the procedure. If all the previous intermediate subsets are linearly dependent, we get to the step

$$V = \langle B \setminus \{v_{j_1}, \ldots, v_{j_{n-1}}\} \rangle = \langle v_{j_n} \rangle.$$

If v_{j_n} were linearly dependent, then $v_{j_n} = 0$, hence $V = \langle v_{j_n} \rangle = \{0\}$, contradiction. Hence v_{j_n} is linearly independent and thus forms a single element basis of V. □

Remark 2.7.3 We are going to see that a vector space may have more than one basis.

Let us give now a characterization theorem for a basis of a vector space.

Theorem 2.7.4 *Let V be a vector space over K. A list $B = (v_1, \ldots, v_n)$ of vectors in V is a basis of V if and only if every vector $v \in V$ can be uniquely written as a linear combination of the vectors $v_1, \ldots, v_n$, that is,*

$$v = k_1 v_1 + \cdots + k_n v_n$$

for some unique $k_1, \ldots, k_n \in K$.

Proof. $\boxed{\Longrightarrow}$ Assume that B is a basis of V. Hence B is linearly independent and $\langle B \rangle = V$. The second condition assures us that every vector $v \in V$ can be written as a linear

combination of the vectors of B. Suppose now that $v = k_1v_1 + \cdots + k_nv_n$ and $v = k'_1v_1 + \cdots + k'_nv_n$ for some $k_1, \ldots, k_n, k'_1, \ldots, k'_n \in K$. It follows that

$$(k_1 - k'_1)v_1 + \cdots + (k_n - k'_n)v_n = 0\,.$$

By the linear independence of B, we must have $k_i = k'_i$ for each $i \in \{1, \ldots, n\}$. Thus, we have proved the uniqueness of writing.

$\boxed{\Longleftarrow}$ Assume that every vector $v \in V$ can be uniquely written as a linear combination of the vectors of B. Then clearly, $V = \langle B\rangle$. For $k_1, \ldots, k_n \in K$, we have by the uniqueness of writing

$$k_1v_1 + \cdots + k_nv_n = 0 \Longrightarrow k_1v_1 + \cdots + k_nv_n = 0 \cdot v_1 + \cdots + 0 \cdot v_n \Longrightarrow$$

$$\Longrightarrow k_1 = \cdots = k_n = 0\,,$$

hence B is linearly independent. Consequently, B is a basis of V. □

Definition 2.7.5 Let V be a vector space over K, $B = (v_1, \ldots, v_n)$ a basis of V and $v \in V$. Then the scalars $k_1, \ldots, k_n \in K$ appearing in the unique writing of v as a linear combination

$$v = k_1v_1 + \cdots + k_nv_n$$

of the vectors of B are called the *coordinates of v in the basis B*.

Example 2.7.6 (a) If K is a field and $n \in \mathbb{N}^*$, then the list $E = (e_1, \ldots, e_n)$ of vectors of K^n, where

$$\begin{cases} e_1 = (1, 0, 0, \ldots, 0) \\ e_2 = (0, 1, 0, \ldots, 0) \\ \cdots\cdots\cdots \\ e_n = (0, 0, 0, \ldots, 1) \end{cases}$$

is a basis of the canonical vector space K^n over K, called the *canonical basis* (or *standard basis*). Indeed, each vector $v = (x_1, \ldots, x_n) \in K^n$ has a unique writing $v = x_1e_1 + \cdots + x_ne_n$ as a linear combination of the vectors of E, hence E is a basis of V by Theorem 2.7.4.

Notice that the coordinates of a vector in the canonical basis are just the components of that vector, fact that is not true in general.

In particular, the canonical vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$ has the above canonical basis $E = (e_1, \ldots, e_n)$, where 0 and 1 are just the elements $\widehat{0}$ and $\widehat{1}$ of $\mathbb{Z}_2$.

Also, if $n = 1$, the set $\{1\}$ is a basis of the canonical vector space K over K. For instance, $\{1\}$ is a basis of the vector space $\mathbb{C}$ over $\mathbb{C}$.

(b) Consider the canonical real vector space $\mathbb{R}^2$. We already know a basis of $\mathbb{R}^2$, namely the canonical basis $((1, 0), (0, 1))$. But it is easy to show that the list $((1, 1), (0, 1))$ is also a basis of $\mathbb{R}^2$. Therefore, a vector space may have more than one basis.

Also, note that $\{e_1\}$ is linearly independent, but not a system of generators, while the list $(e_1, e_2, e_1 + e_2)$ is a system of generators, but not linearly independent. Hence none of the two lists is a basis of the canonical real vector space $\mathbb{R}^2$.

(*c*) Let V_3 be the real vector space of all vectors (in the classical sense) in the space with a fixed origin O. Then a basis of V_3 consists of the three pairwise orthogonal *unit vectors* $\overrightarrow{i}$, $\overrightarrow{j}$, $\overrightarrow{k}$.

(*d*) Let K be a field and $n \in \mathbb{N}$. Then the list

$$E = (1, X, X^2, \ldots, X^n)$$

is a basis of the vector space $K_n[X] = \{f \in K[X] \mid \text{degree}(f) \leq n\}$ over K, because every vector (polynomial) $f \in K_n[X]$ can be uniquely written as a linear combination $a_0 \cdot 1 + a_1 \cdot X + \cdots + a_n \cdot X^n$ ($a_0, \ldots, a_n \in K$) of the vectors of E (see Theorem 2.7.4).

In this case, the coordinates of a vector $f \in K_n[X]$ in the basis B are just its coefficients as a polynomial.

(*e*) Consider the real vector space $\mathbb{R}_2[X] = \{f \in \mathbb{R}[X] \mid \text{degree}(f) \leq 2\}$. We have seen that the list $E = (1, X, X^2)$ is a basis of $\mathbb{R}_2[X]$. Let us show that the list

$$B = (1, X - 1, (X - 1)^2)$$

is also a basis of $\mathbb{R}_2[X]$. Let $g = a_0 + a_1X + a_2X^2 \in \mathbb{R}_2[X]$. We look for unique $k_1, k_2, k_3 \in \mathbb{R}$ such that

$$g = k_1 \cdot 1 + k_2 \cdot (X - 1) + k_3 \cdot (X - 1)^2.$$

The equality is equivalent to the linear system of equations

$$\begin{cases} k_1 - k_2 + k_3 &= a_0 \\ k_2 - 2k_3 &= a_1 \\ k_3 &= a_2 \end{cases}$$

which has the unique solution $k_1 = a_0 + a_1 + a_2$, $k_2 = a_1 + 2a_2$, $k_3 = a_2$. Hence B is a basis of $\mathbb{R}_2[X]$, and the coordinates of a vector $g = a_0 + a_1X + a_2X^2 \in \mathbb{R}_2[X]$ in the basis B are $a_0 + a_1 + a_2$, $a_1 + 2a_2$, a_2.

(*f*) Let K be a field. The list

$$E = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

is a basis of the vector space $M_2(K)$ over K.

More generally, let $m, n \in \mathbb{N}$, $m, n \geq 2$ and consider the matrices $E_{ij} = (a_{kl})$, where

$$a_{kl} = \begin{cases} 1 & \text{if } k = i \text{ and } l = j \\ 0 & \text{otherwise} \end{cases}.$$

Then the list consisting of all matrices E_{ij} is a basis of the vector space $M_{m,n}(K)$ over K.

In this case, the coordinates of a vector $A \in M_{m,n}(K)$ in the above basis are just the entries of that matrix.

(*g*) Consider the real vector space $M_2(\mathbb{R})$. We have seen that

$$E = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

is a basis of $M_2(\mathbb{R})$. Let us show that the list

$$B = \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right)$$

is also a basis of $M_2(\mathbb{R})$. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R})$. We look for unique $k_1, k_2, k_3, k_4 \in \mathbb{R}$ such that

$$A = k_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + k_2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + k_3 \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + k_4 \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

The equality is equivalent to the linear system of equations

$$\begin{cases} k_1 + k_2 + k_3 = a \\ k_4 = b \\ k_3 = c \\ k_2 + k_4 = d \end{cases}$$

which has the unique solution

$$\begin{cases} k_1 = a - d + b - c \\ k_2 = d - b \\ k_3 = c \\ k_4 = b \end{cases}.$$

Hence B is a basis of $M_2(\mathbb{R})$, and the coordinates of a vector $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{R})$ in the basis B are $a - d + b - c$, $d - b$, c, b.

(h) Since $\forall z \in \mathbb{C}$, $\exists! x, y \in \mathbb{R}$ such that $z = x \cdot 1 + y \cdot i$, the list $B = (1, i)$ is a basis of the vector space $\mathbb{C}$ over $\mathbb{R}$. The coordinates of a vector $z \in \mathbb{C}$ in the basis B are just its real and its imaginary part.

Theorem 2.7.7 *Let $f : V \to V'$ be a K-linear map and let $B = (v_1, \ldots, v_n)$ be a basis of V. Then f is determined by its values on the vectors of the basis B.*

Proof. Let $v \in V$. Since B is a basis of V, $\exists! k_1, \ldots, k_n \in K$ such that $v = k_1 v_1 + \cdots + k_n v_n$. Then

$$f(v) = f(k_1 v_1 + \cdots + k_n v_n) = k_1 f(v_1) + \cdots + k_n f(v_n),$$

that is, f is determined by $f(v_1), \ldots, f(v_n)$. □

Corollary 2.7.8 *Let $f, g : V \to V'$ be K-linear maps and let $B = (v_1, \ldots, v_n)$ be a basis of V. If $f(v_i) = g(v_i)$, $\forall i \in \{1, \ldots, n\}$, then $f = g$.*

Proof. Let $v \in V$. Then $v = k_1 v_1 + \cdots + k_n v_n$ for some $k_1, \ldots, k_n \in K$, hence

$$f(v) = f(k_1 v_1 + \cdots + k_n v_n) = k_1 f(v_1) + \cdots + k_n f(v_n) = k_1 g(v_1) + \cdots + k_n g(v_n) = g(v).$$

Therefore, $f = g$. □

Theorem 2.7.9 *Let $f : V \to V'$ be a K-linear map, and let $X = (v_1, \ldots, v_n)$ be a list of vectors in V.*
(i) If f is injective and X is linearly independent in V, then $f(X)$ is linearly independent in V'.
(ii) If f is surjective and X is a system of generators for V, then $f(X)$ is a system of generators for V'.
(iii) If f is bijective and X is a basis of V, then $f(X)$ is a basis of V'.

Proof. We have $f(X) = (f(v_1), \ldots, f(v_n))$.

(i) Let $k_1, \ldots, k_n \in K$ be such that

$$k_1 f(v_1) + \cdots + k_n f(v_n) = 0'.$$

Since f is a K-linear map, it follows that

$$f(k_1 v_1 + \cdots + k_n v_n) = f(0),$$

whence by the injectivity of f we get

$$k_1 v_1 + \cdots + k_n v_n = 0.$$

But since X is linearly independent in V, we have $k_1 = \cdots = k_n = 0$. Hence $f(X)$ is linearly independent in V'.

(ii) Since X is a system of generators for V, we have $\langle X \rangle = V$. By the surjectivity of f we have:

$$\langle f(X) \rangle = f(\langle X \rangle) = f(V) = V',$$

that is, $f(X)$ is a system of generators for V'.

(iii) This follows by (i) and (ii). □

Extra: Lossy compression

Following [11], we present a way of achieving lossy compression of images.

Definition 2.7.10 Let $k, n \in \mathbb{N}^*$ be such that $k < n$, and let u be a vector of the canonical vector space K^n over K. Then the *closest k-sparse* vector associated to u is defined as the vector obtained from u by replacing all but its k largest magnitude components by zero.

Example 2.7.11 Consider an image consisting of a single row of four pixels with intensities 200, 50, 200 and 75 respectively. We know that such an image can be viewed as a vector $u = (200, 50, 200, 75)$ in the real canonical vector space $\mathbb{R}^4$. The closest 2-sparse vector associated to u is the vector $\tilde{u} = (200, 0, 200, 0)$.

Suppose that we need to store a grayscale image of (say) $n = 2000 \times 1000$ pixels more compactly. We can view it as a vector v in the real canonical vector space $\mathbb{R}^n$. If we just store its associated closest k-sparse vector, then the compressed image may be far from the original.

One may use the following *lossy compression algorithm*:

Step 1. *Consider a* suitable *basis* $B = (v_1, \ldots, v_n)$ *of the real canonical vector space* $\mathbb{R}^n$.

Step 2. *Determine the* n*-tuple* u *(which is desired to have as many zeros as possible) of the coordinates of* v *in the basis* B.

Step 3. *Replace* u *by the closest* k*-sparse* n*-tuple* $\tilde{u}$ *for a* suitable k*, and store* $\tilde{u}$.

Step 4. *In order to recover an image from* $\tilde{u}$*, compute the corresponding linear combination of the vectors of* B *with scalars the components of* $\tilde{u}$.

Consider the following image:

First, use the closest sparse vector which supresses all but 10% of the components of v, and secondly, use the lossy compression algorithm which supresses all but 10% of the components of u in order to get the following images respectively:

2.8 Dimension

Recall that we consider only finitely generated vector spaces. Let us begin with a very useful lemma, that will be often implicitly used.

Lemma 2.8.1 *Let* V *be a vector space over* K *and let* $Y = \langle y_1, \ldots, y_n, z \rangle$. *If* $z \in \langle y_1, \ldots, y_n \rangle$, *then* $Y = \langle y_1, \ldots, y_n \rangle$.

Proof. The generated subspace Y is the set of all linear combinations of the vectors $y_1, \ldots, y_n, z$ (see Theorem 2.3.7). Since $z \in \langle y_1, \ldots, y_n \rangle$, z is a linear combination of the vectors $y_1, \ldots, y_n$. It follows that every vector in Y can be written as a linear combination only of the vectors $y_1, \ldots, y_n$. Consequently, $Y = \langle y_1, \ldots, y_n \rangle$. □

The following result is a key theorem for proving that any two bases of a vector space have the same number of elements. But it is worth mentioning that it has a much broader importance in Linear Algebra.

Theorem 2.8.2 (Steinitz Theorem, Exchange Theorem) *Let V be a vector space over K, $X = (x_1, \ldots, x_m)$ a linearly independent list of vectors of V and $Y = (y_1, \ldots, y_n)$ a system of generators of V. Then:*
(i) $m \leq n$.
(ii) m vectors of Y can be replaced by the vectors of X obtaining again a system of generators for V.

Proof. We prove this result by induction on m.

The first step is to check it for $m = 1$. Then clearly $m \leq n$. Since Y is a system of generators for V, we have $x_1 = \sum_{i=1}^{n} k_i y_i$ for some $k_1, \ldots, k_n \in K$. The list $X = \{x_1\}$ is linearly independent, hence $x_1 \neq 0$. It follows that $\exists j \in \{1, \ldots, n\}$ such that $k_j \neq 0$. Then

$$y_j = k_j^{-1} x_1 - \sum_{\substack{i=1 \\ i \neq j}}^{n} k_j^{-1} k_i y_i \,,$$

that is, y_j is a linear combination of the vectors $y_1, \ldots, y_{j-1}, x_1, y_{j+1}, \ldots, y_n$. Hence, in any linear combination of $y_1, \ldots, y_n$, the vector y_j can be expressed as a linear combination of the other vectors and x_1. Therefore, we have

$$V = \langle y_1, \ldots, y_n \rangle = \langle y_1, \ldots, y_{j-1}, x_1, y_{j+1}, \ldots, y_n \rangle \,.$$

Thus, we have obtained again a system of n generators for V containing x_1.

Let us now move on to the second step of the induction. We suppose the conclusion is true for $m - 1$ and prove it for m. Let $X = (x_1, \ldots, x_m)$ be a linearly independent list in V. Then $(x_1, \ldots, x_{m-1})$ must be also linearly independent in V. By the induction hypothesis, we have $m - 1 \leq n$ and, after a renumbering,

$$V = \langle x_1, \ldots, x_{m-1}, y_m, \ldots, y_n \rangle \,.$$

If $m - 1 = n$, then $V = \langle x_1, \ldots, x_{m-1} \rangle$, whence it follows that $x_m \in \langle x_1, \ldots, x_{m-1} \rangle$, which contradicts the fact that X is linearly independent in V. Thus $m - 1 < n$, so that $m \leq n$.

We have $x_m \in V = \langle x_1, \ldots, x_{m-1}, y_m, \ldots, y_n \rangle$, whence

$$x_m = \sum_{i=1}^{m-1} k_i x_i + \sum_{i=m}^{n} k_i y_i$$

for some $k_1, \ldots, k_n \in K$. The list X being linearly independent in V, it follows that $\exists m \leq j \leq n$ such that $k_j \neq 0$ (otherwise, $x_m = \sum_{i=1}^{m-1} k_i x_i$ and the list X would be linearly dependent in V). For simplicity of writing, assume that $j = m$. It follows that

$$y_m = k_m^{-1} x_m - \sum_{i=1}^{m-1} k_m^{-1} k_i x_i - \sum_{i=m+1}^{n} k_m^{-1} k_i y_i \,.$$

Thus, y_m is a linear combination of the vectors $x_1, \ldots, x_m, y_{m+1}, \ldots, y_n$, that is, we have $y_m \in \langle x_1, \ldots, x_m, y_{m+1}, \ldots, y_n \rangle$. Therefore, it follows that

$$V = \langle x_1, \ldots, x_{m-1}, y_m, \ldots, y_n \rangle = \langle x_1, \ldots, x_m, y_{m+1}, \ldots, y_n \rangle .$$

Thus, we have obtained again a system of generators for V, where m vectors of the list Y have been replaced by the vectors of the list X. This completes the proof. □

Remark 2.8.3 Let us point out that in Steinitz Theorem not necessarily the first m vectors of Y can be replaced by the m vectors of X.

Theorem 2.8.4 *Any two bases of a vector space have the same number of elements.*

Proof. Let V be a vector space over K and let $B = (v_1, \ldots, v_m)$ and $B' = (v'_1, \ldots, v'_n)$ be bases of V. Since B is linearly independent in V and B' is a system of generators for V, we have $m \leq n$ by Theorem 2.8.2. Since B is a system of generators for V and B' is linearly independent in V, we have $n \leq m$ by the same Theorem 2.8.2. Hence $m = n$. □

Definition 2.8.5 Let V be a vector space over K. Then the number of elements of any of its bases is called the *dimension of* V and is denoted by $\dim_K V$ or simply by $\dim V$.

Remark 2.8.6 If $V = \{0\}$, then V has the basis $\emptyset$ and $\dim V = 0$.

Example 2.8.7 Using the examples of bases given in the previous section, one can easily determine the dimension of each of those vector spaces.

(a) Let K be a field and $n \in \mathbb{N}^*$. Then $\dim_K K^n = n$.

(b) We have seen that the subspaces of $\mathbb{R}^3$ are $\{(0,0,0)\}$, any line containing the origin, any plane containing the origin and $\mathbb{R}^3$. Their dimensions are 0, 1, 2 and 3 respectively.

(c) Let K be a field and $n \in \mathbb{N}$. Then $\dim K_n[X] = n + 1$.

(d) Let K be a field. Then $\dim M_2(K) = 4$.
More generally, if $m, n \in \mathbb{N}$, $m, n \geq 2$, then $\dim M_{m,n}(K) = m \cdot n$.

(e) Consider the subspace

$$S = \{(x, y, z) \in \mathbb{R}^3 \mid x - y - z = 0\}$$

of the canonical real vector space $\mathbb{R}^3$. We have seen that $S = \langle (1,1,0), (1,0,1) \rangle$. Since the vectors $(1,1,0)$ and $(1,0,1)$ are linearly independent, it follows that $B = ((1,1,0), (1,0,1))$ is a basis of S. Hence $\dim S = 2$.

(f) We have $\dim_{\mathbb{C}} \mathbb{C} = 1$ and $\dim_{\mathbb{R}} \mathbb{C} = 2$.

Theorem 2.8.8 *Let V be a vector space over K. Then the following statements are equivalent:*
(i) $\dim V = n$.
(ii) The maximum number of linearly independent vectors in V is n.
(iii) The minimum number of generators for V is n.

Proof. $(i) \implies (ii)$ Assume that $\dim V = n$. Let $B = (v_1, \ldots, v_n)$ be a basis of V. Then B is a list of n linearly independent vectors in V. Since B is a system of generators for V, any linearly independent list in V must have at most n elements by Theorem 2.8.2.

$(ii) \implies (i)$ Assume (ii). Let $B = (v_1, \ldots, v_m)$ be a basis of V and let $(u_1, \ldots, u_n)$ be a linearly independent list in V. Since B is linearly independent, we have $m \leq n$ by hypothesis. Since B is a system of generators for V, we have $n \leq m$ by Theorem 2.8.2. Hence $m = n$ and consequently $\dim V = n$.

$(i) \implies (iii)$ Assume that $\dim V = n$. Let $B = (v_1, \ldots, v_n)$ be a basis of V. Then B is a system of n generators for V. Since B is a linearly independent list in V, any system of generators for V must have at least n elements by Theorem 2.8.2.

$(iii) \implies (i)$ Assume (iii). Let $B = (v_1, \ldots, v_m)$ be a basis of V and let $(u_1, \ldots, u_n)$ be a system of generators for V. Since B is a system of generators for V, we have $n \leq m$ by hypothesis. Since B is linearly independent, we have $m \leq n$ by Theorem 2.8.2. Hence $m = n$ and consequently $\dim V = n$. □

Theorem 2.8.9 *Let V be a vector space over K with $\dim V = n$ and $X = (u_1, \ldots, u_n)$ a list of vectors in V. Then*

$$X \text{ is linearly independent in } V \iff X \text{ is a system of generators for } V\,.$$

Proof. Let $B = (v_1, \ldots, v_n)$ be a basis of V.

$\boxed{\implies}$ Assume that X is linearly independent. Since B is a system of generators for V, we know by Theorem 2.8.2 that n vectors of B, that is, all the vectors of B, can be replaced by the vectors of X and we get another system of generators for V. Hence $\langle X \rangle = V$. Thus, X is a system of generators for V.

$\boxed{\impliedby}$ Assume that X is a system of generators for V. Suppose that X is linearly dependent. Then $\exists j \in \{1, \ldots, n\}$ such that

$$u_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} k_i u_i$$

for some $k_i \in K$. It follows that

$$V = \langle X \rangle = \langle u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_n \rangle\,.$$

But the minimum number of generators for V is n by Theorem 2.8.8, which is a contradiction. Therefore, X is linearly independent. □

Corollary 2.8.10 *Let $n \in \mathbb{N}$, $n \geq 2$. Then n vectors in K^n form a basis of the canonical vector space K^n if and only if the determinant consisting of their components is non-zero.*

Proof. We have seen that n vectors in K^n are linearly independent if and only if the determinant consisting of their components is non-zero. But if this happens, then using the fact that $\dim_K K^n = n$ and Theorem 2.8.9, the vectors are also a system of generators, and thus a basis of K^n. □

Theorem 2.8.11 *Any linearly independent list of vectors in a vector space can be completed to a basis of the vector space.*

Proof. Let V be a vector space over K. Let $B = (v_1, \dots, v_n)$ be a basis of V and let $(u_1, \dots, u_m)$ be a linearly independent list in V. Since B is a system of generators for V, we know by Theorem 2.8.2 that $m \leq n$ and m vectors of B can be replaced by the vectors $(u_1, \dots, u_m)$ obtaining again a system of generators for V, say $(u_1, \dots, u_m, v_{m+1}, \dots, v_n)$. But by Theorem 2.8.9, this is also linearly independent in V and consequently a basis of V. □

Remark 2.8.12 The completion of a linearly independent list to a basis of the vector space is not unique.

Example 2.8.13 The list (e_1, e_2), where $e_1 = (1, 0, 0)$ and $e_2 = (0, 1, 0)$, is linearly independent in the canonical real vector space $\mathbb{R}^3$. It can be completed to the canonical basis of the space, namely (e_1, e_2, e_3), where $e_3 = (0, 0, 1)$. On the other hand, since $\dim_{\mathbb{R}} \mathbb{R}^3 = 3$, in order to obtain a basis of the space it is enough to add to our list a vector v_3 such that (e_1, e_2, v_3) is linearly independent (see Theorem 2.8.9). For instance, we may take $v_3 = (1, 1, 1)$, since the determinant consisting of the components of the three vectors is non-zero.

Corollary 2.8.14 *Let V be a vector space over K and $S \leq V$. Then:*
(i) Any basis of S is a part of a basis of V.
(ii) $\dim S \leq \dim V$.
(iii) $\dim S = \dim V \iff S = V$.

Proof. *(i)* Let $(u_1, \dots, u_m)$ be a basis of S. Since the list is linearly independent, it can be completed to a basis $(u_1, \dots, u_m, v_{m+1}, \dots, v_n)$ of V by Theorem 2.8.11.

(ii) It follows by *(i)*.

(iii) Assume that $\dim S = \dim V = n$. Let $(u_1, \dots, u_n)$ be a basis of S. Then it is linearly independent in V, hence it is a basis of V by Theorem 2.8.9. Thus, if $v \in V$, then $v = k_1 u_1 + \cdots + k_n u_n$ for some $k_1, \dots, k_n \in K$, hence $v \in S$. Therefore, $S = V$. □

Theorem 2.8.15 *Let V be a vector space over K and let $S \leq V$. Then there exists $\overline{S} \leq V$ such that $V = S \oplus \overline{S}$. In particular,*

$$\dim V = \dim S + \dim \overline{S}.$$

Proof. Let $(u_1, \dots, u_m)$ be a basis of S. Then by Corollary 2.8.14, it can be completed to a basis $B = (u_1, \dots, u_m, v_{m+1}, \dots, v_n)$ of V. We consider

$$\overline{S} = \langle v_{m+1}, \dots, v_n \rangle$$

and we prove that $V = S \oplus \overline{S}$.

Let $v \in V$. Then

$$v = \sum_{i=1}^{m} k_i u_i + \sum_{i=m+1}^{n} k_i v_i \in S + \overline{S},$$

for some $k_1, \ldots, k_n \in K$. Hence $V = S + \overline{S}$.

Now let $v \in S \cap \overline{S}$. Then

$$v = \sum_{i=1}^{m} k_i u_i = \sum_{i=m+1}^{n} k_i v_i \,,$$

for some $k_1, \ldots, k_n \in K$. Hence

$$\sum_{i=1}^{m} k_i u_i - \sum_{i=m+1}^{n} k_i v_i = 0 \,,$$

whence $k_i = 0$, $\forall i \in \{1, \ldots, n\}$, because B is a basis. Thus, $v = 0$ and $S \cap \overline{S} = \{0\}$.

Therefore, $V = S \oplus \overline{S}$. □

Remark 2.8.16 This is an extremely important property of a vector space, that allows us to split it in "smaller" subspaces, that can be studied much easier and then to use that information to get information about the entire vector space.

Definition 2.8.17 Let V be a vector space over K and $S \leq V$. Then a subspace $\overline{S}$ of V such that

$$V = S \oplus \overline{S}$$

is called a *complement of S in V*.

Remark 2.8.18 A subspace of a vector space may have more than one complement (see also the remark following Theorem 2.8.11).

Example 2.8.19 Consider the subspace $S = \langle e_1, e_2 \rangle$ of the canonical real vector space $\mathbb{R}^3$, where $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$. Then clearly (e_1, e_2) is a basis of S. Now by Example 2.8.13, it can be completed to a basis of $\mathbb{R}^3$, with the vector $e_3 = (0, 0, 1)$ or with the vector $v_3 = (1, 1, 1)$. Following the proof of Theorem 2.8.15, a complement in V of the subspace $S = \langle e_1, e_2 \rangle$ is $\langle e_3 \rangle$ or $\langle v_3 \rangle$.

2.9 Dimension theorems

Theorem 2.9.1 *Let V and V' be vector spaces over K. Then*

$$V \simeq V' \iff \dim V = \dim V' \,.$$

Proof. $\boxed{\Longrightarrow}$ Let $f : V \to V'$ be a K-isomorphism and let $B = (v_1, \ldots, v_n)$ be a basis of V. Note that, since f is injective, we have $f(v_i) \neq f(v_j)$ for every $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Hence the list

$$B' = f(B) = (f(v_1), \ldots, f(v_n))$$

has n elements. By Theorem 2.7.9, B' is a basis of V'. Now it follows that $\dim V = \dim V'$.

$\boxed{\Longleftarrow}$ Assume that $\dim V = \dim V' = n$. Let $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_n)$ be bases of V and V' respectively. Define a function $f : V \to V'$ in the following way. For every $v = k_1v_1 + \cdots + k_nv_n \in V$ (where $k_1, \ldots, k_n \in K$ are uniquely determined), define

$$f(v) = k_1v'_1 + \cdots + k_nv'_n.$$

Let us first prove that f is a K-linear map. Let $\alpha, \beta \in K$ and $v, w \in V$. Then $v = k_1v_1 + \cdots + k_nv_n$ and $w = l_1v_1 + \cdots + l_nv_n$ for some unique $k_1, \ldots, k_n, l_1, \ldots, l_n \in K$. It follows that

$$\begin{aligned} f(\alpha v + \beta w) &= f((\alpha k_1 + \beta l_1)v_1 + \cdots + (\alpha k_n + \beta l_n)v_n) \\ &= (\alpha k_1 + \beta l_1)v'_1 + \cdots + (\alpha k_n + \beta l_n)v'_n \\ &= \alpha(k_1v'_1 + \cdots + k_nv'_n) + \beta(l_1v'_1 + \cdots + l_nv'_n) \\ &= \alpha f(v) + \beta f(w). \end{aligned}$$

Hence f is a K-linear map. In particular, we have $f(v_i) = v'_i$ for every $i \in \{1, \ldots, n\}$.

Now let us prove that f is bijective. Let $v' = k'_1v'_1 + \cdots + k'_nv'_n \in V'$ (where $k'_1, \ldots, k'_n \in K$ are uniquely determined). Using the fact that $f(v_i) = v'_i$ for every $i \in \{1, \ldots, n\}$, it follows that

$$v' = k'_1f(v_1) + \cdots + k'_nfv_n) = f(k'_1v_1 + \cdots + k'_nv_n),$$

where the vector $k'_1v_1 + \cdots + k'_nv_n \in V$ is uniquely determined. Hence f is bijective, and thus f is a K-isomorphism. □

We may immediately deduce the following result.

Theorem 2.9.2 *Any vector space V over K with $\dim V = n$ is isomorphic to the canonical vector space K^n over K.*

Remark 2.9.3 Theorem 2.9.2 is a very important structure theorem, saying that, up to an isomorphism, *any finite dimensional vector space over K is, in fact, the canonical vector space K^n over K*. For instance, we have the K-isomorphisms $K_n[X] \simeq K^{n+1}$ and $M_{m,n}(K) \simeq K^{mn}$. Now we have an explanation why we have used so often the canonical vector spaces: not only because the operations are very nice and easily defined, but they are, up to an isomorphism, the only types of finite dimensional vector spaces.

Definition 2.9.4 Let $f : V \to V'$ be a K-linear map. Then:
(1) $\dim(\mathrm{Ker} f)$ is called the *nullity* of f, and is denoted by $\mathrm{null}(f)$.
(2) $\dim(\mathrm{Im} f)$ is called the *rank* of f, and is denoted by $\mathrm{rank}(f)$.

Next we present an important theorem relating the nullity and the rank of a linear map.

Theorem 2.9.5 (First Dimension Theorem) *Let $f : V \to V'$ be a K-linear map. Then*

$$\dim V = \dim(\mathrm{Ker} f) + \dim(\mathrm{Im} f)\,.$$

In other words, $\dim V = \mathrm{null}(f) + \mathrm{rank}(f)$.

Proof. Let $(u_1, \ldots, u_m)$ be a basis of the subspace $\mathrm{Ker}f$ of V. Then by Corollary 2.8.14, it can be completed to a basis $B = (u_1, \ldots, u_m, v_{m+1}, \ldots, v_n)$ of V. We are going to prove that a basis of $\mathrm{Im}f$ is

$$B' = (f(v_{m+1}), \ldots, f(v_n)).$$

We first prove that B' is linearly independent in $\mathrm{Im}f$. By the K-linearity of f we have:

$$\sum_{i=m+1}^{n} k_i f(v_i) = 0 \Longrightarrow f\left(\sum_{i=m+1}^{n} k_i v_i\right) = 0 \Longrightarrow \sum_{i=m+1}^{n} k_i v_i \in \mathrm{Ker}f\,,$$

where $k_{m+1}, \ldots, k_n \in K$. Since $(u_1, \ldots, u_m)$ is a basis of $\mathrm{Ker}f$, $\exists k_1, \ldots, k_m \in K$ such that

$$\sum_{i=m+1}^{n} k_i v_i = \sum_{i=1}^{m} k_i u_i\,,$$

that is,

$$\sum_{i=1}^{m} k_i u_i - \sum_{i=m+1}^{n} k_i v_i = 0\,.$$

But $B = (u_1, \ldots, u_m, v_{m+1}, \ldots, v_n)$ is a basis of V, hence it follows that $k_i = 0$, $\forall i \in \{1, \ldots, n\}$. Therefore, B' is linearly independent in $\mathrm{Im}f$.

Let us now prove that B' is a system of generators for $\mathrm{Im}f$. Let $v' \in \mathrm{Im}f$. Then $v' = f(v)$ for some $v \in V$. Using the basis B of V, we can write

$$v = \sum_{i=1}^{m} k_i u_i + \sum_{i=m+1}^{n} k_i v_i\,,$$

for some $k_1, \ldots, k_n \in K$. Then by the K-linearity of f and the fact that $u_1, \ldots, u_m \in \mathrm{Ker}f$, it follows that

$$v' = f(v) = f\left(\sum_{i=1}^{m} k_i u_i + \sum_{i=m+1}^{n} k_i v_i\right) = \sum_{i=1}^{m} k_i f(u_i) + \sum_{i=m+1}^{n} k_i f(v_i) = \sum_{i=m+1}^{n} k_i f(v_i)\,.$$

Hence B' is a system of generators for $\mathrm{Im}f$.

Therefore, B' is a basis of $\mathrm{Im}f$ and consequently, we have

$$\dim V = n = m + (n - m) = \dim(\mathrm{Ker}f) + \dim(\mathrm{Im}f)\,,$$

which finishes the proof. □

Corollary 2.9.6 *Let $f \in \mathrm{End}_K(V)$. Then the following statements are equivalent:*
(i) f is injective.
(ii) f is surjective.
(iii) f is bijective.

Proof. It is enough to prove $(i) \Longleftrightarrow (ii)$.

$(i) \Longrightarrow (ii)$ Assume that f is injective. Then $\mathrm{Ker}f = \{0\}$ by Theorem 2.4.8, hence $\dim(\mathrm{Ker}f) = 0$. By Theorem 2.9.5, it follows that $\dim(\mathrm{Im}f) = \dim V$. But $\mathrm{Im}f \leq V$, so that $\mathrm{Im}f = V$ by Corollary 2.8.14.

$(ii) \Longrightarrow (i)$ Assume that f is surjective. Since $\mathrm{Im} f = V$, it follows by Theorem 2.9.5 that $\dim(\mathrm{Ker} f) = 0$, whence $\mathrm{Ker} f = \{0\}$. By Theorem 2.4.8, f is injective. □

Theorem 2.9.7 (Second Dimension Theorem) *Let V be a vector space over K and let S, T be subspaces of V. Then*

$$\dim S + \dim T = \dim(S \cap T) + \dim(S + T)\,.$$

Proof. Consider the direct product vector space $S \times T$ over K. Let $(s_1, \ldots, s_m)$ and $(t_1, \ldots, t_n)$ be bases of S and T respectively. For every $(s,t) \in S \times T$, we have some unique writings $s = \sum_{i=1}^m k_i s_i$ and $t = \sum_{j=1}^n k'_j t_j$ with $k_i, k'_j \in K$ for every $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. It follows that we have a unique writing

$$(s,t) = (s,0) + (0,t) = \sum_{i=1}^{m} k_i(s_i, 0) + \sum_{j=1}^{n} k'_j(0, t_j).$$

Then $((s_1,0), \ldots, (s_m,0), (0,t_1), \ldots, (0,t_n))$ is a basis of $S \times T$, and thus we have

$$\dim(S \times T) = m + n = \dim S + \dim T.$$

Now consider $f : S \times T \to S + T$ defined by $f(s,t) = s + t$. For every $(s_1,t_1), (s_2,t_2) \in S \times T$ and every $k_1, k_2 \in K$ we have

$$\begin{aligned} f(k_1(s_1,t_1 + k_2(s_2,t_2)) &= f(k_1 s_1 + k_2 s_2, k_1 t_1 + k_2 t_2) \\ &= k_1 s_1 + k_2 s_2 + k_1 t_1 + k_2 t_2 \\ &= k_1(s_1 + t_1) + k_2(s_2 + t_2) \\ &= k_1 f(s_1,t_1) + k_2 f(s_2,t_2). \end{aligned}$$

Hence f is a K-linear map. We have

$$\begin{aligned} \mathrm{Ker}\, f &= \{(s,t) \in S \times T \mid f(s,t) = 0\} \\ &= \{(s,t) \in S \times T \mid s + t = 0\} \\ &= \{(s,-s) \mid s \in S \cap T\}. \end{aligned}$$

Note that $g : S \cap T \to \mathrm{Ker}\, f$ defined by $g(s) = (s,-s)$ is a K-isomorphism. Hence we have $\dim(S \cap T) = \dim(\mathrm{Ker}\, f)$ by Theorem 2.9.1. Also, we have $\mathrm{Im}\, f = S + T$.

Now by the First Dimension Theorem we deduce that

$$\dim S + \dim T = \dim(S \times T) = \dim(\mathrm{Ker}\, f) + \dim \mathrm{Im}\, f = \dim(S \cap T) + \dim(S + T),$$

which shows the conclusion. □

Corollary 2.9.8 *Let V be a vector space over K, and let S and T be subspaces of V such that $V = S \oplus T$. Then*

$$\dim V = \dim S + \dim T\,.$$

Extra: Checksum function

Following [11], we present a checksum function for detecting corrupted files.

Definition 2.9.9 Let $u = (x_1, \ldots, x_n), v = (y_1, \ldots, y_n) \in K^n$. Then the *dot-product* (or *scalar product*) of u and v is the scalar

$$u \cdot v = x_1 y_1 + \cdots + x_n y_n \in K.$$

Example 2.9.10 We give an example of a checksum function which may detect accidental random corruption of a file during transmission or storage.
Let $a_1, \ldots, a_{64} \in \mathbb{Z}_2^n$ and let $f : \mathbb{Z}_2^n \to \mathbb{Z}_2^{64}$ be the $\mathbb{Z}_2$-linear map defined by

$$f(v) = (a_1 \cdot v, \ldots, a_{64} \cdot v).$$

Suppose that v is a "file". We model corruption as the addition of a random vector $e \in \mathbb{Z}_2^n$ (the error), so the corrupted version of the file is $v + e$. We look for a formula for the probability that the corrupted file has the same checksum as the original file.
The checksum of the original file v is taken to be $f(v)$, hence the checksum of the corrupted file $v + e$ is $f(v + e)$. The original file and the corrupted file have the same checksum if and only if $f(v) = f(v + e)$ if and only if $f(e) = 0$ if and only if $e \in \operatorname{Ker} f$.
Every vector space V over the field $\mathbb{Z}_2$ with $\dim V = n$ is isomorphic to $\mathbb{Z}_2^n$, hence it has 2^n vectors. In particular, $\operatorname{Ker} f$ has 2^k vectors, where $k = \dim(\operatorname{Ker} f)$.
If the error is chosen according to the uniform distribution, the probability that $v + e$ has the same checksum as v is the following:

$$P = \frac{\text{number of vectors in } \operatorname{Ker} f}{\text{number of vectors in } \mathbb{Z}_2^n} = \frac{2^k}{2^n}.$$

One may show that $\dim(\operatorname{Im} f)$ is close to $\min(n, 64)$. So if we choose $n \geq 64$, we may assume that $\dim(\operatorname{Im} f) = 64$. By the First Dimension Theorem, we have

$$k = \dim(\operatorname{Ker} f) = \dim \mathbb{Z}_2^n - \dim(\operatorname{Im} f) = n - 64.$$

Hence

$$P = \frac{2^{n-64}}{2^n} = \frac{1}{2^{64}},$$

and thus there is only a very tiny chance that the change is undetected.

Chapter 2 quiz

Decide whether the following statements are **true** or **false**.

1. The scalar multiplication on a vector space V over K is a function $V \times V \to V$.

2. If V is a vector space over K, $k_1, k_2 \in K$ and $v \in V$, then

$$(k_1 + k_2) \cdot v = k_2 \cdot v + k_1 \cdot v.$$

3. If V is a vector space over K, $k_1, k_2 \in K$ and $v \in V$, then

$$(k_1 \cdot k_2) \cdot v = k_2 \cdot (k_1 \cdot v).$$

4. If V is a non-zero vector space, then V contains infinitely many vectors.

5. If V is a non-zero vector space over an infinite field, then V contains infinitely many vectors.

6. A subset S of a vector space V is a subspace of V if S contains the zero vector of V.

7. The intersection of any subspaces of a vector space is a subspace.

8. The union of any subspaces of a vector space is a subspace.

9. The subspace generated by $\emptyset$ in a vector space is $\emptyset$.

10. Any linear combination of linear combinations of vectors in a vector space is a linear combination.

11. The union of two subspaces of a vector space is included in their sum.

12. Every linear map is a group homomorphism.

13. The inverse of an isomorphism of vector spaces is a linear map.

14. If a list of vectors in a vector space is linearly dependent, then every vector of the list is a linear combination of the other vectors.

15. If $v_1, \ldots, v_n$ are linearly dependent vectors in a vector space V, then there are non-zero scalars $k_1, \ldots, k_n \in K$ such that

$$k_1 v_1 + \cdots + k_n v_n = 0.$$

16. Let v_1, v_2, v_3 be vectors of a vector space V. If v_1, v_2 are linearly independent in V and v_2, v_3 are linearly independent in V, then v_1, v_3 are linearly independent in V.

17. Every vector space has a system of generators.

18. If X is a system of generators for a vector space V and one vector from X is a linear combination of the other vectors from X, then those other vectors form a system of generators for V.

19. Every vector space has a unique basis.

20. The zero vector space has no basis.

21. A vector space of dimension n may be generated by $n - 1$ vectors.

22. A vector space of dimension n may have $n + 1$ linearly independent vectors.

23. Every subspace of a vector space has a unique complement.

24. Every system of generators of a vector space can be completed to a basis of the vector space.

25. Let V be a vector space and let S, T be subspaces of V. Then

$$\dim S + \dim T \leq \dim (S + T).$$

Chapter 2 projects

Use a programming language of your choice and implement the following projects.

Project 2.1

- *Input*: non-zero natural number n
- *Output*:
  1. the number of bases of the vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$
  2. the vectors of each such basis (for $n \leq 4$)

Example: The vector space $\mathbb{Z}_2^2$ over $\mathbb{Z}_2$ has 4 vectors, namely

$$(0,0), (0,1), (1,0), (1,1).$$

Its dimension is 2, so every basis has two vectors v_1 and v_2. The vector v_1 may be chosen in 3 ways ($v_1 \neq 0$), while v_2 may be chosen in 2 ways ($v_2 \neq (0,0)$ and v_2 different of any linear combination of other non-zero vectors). Hence there are $3 \cdot 2 = 6$ bases.

- *Input*: $n = 2$
- *Output*:
  1. the number of bases of the vector space $\mathbb{Z}_2^2$ over $\mathbb{Z}_2$ is 6
  2. the vectors of each such basis are:

((0,1),(1,0))	((1,0),(1,1))
((0,1),(1,1))	((1,1),(0,1))
((1,0),(0,1))	((1,1),(1,0))

Project 2.2

- *Input*: non-zero natural numbers k and n with $k \leq n$
- *Output*:
  1. the number of k-dimensional subspaces of the vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$
  2. a basis of each such subspace (for $1 \leq k \leq n \leq 6$)

Example: The vector space $\mathbb{Z}_2^3$ over $\mathbb{Z}_2$ has 8 vectors, namely

$$(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1).$$

Any 2-dimensional subspace has a basis with two vectors. There are $C_7^2 = 21$ possibilities to choose 2 vectors out of the 8 vectors of $\mathbb{Z}_2^3$, but some of them will generate the same subspace. Only 7 choices will generate different subspaces.

- *Input*: $k = 2$, $n = 3$
- *Output*:
  1. the number of 2-dimensional subspaces of the vector space $\mathbb{Z}_2^3$ over $\mathbb{Z}_2$ is 7
  2. a basis of each such subspace is:

((0,0,1),(0,1,0))	((0,1,0),(1,0,1))
((0,0,1),(1,0,0))	((0,1,1),(1,0,0))
((0,0,1),(1,1,0))	
((0,1,0),(1,0,0))	((0,1,1),(1,0,1))

Chapter 3

Matrices and linear systems

In this chapter we present the essential connection between linear maps and matrices, developing a more computational apparatus. Using elementary operations, we are able to give practical methods for computing the rank or the inverse of a matrix as well as for solving linear systems of equations. We also study eigenvalues and eigenvectors, which among many applications, offer tools for computing powers of a matrix.

Throughout the present chapter K will always denote a field, and $m, n \in \mathbb{N}^*$.

3.1 Elementary operations

Definition 3.1.1 By an *elementary operation* on a list of vectors in a vector space we understand one of the following three processes:

(1) To interchange any two vectors of the list.

(2) To multiply a vector of the list by a non-zero scalar.

(3) To multiply a vector of the list by a scalar and add the result to another vector of the list.

Let us rewrite the things more formally.

Definition 3.1.2 Let V be a vector space over K. Then an *elementary operation* is one of the functions ε_{ij}, $\varepsilon_{i\alpha}$, $\varepsilon_{ij\alpha} : V^n \to V^n$ defined for every $(v_1, \ldots, v_n) \in V^n$ by:

$$\varepsilon_{ij}(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = (v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n)\,, \tag{1}$$

$$\varepsilon_{i\alpha}(v_1, \ldots, v_i, \ldots, v_n) = (v_1, \ldots, \alpha v_i, \ldots, v_n), \quad \alpha \in K^*\,, \tag{2}$$

$$\varepsilon_{ij\alpha}(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = (v_1, \ldots, v_i + \alpha v_j, \ldots, v_j, \ldots, v_n)\,, \quad \alpha \in K. \tag{3}$$

Theorem 3.1.3 *Using the previous notation, we have* $\varepsilon_{ij}, \varepsilon_{i\alpha}, \varepsilon_{ij\alpha} \in \mathrm{Aut}_K(V^n)$.

Proof. It is easy to show that V^n has a structure of a vector space over K, where the operations are defined by

$$(v_1, \ldots, v_n) + (v_1', \ldots, v_n') = (v_1 + v_1', \ldots, v_n + v_n')\,,$$
$$k(v_1, \ldots, v_n) = (kv_1, \ldots, kv_n)\,,$$

$\forall k \in K$ and $\forall (v_1, \ldots, v_n), (v'_1, \ldots, v'_n) \in V^n$. Also, it is easy to check that ε_{ij}, $\varepsilon_{i\alpha}, \varepsilon_{ij\alpha}$ are K-linear maps. They are also bijections, having the inverses

$$(\varepsilon_{ij})^{-1} = \varepsilon_{ji}\,, \quad (\varepsilon_{i\alpha})^{-1} = \varepsilon_{i\alpha^{-1}}\,, \quad (\varepsilon_{ij\alpha})^{-1} = \varepsilon_{ij(-\alpha)}\,,$$

which imply that they are automorphisms. □

Definition 3.1.4 Let V be a vector space over K. Then two lists X and X' of vectors in the vector space V^n over K are called *equivalent* if one of them can be obtained from the other by applying a finite number of elementary operations, that is, there exists a finite composition $\varphi : V^n \to V^n$ of elementary operations such that

$$\varphi(X) = X' \quad \text{or} \quad \varphi(X') = X.$$

Remark 3.1.5 (1) The composition $\varphi : V^n \to V^n$ of elementary operations is obviously bijective, hence by Theorem 3.1.3, if one of the lists can be obtained from the other by applying a finite number of elementary operations, then the other one also can.

(2) The name "equivalent" is justified by the fact that the previously defined relation is an equivalence relation (reflexive, transitive and symmetric).

Theorem 3.1.6 *Let V be a vector space over K and let X and X' be equivalent lists of vectors in the vector space V^n over K. Then:*
(i) X is linearly independent in $V^n \iff X'$ is linearly independent in V^n.
(ii) X is a system of generators for $V^n \iff X'$ is a system of generators for V^n.
(iii) X is a basis of $V^n \iff X'$ is a basis of V^n.

Proof. Since the lists X and X' are equivalent, there exists a finite composition $\varphi : V^n \to V^n$ of elementary operations such that $\varphi(X) = X'$. Since by Theorem 3.1.3, each elementary operation is an isomorphism, it follows that φ is an isomorphism. But we know that isomorphisms preserve linearly independent lists, systems of generators, and bases by Theorem 2.7.9. □

Remark 3.1.7 In what follows, let us apply the previously presented theory of elementary operations in the case of the vector space $M_{m,n}(K)$ of $m \times n$-matrices over K. In order to do that, we will see a matrix $A = (a_{ij}) \in M_{m,n}(K)$ as a list of vectors $(a^1, \ldots, a^n)$, each of them being a column

$$a^j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Then we get the following well-known elementary operations for a matrix:

(1) Interchanging any two columns of the matrix.

(2) Multiplying a column of the matrix by a non-zero scalar.

(3) Multiplying a column of the matrix by a scalar and add the result to another column of the matrix.

Applying Definition 3.1.4 in this case, we say that two matrices are *equivalent* if one of them can be obtained from the other by a finite number of the previous elementary operations on columns.

Theorem 3.1.8 *The value of an elementary operation applied on a matrix $A = (a_{ij}) \in M_{m,n}(K)$, seen as a list of column-vectors $(a^1, \dots, a^n)$, is equal to A multiplied on the right hand side by the matrix obtained from the identity matrix I_n, also seen as a list of columns, by applying the same elementary operation.*

Proof. For simplicity of writing we are going to prove the theorem for the first two columns involved in the elementary operations. We have

$$\begin{aligned}
\varepsilon_{12}(A) &= \varepsilon_{12}(a^1, a^2, a^3, \dots, a^n) \\
&= (a^2, a^1, a^3, \dots, a^n) \\
&= \begin{pmatrix} a_{12} & a_{11} & a_{13} & \dots & a_{1n} \\ a_{22} & a_{21} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m2} & a_{m1} & a_{m3} & \dots & a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \\
&= A \cdot E_{12}\,,
\end{aligned}$$

where E_{12} is the matrix obtained from the identity matrix I_n by interchanging the first two columns.

Furthermore, $\forall \alpha \in K^*$,

$$\begin{aligned}
\varepsilon_{1\alpha}(A) &= \varepsilon_{1\alpha}(a^1, a^2, \dots, a^n) \\
&= (\alpha a^1, a^2, \dots, a^n) \\
&= \begin{pmatrix} \alpha a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \alpha a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} \alpha & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & \vdots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \\
&= A \cdot E_{1\alpha}\,,
\end{aligned}$$

where $E_{1\alpha}$ is the matrix obtained from the identity matrix I_n by multiplying the first column by α.

Finally, $\forall \alpha \in K$,

$$\begin{aligned}
\varepsilon_{12\alpha}(A) &= \varepsilon_{12\alpha}(a^1, a^2, a^3, \ldots, a^n) \\
&= (a^1 + \alpha a^2, a^2, a^3, \ldots, a^n) \\
&= \begin{pmatrix} a_{11} + \alpha a_{12} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} + \alpha a_{22} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} + \alpha a_{m2} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ \alpha & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} \\
&= A \cdot E_{12\alpha}\,,
\end{aligned}$$

where $E_{12\alpha}$ is the matrix obtained from the identity matrix I_n by multiplying the second column by α and adding it to the first column. □

Definition 3.1.9 The matrices of the form E_{ij}, $E_{i\alpha}$, $E_{ij\alpha}$, obtained by applying elementary operations on the identity matrix I_n, are called *elementary matrices*.

Theorem 3.1.10 *The elementary matrices are invertible.*

Proof. We have $\det I_n = 1 \neq 0$. But the determinant remains non-zero by applying elementary operations on I_n. Hence the elementary matrices are invertible. □

Remark 3.1.11 (1) It is also possible to see a matrix $A = (a_{ij}) \in M_{m,n}(K)$ as a list of vectors $(a_1, \ldots, a_m)$, each of them being a row $a_i = (a_{i1} \ldots a_{in})$. In this case, the elementary operations are made on rows and the value of an elementary operation applied on A is equal to A multiplied on the left hand side by the matrix obtained from the identity matrix I_m by applying the same elementary operation.

(2) By the reason of methods of computing the rank of a matrix and solving linear systems of equations, from now on we will apply the elementary operations on the rows of a matrix.

Now we would like to find for a given matrix a "better" equivalent matrix. This better form is described in the following definition.

Definition 3.1.12 We say that a matrix $A \in M_{m,n}(K)$ is in *(row) echelon form* with $r \geq 1$ non-zero rows if:
(1) the rows $1, \ldots, r$ are non-zero and the rows $r+1, \ldots, m$ are zero;
(2)

$$0 \leq N(1) < N(2) < \cdots < N(r)\,,$$

where $N(i)$ denotes the number of zero elements from the beginning of the row i, $\forall i \in \{1, \ldots, r\}$.

Theorem 3.1.13 *Every non-zero matrix $A \in M_{m,n}(K)$ is equivalent to a matrix in echelon form.*

Proof. As we have mentioned, we are going to use elementary operations on the rows of a matrix. Let $A = (a_{ij}) \in M_{m,n}(K)$. Look for a row i with the least $N(i)$, that is, the least number of zero elements from the beginning of the row. Then interchange it with the first row. Let $j_1 = N(1) + 1$, that is, a_{1j_1} is the first non-zero element of the first row. Such an element is sometimes called a *pivot*. Then make zeros on the column j_1 below the element a_{1j_1} by applying elementary operations on rows. In order to do that, multiply the first row by $-a_{kj_1}a_{1j_1}^{-1}$ and add it to the row k, $\forall k \in \{2, \ldots, m\}$. Now look for a row i with the least $N(i)$, where $i \in \{2, \ldots, n\}$. Then interchange it with the second row. Let $j_2 = N(2)+1$, that is, a_{2j_2} is the first non-zero element of the second row. Then make zeros on the column j_2 below the element a_{1j_2} by applying elementary operations on rows. Repeating this procedure, we get a matrix in echelon form. □

Example 3.1.14 Consider the matrix

$$A = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 3 & 2 & -2 & 6 \\ -1 & 1 & 1 & 0 \end{pmatrix} \in M_{3,4}(\mathbb{R})\,.$$

Then by applying elementary operations only on rows, we have the following succession of equivalent matrices (we denote by $\sim$ their equivalence):

$$A = \begin{pmatrix} \boxed{1} & 1 & -1 & 2 \\ 3 & 2 & -2 & 6 \\ -1 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & \boxed{-1} & 1 & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 \end{pmatrix}\,.$$

We have pointed out the pivots by putting them in a box. In the first step, we have multiplied the first row by -3 and then by 1 and add it to the second row and to the third row respectively. In the second step, we have multiplied the second row by 2 and add it to the third row. The last matrix is in (trapezoidal) echelon form and it is equivalent to A.

3.2 Applications of elementary operations

Let us now see how to use elementary operations to compute easier the rank of a matrix and the inverse of a square matrix. Recall that we are going to apply elementary operations on the rows of a matrix.

Lemma 3.2.1 *Let $A = (a_{ij}) \in M_{m,n}(K)$, seen as a list of row-vectors $(a_1, \ldots, a_m)$. Then a sublist of $(a_1, \ldots, a_m)$ consisting of r vectors is linearly independent in K^n if and only if there exists a non-zero minor of order r of the matrix A.*

Proof. For the sake of simplicity, let us consider the sublist $X = (a_1, \ldots, a_r)$ of $(a_1, \ldots, a_m)$. Then X is linearly independent if and only if

$$k_1a_1 + \cdots + k_ra_r = 0 \Longrightarrow k_1 = \cdots = k_r = 0 \quad (k_1, \ldots, k_r \in K)\,.$$

But this is equivalent to the fact that the linear homogeneous system of equations

$$\begin{cases} k_1a_{11} + k_2a_{21} + \cdots + k_ra_{r1} = 0 \\ k_1a_{12} + k_2a_{22} + \cdots + k_ra_{r2} = 0 \\ \cdots\cdots\cdots\cdots \\ k_1a_{1n} + k_2a_{2n} + \cdots + k_ra_{rn} = 0\,. \end{cases}$$

has only the zero solution. And this happens if and only if there exists a non-zero minor of order r of A (see the forthcoming Theorem 3.6.9). □

Theorem 3.2.2 *Let $0 \neq A = (a_{ij}) \in M_{m,n}(K)$, seen as a list of row-vectors $(a_1, \ldots, a_m)$ or as a list of column-vectors $(a^1, \ldots, a^n)$. Then:*
(i) $\operatorname{rank}(A) = \dim < a_1, \ldots, a_m >= \dim < a^1, \ldots, a^n >$.
(ii) $\operatorname{rank}(A) = \operatorname{rank}(C) = r$, *where C is a matrix in echelon form with r non-zero rows equivalent to A.*

Proof. (i) Denote $r = \dim < a_1, \ldots, a_m >$. Then by Theorem 2.8.8, the maximum number of linearly independent vectors in the list $(a_1, \ldots, a_m)$ is r. Then by Lemma 3.2.1, the maximum order of the non-zero minors of A is r. Therefore, $\operatorname{rank}(A) = r$.

Similarly, $\operatorname{rank}(A) = \dim < a^1, \ldots, a^n >$.

(ii) Let us see the matrices A and C as lists of row-vectors $(a_1, \ldots, a_m)$ and $(c_1, \ldots, c_r, c_{r+1}, \ldots, c_m)$. Clearly, $c_{r+1} = \cdots = c_m = 0$ and the list $(c_1, \ldots, c_r)$ is linearly independent. By Theorems 3.1.6 and 3.2.2, the elementary operations preserve the rank of a matrix. It follows that

$$\begin{aligned} \operatorname{rank}(A) &= \dim < a_1, \ldots, a_m > \\ &= \dim < c_1, \ldots, c_m > \\ &= \dim < c_1, \ldots, c_r >= r\,. \end{aligned}$$

Clearly, $\operatorname{rank}(C) = r$. □

Example 3.2.3 Consider the matrix

$$A = \begin{pmatrix} -3 & 5 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -3 \end{pmatrix} \in M_{3,4}(\mathbb{R})\,.$$

Then

$$A \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ -1 & 1 & 0 & 1 \\ -3 & 5 & -1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & 2 & -1 & -2 \\ 0 & 8 & -4 & -8 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & 2 & -1 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix}\,,$$

hence $\operatorname{rank}(A) = 2$.

Theorem 3.2.4 *Let $A \in M_n(K)$ with $\det(A) \neq 0$. Then A is equivalent to the identity matrix I_n and the inverse matrix A^{-1} of A is obtained from the identity matrix I_n by applying the same elementary operations as one does to obtain I_n from A.*

Proof. Since $\det(A) \neq 0$, $A \in M_n(K)$ is invertible, hence we have $\text{rank}(A) = n$. Then by applying elementary operations we get to a matrix C in echelon form, having n non-zero rows. The matrix C has all the elements below the principal diagonal zero. Then one can make zeros above the principal diagonal, by applying elementary operations on rows starting with the last row. Thus, A is equivalent to a matrix in diagonal form. But since its rank is n, all the elements on the diagonal are non-zero, hence we may multiply by their inverses to get I_n. Therefore, A is equivalent to I_n.

But by Theorem 3.1.8 this means that there exist some elementary matrices $E_1, \ldots, E_k$ such that $E_k \ldots E_1 A = I_n$. It follows that

$$A^{-1} = E_k \ldots E_1 I_n \,,$$

that is, A^{-1} is obtained from the identity matrix I_n by applying the same elementary operations as one does to obtain I_n from A. □

Example 3.2.5 Consider the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \in M_3(\mathbb{R}) \,.$$

Then $\det(A) = 1 \neq 0$, hence A is invertible. Let us determine its inverse with the above described method. For simplicity of writing, we will put in the same matrix both A and the identity matrix I_3 and we will apply in parallel elementary operations on rows. We have

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{array}\right)$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & 1 & -1 \end{array}\right) .$$

We read the inverse matrix A^{-1} from the right hand half of the last matrix, hence we have

$$A^{-1} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} .$$

Now consider the same matrix with entries in $\mathbb{Z}_2$, that is,

$$A = \begin{pmatrix} \widehat{0} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{0} & \widehat{1} \end{pmatrix} \in M_3(\mathbb{Z}_2) \,.$$

The above computations, considered now in $\mathbb{Z}_2$ (where $\widehat{1} + \widehat{1} = \widehat{0}$, hence $\widehat{-1} = \widehat{1}$), give

$$A^{-1} = \begin{pmatrix} \widehat{1} & \widehat{1} & \widehat{0} \\ \widehat{0} & \widehat{1} & \widehat{1} \\ \widehat{1} & \widehat{1} & \widehat{1} \end{pmatrix} .$$

Extra: LU decomposition

We present a matrix decomposition which offers more efficient ways for computing the inverse or the determinant of a matrix or for solving square linear systems of equations.

Definition 3.2.6 A matrix $A \in M_n(K)$ has an *LU decomposition* if it may be written as $A = L \cdot U$ for some lower triangular matrix L (that is, a matrix all of whose elements above its principal diagonal are zero) and upper triangular matrix U (that is, a matrix all of whose elements under its principal diagonal are zero).

Theorem 3.2.7 *If $A \in M_n(K)$ can be reduced to an echelon form without interchanging any rows, then A has an LU decomposition, not necessarily unique. More generally, for every $A \in M_n(K)$ there is a permutation matrix P (that is, a matrix obtained by repeatedly interchanging the rows and columns of an identity matrix) such that $P \cdot A$ has an LU decomposition.*

Remark 3.2.8 If A has an LU decomposition, then

$$\det(A) = \det(L) \cdot \det(U).$$

Moreover, if A is also invertible, then

$$A^{-1} = U^{-1} \cdot L^{-1}.$$

The determinants and the inverses of L and U are computed much easier than the determinant and the inverse of A.

Example 3.2.9 We have already seen in Example 3.2.5 that the matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \in M_3(\mathbb{R})$$

is invertible and $\det(A) = 1$. Alternatively, note that there is the permutation matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{R})$$

(corresponding to interchanging the first two rows) such that

$$P \cdot A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} = L \cdot U$$

as a product of a lower triangular matrix L and an upper triangular matrix U. Then

$$\det(A) = \det(P)^{-1} \cdot \det(L) \cdot \det(U) = 1,$$

and

$$\begin{aligned} A^{-1} &= U^{-1} \cdot L^{-1} \cdot P \\ &= \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}. \end{aligned}$$

3.3 The matrix of a list of vectors

In the previous section, we have seen a matrix as a list of row-vectors. Now we discuss a converse, namely we define the matrix associated to a list of vectors, with respect to a basis.

Definition 3.3.1 Let V be a vector space over K, $B = (v_1, \dots, v_n)$ a basis of V and $X = (u_1, \dots, u_m)$ a list of vectors in V. Let

$$\begin{cases} u_1 = a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ u_2 = a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \cdots\cdots\cdots\cdots\cdots \\ u_m = a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{cases}$$

be the unique writings of the vectors in X as linear combinations of vectors of the basis B, for some $a_{ij} \in K$. The *matrix of the list of vectors* X in the basis B is the matrix having as its rows the coordinates of the vectors in X in the basis B, that is,

$$[X]_B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Example 3.3.2 Consider the canonical basis $B = (e_1, e_2, e_3, e_4)$ and the list $X = (u_1, u_2, u_3)$ in the canonical real vector space $\mathbb{R}^4$, where

$$\begin{cases} u_1 &= (1, 2, 3, 4) \\ u_2 &= (5, 6, 7, 8) \\ u_3 &= (9, 10, 11, 12) \end{cases}.$$

Since the coordinates of a vector in the canonical basis are just its components, we get

$$[X]_B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}.$$

Now we give a theorem which allows one to determine the dimension of the subspace generated by a list of vectors.

Theorem 3.3.3 *Let V be a vector space over K, $B = (v_1, \dots, v_n)$ a basis of V and $X = (u_1, \dots, u_m)$ a list of vectors in V having the matrix A in the basis B. Then:*
(i) $\dim\langle X\rangle = \operatorname{rank}(A)$.
(ii) A basis of $\langle X\rangle$ is the list of non-zero row-vectors $(c_1, \dots, c_r)$ of an echelon form C equivalent to A.

Proof. These follow by Theorem 3.2.2 and its proof. □

Example 3.3.4 Let us determine the dimensions of the subspaces S, T, $S+T$ and $S\cap T$ of the canonical real vector space $\mathbb{R}^4$, where

$$S = \langle(-3,5,-1,1),(-1,1,0,1),(1,1,-1,-3)\rangle\,,$$
$$T = \langle(1,0,2,0),(2,1,-1,2)\rangle\,.$$

One can easily show that the ranks of the matrices in the canonical basis corresponding to the vectors from S and from T respectively are both 2. Hence $\dim S = \dim T = 2$.

Furthermore, we have $S+T = \langle S\cup T\rangle$. We write the matrix of $S\cup T$ in the canonical basis and we have

$$\begin{pmatrix} -3 & 5 & -1 & 1 \\ -1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -3 \\ 1 & 0 & 2 & 0 \\ 2 & 1 & -1 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ -1 & 1 & 0 & 1 \\ -3 & 5 & -1 & 1 \\ 1 & 0 & 2 & 0 \\ 2 & 1 & -1 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & 2 & -1 & -2 \\ 0 & 8 & -4 & -8 \\ 0 & -1 & 3 & 3 \\ 0 & -1 & 1 & 8 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & -1 & 3 & 3 \\ 0 & 2 & -1 & -2 \\ 0 & 2 & -1 & -2 \\ 0 & -1 & 1 & 8 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & -1 & 3 & 3 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & -2 & 5 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & -1 & 3 & 3 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & \frac{33}{5} \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & -3 \\ 0 & -1 & 3 & 3 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 33 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then by Theorem 3.3.3, $\dim(S+T) = 4$ and a basis of $S+T$ consists of the non-zero row-vectors from the echelon form, that is,

$$((1,1,-1,-3),(0,-1,3,3),(0,0,5,4),(0,0,0,33))\,.$$

Now by the Second Dimension Theorem, it follows that

$$\dim(S\cap T) = \dim S + \dim T - \dim(S+T) = 2+2-4 = 0\,.$$

Now we are going to define the matrix of a vector in a basis of a vector space. Even if one might expect to define it as a row-matrix, by considering a single vector list, it is more convenient to define it as a column-matrix for our purposes concerning linear maps in order to avoid formulas involving transposes.

Definition 3.3.5 Let V be a vector space over K, $v\in V$ and $B=(v_1,\dots,v_n)$ a basis of V. If $v = k_1v_1+\cdots+k_nv_n$ $(k_1,\dots,k_n\in K)$ is the unique writing of v as a linear combination of the vectors of the basis B, then the *matrix of the vector* v in the basis B is

$$[v]_B = \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}.$$

Example 3.3.6 Consider the vector $v = (1, 2, 3)$ in the canonical real vector space $\mathbb{R}^3$, and let E be the canonical basis. Then

$$[v]_E = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Extra: Hill cipher

Let $n \in \mathbb{N}^*$ and consider the canonical vector space $V = \mathbb{Z}_2^n$ over $\mathbb{Z}_2$ with canonical basis E. The vectors of V may be identified with n-bit binary strings. Suppose that Alice needs to send an n-bit plaintext $p \in \mathbb{Z}_2^n$ to Bob.

Hill cipher:

1. (*Key establishment*) Alice and Bob randomly choose an invertible matrix $K \in M_n(\mathbb{Z}_2)$ as a key, and compute its inverse.
2. (*Encryption*) Alice computes the ciphertext c according to the formula
$$[c]_E^T = [p]_E^T \cdot K.$$
3. (*Decryption*) Bob computes the plaintext p according to the formula
$$[p]_E^T = [c]_E^T \cdot K^{-1}.$$

Remark 3.3.7 The Hill cipher, which is nowadays insecure, was the first application of linear algebra to cryptography.

Example 3.3.8 Alice wants to send the message $p = (1, 0, 1) \in \mathbb{Z}_2^3$ to Bob.
Alice and Bob agree on the matrix

$$K = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in M_3(\mathbb{Z}_2)$$

as a key, and compute its inverse

$$K^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{Z}_2).$$

Alice encrypts the message by computing the ciphertext c as:

$$[c]_E^T = [p]_E^T \cdot K = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}.$$

Bob decrypts the message by computing the plaintext p as:

$$[p]_E^T = [c]_E^T \cdot K^{-1} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}.$$

3.4 The matrix of a linear map

Definition 3.4.1 Let $f : V \to V'$ be a K-linear map, $B = (v_1, \ldots, v_n)$ a basis of V and $B' = (v'_1, \ldots, v'_m)$ a basis of V'. Then we can uniquely write the vectors in $f(B)$ as linear combinations of the vectors of the basis B', say

$$\begin{cases} f(v_1) = a_{11}v'_1 + a_{21}v'_2 + \cdots + a_{m1}v'_m \\ f(v_2) = a_{12}v'_1 + a_{22}v'_2 + \cdots + a_{m2}v'_m \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ f(v_n) = a_{1n}v'_1 + a_{2n}v'_2 + \cdots + a_{mn}v'_m \end{cases}$$

for some $a_{ij} \in K$.
Then the *matrix of the K-linear map f* in the bases B and B' is the matrix having as its columns the coordinates of the vectors of $f(B)$ in the basis B', that is,

$$[f]_{BB'} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

If $V = V'$ and $B = B'$, then we simply denote $[f]_B = [f]_{BB'}$.

Remark 3.4.2 We have to emphasize that we put the coordinates on the columns of the matrix of a linear map and not on the rows as we did for the matrix of a list of vectors.

Example 3.4.3 Consider the $\mathbb{R}$-linear map $f : \mathbb{R}^4 \to \mathbb{R}^3$ defined by

$$f(x, y, z, t) = (x + y + z, y + z + t, z + t + x)\,,\ \forall (x, y, z, t) \in \mathbb{R}^4\,.$$

Let $E = (e_1, e_2, e_3, e_4)$ and $E' = (e'_1, e'_2, e'_3)$ be the canonical bases in $\mathbb{R}^4$ and $\mathbb{R}^3$ respectively. Since

$$\begin{cases} f(e_1) = f(1,0,0,0) = (1,0,1) = e'_1 + e'_3 \\ f(e_2) = f(0,1,0,0) = (1,1,0) = e'_1 + e'_2 \\ f(e_3) = f(0,0,1,0) = (1,1,1) = e'_1 + e'_2 + e'_3 \\ f(e_4) = f(0,0,0,1) = (0,1,1) = e'_2 + e'_3 \end{cases}$$

it follows that the matrix of f in the bases E and E' is

$$[f]_{EE'} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Theorem 3.4.4 *Let $f : V \to V'$ be a K-linear map, $B = (v_1, \ldots, v_n)$ a basis of V, $B' = (v'_1, \ldots, v'_m)$ a basis of V' and $v \in V$. Then*

$$[f(v)]_{B'} = [f]_{BB'} \cdot [v]_B\,.$$

Proof. Let $[f]_{BB'} = (a_{ij}) \in M_{m,n}(K)$. Let $v = \sum_{j=1}^{n} k_j v_j$ and

$$f(v) = \sum_{i=1}^{m} k_i' v_i'$$

for some $k_i, k_i' \in K$. On the other hand, using the definition of the matrix of f in the bases B and B', we have

$$f(v) = f\left(\sum_{j=1}^{n} k_j v_j\right) = \sum_{j=1}^{n} k_j f(v_j) = \sum_{j=1}^{n} k_j \left(\sum_{i=1}^{m} a_{ij} v_i'\right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} k_j\right) v_i' .$$

But the writing of $f(v)$ as a linear combination of the vectors of the basis B' is unique, hence we must have $k_i' = \sum_{j=1}^{n} a_{ij} k_j$ for every $i \in \{1, \ldots, m\}$. Therefore, $[f(v)]_{B'} = [f]_{BB'} \cdot [v]_B$. □

Now we give a connection between the ranks of a linear map and of its matrix in a pair of bases.

Theorem 3.4.5 *Let $f : V \to V'$ be a K-linear map. Then*

$$\operatorname{rank}(f) = \operatorname{rank}([f]_{BB'}) ,$$

where B and B' are any bases of V and V' respectively.

Proof. Let $B = (v_1, \ldots, v_n)$ and $[f]_{BB'} = A$. Using our results relating ranks and dimensions, we have

$$\begin{aligned} \operatorname{rank}(f) &= \dim(\operatorname{Im} f) = \dim f(V) = \dim f(\langle v_1, \ldots, v_n \rangle) \\ &= \dim\langle f(v_1), \ldots, f(v_n)\rangle = \operatorname{rank}(\ A^T) = \operatorname{rank}(A) = \operatorname{rank}([f]_{BB'}) . \end{aligned}$$

Now take some other bases $B_1 = (u_1, \ldots, u_n)$ of V and B_1' of V' and denote $[f]_{B_1 B_1'} = A_1$. Then

$$\begin{aligned} \operatorname{rank}([f]_{B_1 B_1'}) &= \operatorname{rank}(A_1) = \operatorname{rank}(\ A_1^T) = \dim\langle f(u_1), \ldots, f(u_n)\rangle \\ &= \dim(\operatorname{Im} f) = \dim\langle f(v_1), \ldots, f(v_n)\rangle = \operatorname{rank}([f]_{BB'}) . \end{aligned}$$

This shows the result. □

Remark 3.4.6 Notice that the rank of a linear map does not depend on the pair of bases in which we write its matrix. Also notice that, considering matrices of a linear map in different pairs of bases, their ranks are the same. Some other connection between matrices of a linear map in different pairs of bases will be discussed in the next section.

Example 3.4.7 Consider the $\mathbb{R}$-linear map $f : \mathbb{R}^4 \to \mathbb{R}^3$ defined by

$$f(x, y, z, t) = (x + y + z, y + z + t, z + t + x) , \ \forall (x, y, z, t) \in \mathbb{R}^4 .$$

Let $E = (e_1, e_2, e_3, e_4)$ and $E' = (e_1', e_2', e_3')$ be the canonical bases in $\mathbb{R}^4$ and $\mathbb{R}^3$ respectively. Using Example 3.4.3 it follows that

$$[f]_{EE'} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix} .$$

Now by Theorem 3.4.5 it follows that $\operatorname{rank}(f) = \operatorname{rank}([f]_{EE'}) = 3$.

We end this section with a key result in Linear Algebra, connecting linear maps and matrices.

Theorem 3.4.8 *Let V, V' and V'' be vector spaces over K with $\dim V = n$, $\dim V' = m$ and $\dim V'' = p$ and let $B = (v_1, \dots, v_n)$, $B' = (v'_1, \dots, v'_m)$ and $B'' = (v''_1, \dots, v''_p)$ be bases of V, V' and V'' respectively. Then $\forall f, g \in \mathrm{Hom}_K(V, V')$, $\forall h \in \mathrm{Hom}_K(V', V'')$ and $\forall k \in K$, we have*

$$\begin{aligned} [f+g]_{BB'} &= [f]_{BB'} + [g]_{BB'}\,, \\ [kf]_{BB'} &= k \cdot [f]_{BB'}\,, \\ [h \circ f]_{BB''} &= [h]_{B'B''} \cdot [f]_{BB'}\,. \end{aligned}$$

Proof. Let $[f]_{BB'} = (a_{ij}) \in M_{m,n}(K)$, $[g]_{BB'} = (b_{ij}) \in M_{m,n}(K)$ and $[h]_{B'B''} = (c_{ki}) \in M_{pm}(K)$. Then

$$f(v_j) = \sum_{i=1}^{m} a_{ij} v'_i\,, \quad g(v_j) = \sum_{i=1}^{m} b_{ij} v'_i\,, \quad h(v'_i) = \sum_{k=1}^{p} c_{ki} v''_k$$

$\forall j \in \{1, \dots, n\}$ and $\forall i \in \{1, \dots, m\}$.

Then $\forall k \in K$ and $\forall j \in \{1, \dots, n\}$ we have

$$\begin{aligned} (f+g)(v_j) = f(v_j) + g(v_j) &= \sum_{i=1}^{m} a_{ij} v'_i + \sum_{i=1}^{m} b_{ij} v'_i = \sum_{i=1}^{m} (a_{ij} + b_{ij}) v'_i\,, \\ (kf)(v_j) = kf(v_j) &= k \cdot \left(\sum_{i=1}^{m} a_{ij} v'_i \right) = \sum_{i=1}^{m} (k a_{ij}) v'_i\,, \end{aligned}$$

hence $[f+g]_{BB'} = [f]_{BB'} + [g]_{BB'}$ and $[kf]_{BB'} = k \cdot [f]_{BB'}$.

Finally, $\forall j \in \{1, \dots, n\}$ we have

$$\begin{aligned} (h \circ f)(v_j) = h(f(v_j)) &= h\left(\sum_{i=1}^{m} a_{ij} v'_i \right) = \sum_{i=1}^{m} a_{ij} h(v'_i) \\ &= \sum_{i=1}^{m} a_{ij} \left(\sum_{k=1}^{p} c_{ki} v''_k \right) = \sum_{k=1}^{p} \sum_{i=1}^{m} (c_{ki} a_{ij}) v''_k\,, \end{aligned}$$

hence $[h \circ f]_{BB''} = [h]_{B'B''} \cdot [f]_{BB'}$. □

Theorem 3.4.9 *Let V and V' be vector spaces over K with $\dim V = n$ and $\dim V' = m$, and let B and B' be bases of V and V' respectively. Then the map*

$$\varphi : \mathrm{Hom}_K(V, V') \to M_{m,n}(K), \quad \varphi(f) = [f]_{BB'}\,, \ \forall f \in \mathrm{Hom}_K(V, V')$$

is an isomorphism of vector spaces.

Proof. We have seen that $\mathrm{Hom}_K(V, V')$ is a vector space over K with respect to the following addition and scalar multiplication: $\forall f, g \in \mathrm{Hom}_K(V, V')$ and $\forall k \in K$, $f+g, k \cdot$

$f \in \mathrm{Hom}_K(V, V')$, where $\forall x \in V$,

$$(f+g)(x) = f(x) + g(x)\,,$$
$$(kf)(x) = kf(x)\,.$$

Also, $M_{m,n}(K)$ is a vector space over K. By Theorem 3.4.8 it follows that φ is a K-linear map.

Finally, let us prove that φ is bijective. Consider $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_m)$. Let $f, g \in \mathrm{Hom}_K(V, V')$ be such that $\varphi(f) = \varphi(g)$. Then $[f]_{BB'} = [g]_{BB'} = (a_{ij}) \in M_{m,n}(K)$, hence

$$f(v_j) = a_{1j}v'_1 + a_{2j}v'_2 + \cdots + a_{mj}v'_m = g(v_j)\,,$$

$\forall j \in \{1, \ldots, n\}$. We have seen that two K-linear maps are equal if and only if they have the same values at all vectors of a basis. Hence $f = g$, which shows that φ is injective. Now let $A = (a_{ij}) \in M_{m,n}(K)$, seen as a list of column-vectors $(a^1, \ldots, a^n)$, where $a^j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$. Define a K-linear map $f : V \to V'$ on the basis of the domain by

$$f(v_j) = a_{1j}v'_1 + \cdots + a_{mj}v'_m\,,$$

$\forall j \in \{1, \ldots, n\}$. Then $\varphi(f) = [f]_{BB'} = (a_{ij}) = A$. Thus, φ is surjective. □

Remark 3.4.10 The extremely important isomorphism given in Theorem 3.4.9 allows us to work with matrices instead of linear maps, which is much simpler from a computational point of view. Under this isomorphism, the kernel and the image of a linear map $f : V \to V'$, where V and V' are vector spaces over K with $\dim(V) = n$ and $\dim(V') = m$, and bases B and B' respectively, correspond to the *null space* and to the *column space* of its associated matrix $A = [f]_{BB'} \in M_{m,n}(K)$ respectively. Thus, the *null space* of A consists of vectors $x \in K^n$ such that $Ax = 0$, while the *column space* of A consists of all linear combinations of the columns of A. A vector $b \in K^m$ belongs to the column space of A if and only if the system $Ax = b$ has a solution. By the First Dimension Theorem it follows that the sum of the dimensions of the null space and the column space of A equals n.

Theorem 3.4.11 *Let V be a vector space over K with $\dim V = n$, and let B be a basis of V. Then the map*

$$\varphi : \mathrm{End}_K(V) \to M_n(K), \quad \varphi(f) = [f]_B\,,\ \forall f \in \mathrm{End}_K(V)$$

is an isomorphism of vector spaces and of rings.

Proof. Note that $(\mathrm{End}_K(V), +, \circ)$ and $(M_n(K), +, \cdot)$ are rings. The required isomorphisms follow by Theorem 3.4.9. □

Corollary 3.4.12 *Let $f \in \mathrm{End}_K(V)$. Then*

$$f \in \mathrm{Aut}_K(V) \iff \det([f]_B) \neq 0\,,$$

where B is any basis of V.

Proof. Let B a basis of V. By Theorem 3.4.11, $f \in \mathrm{Aut}_K(V) \iff f$ is invertible in the ring $(\mathrm{End}_K(V), +, \circ) \iff [f]_B$ is invertible in the ring $(M_n(K), +, \cdot) \iff \det([f]_B) \neq 0$. □

Extra: Image transformations

Suppose that we have a 2D-image that we want to rotate counterclockwise with θ degrees around the origin. By such a rotation, the point of coordinates $(1, 0)$ becomes the point of coordinates $(\cos\theta, \sin\theta)$, while the point of coordinates $(0, 1)$ becomes the point of coordinates $(-\sin\theta, \cos\theta)$.
We look for an $\mathbb{R}$-linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ satisfying the following conditions:

$$f(1,0) = (\cos\theta, \sin\theta),$$
$$f(0,1) = (-\sin\theta, \cos\theta).$$

Recall that every linear map is determined by its values at the elements of a basis (the canonical basis in our case). Hence the matrix of the linear map f in the canonical basis E of the canonical real vector space $\mathbb{R}^2$ is:

$$[f]_E = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

For any point $v = (x, y) \in \mathbb{R}^2$ of a 2D-image, its corresponding point in the rotated image is computed as $f(v) = (x', y') \in \mathbb{R}^2$, where

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = [f(v)]_E = [f]_E \cdot [v]_E = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix}.$$

For instance, for a counterclockwise rotation of 90^0 around the origin one has the matrix:

$$[f]_E = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Extra: Graphs and networks

Definition 3.4.13 Consider a directed graph (V, E), having the set of vertices V with $|V| = n$ and the set of edges E with $|E| = m$. One may assume that the graph has no loop and no multiple edges between any two given vertices.
Its associated *incidence matrix* $A = (a_{ij}) \in M_{m,n}(\{-1, 1, 0\})$ is defined by

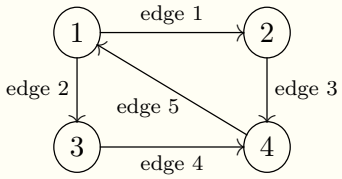
$$a_{ij} = \begin{cases} -1 & \text{if the edge } i \text{ starts at the vertex } j \\ 1 & \text{if the edge } i \text{ ends at the vertex } j \\ 0 & \text{otherwise} \end{cases}$$

for every $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$, hence each row has exactly two non-zero entries. The directed graph (V, E) is called a *network* if some numbers $l_1, \ldots, l_m$ (e.g., length, capacity etc.) are associated to the edges of the graph. To this information one may associate a diagonal matrix $C = (c_{ij}) \in M_m(\mathbb{R})$ defined by $c_{ii} = l_i$ for every $i \in \{1, \ldots, m\}$ and $c_{ij} = 0$ for every $i, j \in \{1, \ldots, m\}$ with $i \neq j$.

The matrices A and C describe the nework (e.g., a computer network) and are useful algebraic tools in network theory.

Theorem 3.4.14 *With the above notation, the column space of A has dimension $n-1$ and the null space of A has dimension 1.*

Example 3.4.15 Consider the following directed graph:



Its incidence matrix is

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix} \in M_{5,4}(\{-1,1,0\}).$$

The dimensions of the column space and the null space of A are 3 and 1 respectively.

3.5 Change of bases

We would like now to establish a connection between the matrices of a vector or of a linear map in different bases of a vector space.

Definition 3.5.1 Let V be a vector space over K, and let $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_n)$ be bases of V. Then we can uniquely write

$$\begin{cases} v'_1 = t_{11}v_1 + t_{21}v_2 + \cdots + t_{n1}v_n \\ v'_2 = t_{12}v_1 + t_{22}v_2 + \cdots + t_{n2}v_n \\ \ldots\ldots\ldots\ldots \\ v'_n = t_{1n}v_1 + t_{2n}v_2 + \cdots + t_{nn}v_n \end{cases}$$

for some $t_{ij} \in K$. Then the matrix $(t_{ij}) \in M_n(K)$, having as its columns the coordinates of the vectors of the basis B' in the basis B, is called the *change matrix* (or *transition matrix*) *from the basis B to the basis B'* and is denoted by $T_{BB'}$.

Remark 3.5.2 (1) B and B' are referred to as the "old" basis and the "new" basis respectively.

(2) The change matrix may be related to the matrix of a linear map as follows. For every $j \in \{1, \ldots, n\}$, the j^{th} column of $T_{BB'}$ consists of the coordinates of $v'_j = 1_V(v'_j)$ in the basis B, hence $T_{BB'} = [1_V]_{B'B}$.

Theorem 3.5.3 *Let V be a vector space over K, and let $B = (v_1, \dots, v_n)$, $B' = (v'_1, \dots, v'_n)$ and $B'' = (v''_1, \dots, v''_n)$ be bases of V. Then*

$$T_{BB''} = T_{BB'} \cdot T_{B'B''}.$$

Proof. Using Theorem 3.4.8, we have

$$T_{BB'} \cdot T_{B'B''} = [1_V]_{B'B} \cdot [1_V]_{B''B'} = [1_V \circ 1_V]_{B''B} = [1_V]_{B''B} = T_{BB''}.$$

We also present a direct proof. Denote $T_{BB'} = (t_{ij}) \in M_n(K)$, $T_{B'B''} = (t'_{jk}) \in M_n(K)$ and $T_{BB''} = (t''_{ik}) \in M_n(K)$. Then we have

$$v'_j = \sum_{i=1}^{n} t_{ij} v_i \,, \quad \forall j \in \{1, \dots, n\},$$

$$v''_k = \sum_{j=1}^{n} t'_{jk} v'_j \,, \quad \forall k \in \{1, \dots, n\} \,,$$

$$v''_k = \sum_{i=1}^{n} t''_{ik} v_i \,, \quad \forall k \in \{1, \dots, n\}.$$

For every $k \in \{1, \dots, n\}$ it follows that

$$v''_k = \sum_{j=1}^{n} t'_{jk} \left(\sum_{i=1}^{n} t_{ij} v_i \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} t_{ij} t'_{jk} \right) v_i.$$

By the uniqueness of writing of each v''_k as a linear combination of the vectors of the basis B, for every $i, k \in \{1, \dots, n\}$ it follows that

$$t''_{ik} = \sum_{j=1}^{n} t_{ij} t'_{jk}.$$

This shows that $T_{BB''} = T_{BB'} \cdot T_{B'B''}$. □

Theorem 3.5.4 *Let V be a vector space over K, and let B and B' be bases of V. Then the change matrix $T_{BB'}$ is invertible and its inverse is the change matrix $T_{B'B}$.*

Proof. Using Theorem 3.5.3 for $B'' = B$, we have

$$T_{BB'} T_{B'B} = T_{BB} = I_n.$$

Using again Theorem 3.5.3 and changing the roles for B, B' and B'' by B', B and B' respectively, we have

$$T_{B'B} T_{BB'} = T_{B'B'} = I_n.$$

Hence $T_{BB'}$ is invertible and $T_{BB'}^{-1} = T_{B'B}$. □

Let us now see how one can use the change matrix from one basis to another in order to compute the coordinates of a vector in different bases or the matrix of a linear map in different bases.

Theorem 3.5.5 *Let V be a vector space over K, let $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_n)$ be bases of V and let $v \in V$. Then*

$$[v]_B = T_{BB'} \cdot [v]_{B'} .$$

Proof. Using Theorem 3.4.4, we have

$$T_{BB'} \cdot [v]_{B'} = [1_V]_{B'B} \cdot [v]_{B'} = [1_V(v)]_B = [v]_B.$$

We also present a direct proof. Consider the writings of the vector $v \in V$ in the two bases B and B', say $v = \sum_{i=1}^n k_i v_i$ and $v = \sum_{j=1}^n k'_j v'_j$ for some $k_i, k'_j \in K$. Since $T_{BB'} = (t_{ij}) \in M_n(K)$, we have

$$v'_j = \sum_{i=1}^n t_{ij} v_i , \quad \forall j \in \{1, \ldots, n\}.$$

It follows that

$$v = \sum_{j=1}^n k'_j \left(\sum_{i=1}^n t_{ij} v_i \right) = \sum_{i=1}^n \left(\sum_{j=1}^n t_{ij} k'_j \right) v_i .$$

By the uniqueness of writing of v as a linear combination of the vectors of the basis B, it follows that $k_i = \sum_{j=1}^n t_{ij} k'_j$, whence $[v]_B = T_{BB'} \cdot [v]_{B'}$. □

Remark 3.5.6 Usually, we are interested in computing the coordinates of a vector v in the new basis B', knowing the coordinates of the same vector v in the old basis B and the change matrix from B to B'. Then by Theorem 3.5.5, we have

$$[v]_{B'} = T_{BB'}^{-1} \cdot [v]_B = T_{B'B} \cdot [v]_B .$$

Example 3.5.7 Consider the bases $E = (e_1, e_2, e_3)$ and $B = (v_1, v_2, v_3)$ of the canonical real vector space $\mathbb{R}^3$, where E is the canonical basis and $v_1 = (0, 1, 1)$, $v_2 = (1, 1, 2)$, $v_3 = (1, 1, 1)$. Let us determine the change matrices from E to B and viceversa. We have

$$\begin{cases} v_1 = \quad e_2 + e_3 \\ v_2 = e_1 + e_2 + 2e_3 \\ v_3 = e_1 + e_2 + e_3 \end{cases}$$

which implies

$$\begin{cases} e_1 = -v_1 \quad + v_3 \\ e_2 = v_1 - v_2 + v_3 \\ e_3 = \quad v_2 - v_3 \end{cases} .$$

Hence we get

$$T_{EB} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} , \quad T_{BE} = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} .$$

We must have $T_{BE} = T_{EB}^{-1}$, so that we could have obtained T_{BE} by computing the inverse of T_{EB}.

Now consider the vector $u = (1, 2, 3)$. Clearly, its coordinates in the canonical basis E are 1, 2 and 3. By Theorem 3.5.5, it follows that

$$[u]_B = T_{BE} \cdot [u]_E = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} .$$

Hence the coordinates of u in the basis B are 1, 1 and 0.

Next we give a theorem relating the matrices of a linear map in different bases.

Theorem 3.5.8 *Let $f \in \mathrm{Hom}_K(V, V')$, let B_1 and B_2 be bases of V and let B_1' and B_2' be bases of V'. Then*

$$[f]_{B_2 B_2'} = T^{-1}_{B_1' B_2'} \cdot [f]_{B_1 B_1'} \cdot T_{B_1 B_2} .$$

Proof. Using Theorem 3.4.8, we have

$$\begin{aligned} T^{-1}_{B_1' B_2'} \cdot [f]_{B_1 B_1'} \cdot T_{B_1 B_2} &= T_{B_2' B_1'} \cdot [f]_{B_1 B_1'} \cdot T_{B_1 B_2} \\ &= [1_V]_{B_1' B_2'} \cdot [f]_{B_1 B_1'} \cdot [1_V]_{B_2 B_1} \\ &= [1_V \circ f \circ 1_V]_{B_2 B_2'} \\ &= [f]_{B_2 B_2'}, \end{aligned}$$

which shows the result. □

Corollary 3.5.9 *Let $f \in \mathrm{End}_K(V)$, and let B and B' be bases of V. Then*

$$[f]_{B'} = T^{-1}_{BB'} \cdot [f]_B \cdot T_{BB'} .$$

Proof. This follows by Theorem 3.5.8 with $B_1 = B_1' = B$ and $B_2 = B_2' = B'$. □

Example 3.5.10 Consider the bases $E = (e_1, e_2, e_3)$ and $B = (v_1, v_2, v_3)$ of the canonical real vector space $\mathbb{R}^3$, where E is the canonical basis and $v_1 = (0, 1, 1)$, $v_2 = (1, 1, 2)$, $v_3 = (1, 1, 1)$. Also let $f \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^3)$ be defined by

$$f(x, y, z) = (x + y, y - z, z + x) , \ \forall (x, y, z) \in \mathbb{R}^3 .$$

Let us determine the matrix of f in the basis E and in the basis B. We have

$$\begin{cases} f(e_1) = (1, 0, 1) = e_1 + e_3 \\ f(e_2) = (1, 1, 0) = e_1 + e_2 \\ f(e_3) = (0, -1, 1) = -e_2 + e_3 \end{cases}$$

which implies that

$$[f]_E = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} .$$

Using Corollary 3.5.9 and the change matrices T_{EB} and T_{BE}, that we have determined in Example 3.5.7, we have

$$[f]_B = T_{EB}^{-1} \cdot [f]_E \cdot T_{EB} = T_{BE} \cdot [f]_E \cdot T_{EB}$$
$$= \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -3 & -2 \\ 1 & 4 & 2 \\ 0 & -2 & 0 \end{pmatrix}.$$

It is worth to be mentioned that we could have reached the same result using the definition of the matrix of a linear map and expressing the vectors $f(v_1)$, $f(v_2)$ and $f(v_3)$ as linear combinations of the vectors v_1, v_2 and v_3 of the basis B.

3.6 Linear systems of equations

In this section we present some aspects of the theory of linear systems of equations which can be approached by using tools of Linear Algebra.

As usual, throughout this section K will be a field. When needed, we use superior indices to denote vectors in K^n and inferior indices to denote their components. For instance, $x^0 = (x_1^0, \dots, x_n^0) \in K^n$.

Definition 3.6.1 Consider a *linear system* of m equations with n unknowns $x_1, \dots, x_n$:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \qquad (S)$$

where the coefficients are $a_{ij}, b_i \in K$ $(i = 1, \dots, m,\ j = 1, \dots, n)$.
If $b_1 = \cdots = b_m = 0$, then the system (S) is called *homogeneous* and is denoted by (S_0).
The matrix $A = (a_{ij}) \in M_{m,n}(K)$ is called the *matrix of the system* (S).
The matrix

$$\bar{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

is called the *augmented* (or *extended*) *matrix of the system* (S).

Next we present two equivalent forms of a linear system of equations.
First, denote

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Then the systems (S) and (S_0) can be written as:

$$A \cdot x = b\,, \qquad (S)$$

$$A \cdot x = 0\,. \tag{S_0}$$

Furthermore, we know that there exists a bijective correspondence between K-linear maps and matrices. Thus, since $A \in M_{m,n}(K)$, there exists $f_A \in \mathrm{Hom}_K(K^n, K^m)$ such that $[f_A]_{EE'} = A$, where E and E' are the canonical bases in K^n and K^m respectively.

Denoting $x = (x_1, \ldots, x_n) \in K^n$ and $b = (b_1, \ldots, b_m) \in K^m$, it follows that

$$[f_A(x)]_{E'} = [f_A]_{EE'} \cdot [x]_E = A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = [b]_{E'}\,.$$

Hence $f_A(x) = b$. Thus, the systems (S) and (S_0) can be written as:

$$f_A(x) = b\,, \tag{S}$$

$$f_A(x) = 0\,. \tag{S_0}$$

Remark 3.6.2 (1) Thus, for a linear system of equations we have three equivalent forms, namely: the classical one with coefficients and unknowns, the one using matrices and the one using the corresponding linear map.

(2) We have denoted by x and b first column-matrices and then row-matrices to get nicer results, without using any transposed matrices.

Definition 3.6.3 An element $x^0 \in M_{n1}(K)$ ($x^0 \in K^n$) is called a:
(1) *(particular) solution* of (S) if $A \cdot x^0 = b$ (or equivalently $f_A(x^0) = b$).
(2) *(particular) solution* of (S_0) if $A \cdot x^0 = 0$ (or equivalently $f_A(x^0) = 0$).

Denote the sets of solutions of (S) and (S_0) by

$$S = \{x^0 \in M_{n1}(K) \mid A \cdot x^0 = b\} \quad \text{or} \quad S = \{x^0 \in K^n \mid f_A(x^0) = b\}\,,$$

$$S_0 = \{x^0 \in M_{n1}(K) \mid A \cdot x^0 = 0\} \quad \text{or} \quad S_0 = \{x^0 \in K^n \mid f_A(x^0) = 0\}\,.$$

Theorem 3.6.4 *The set S_0 of solutions of the homogeneous linear system of equations (S_0) is a subspace of the canonical vector space K^n over K and*

$$\dim S_0 = n - \mathrm{rank}(A)\,.$$

Proof. Since

$$S_0 = \{x^0 \in K^n \mid f_A(x^0) = 0\} = \mathrm{Ker} f_A$$

and the kernel of a linear map is always a subspace of the domain vector space, it follows that $S_0 \leq K^n$. Now by the First Dimension Theorem, it follows that

$$\dim S_0 = \dim(\mathrm{Ker} f_A) = \dim K^n - \dim(\mathrm{Im} f_A) = n - \mathrm{rank}(f_A) = n - \mathrm{rank}(A)\,,$$

which finishes the proof. □

Theorem 3.6.5 *If $x^1 \in S$ is a particular solution of the system (S), then*

$$S = x^1 + S_0 = \{x^1 + x^0 \mid x^0 \in S_0\}\,.$$

Proof. Since $x^1 \in S$, we have $Ax^1 = b$. We prove the requested equality by double inclusion.

First, let $x^2 \in S$. Then

$$Ax^2 = b \Longrightarrow Ax^2 = Ax^1 \Longrightarrow A(x^2 - x^1) = 0 \Longrightarrow x^2 - x^1 \in S_0 \Longrightarrow x^2 \in x^1 + S_0\,.$$

Conversely, let $x^2 \in x^1 + S_0$. There exists $x^0 \in S_0$ such that $x^2 = x^1 + x^0$. Then:

$$Ax^2 = A(x^1 + x^0) = Ax^1 + Ax^0 = b + 0 = b,$$

and consequently $x^2 \in S$.

Therefore, $S = x^1 + S_0$. □

Remark 3.6.6 By Theorem 3.6.5, the general solution of the system (S) can be obtained by knowing the general solution of the homogeneous system (S_0) and a particular solution of (S).

In the sequel, we are going to see when a linear system of equations has a solution.

Definition 3.6.7 The system (S) is called *compatible* (or *consistent*) if it has at least one solution. A compatible system (S) is called *determinate* if it has a unique solution.

Remark 3.6.8 (1) The system (S) is compatible if and only if $\exists x^0 \in K^n$ such that $f_A(x^0) = b$ if and only if $b \in \mathrm{Im} f_A$.

(2) The system (S_0) is compatible if and only if $\exists x^0 \in K^n$ such that $f_A(x^0) = 0$ if and only if $0 \in \mathrm{Im} f_A$. But the last condition always holds, since $\mathrm{Im} f_A$ is a subspace of K^m. Hence any homogeneous linear system of equations is compatible, having at least the zero (trivial) solution.

Theorem 3.6.9 *The system (S_0) has a non-zero solution if and only if* $\mathrm{rank}(A) < n$.

Proof. By Theorem 3.6.4, we have

$$S_0 = \mathrm{Ker} f_A \neq \{0\} \Longleftrightarrow \dim S_0 \neq 0 \Longleftrightarrow n - \mathrm{rank}(A) \neq 0 \Longleftrightarrow \mathrm{rank}(A) < n\,,$$

which proves the result. □

Corollary 3.6.10 *Let $A \in M_n(K)$. Then*

$$S_0 = \{0\} \Longleftrightarrow \mathrm{rank}(A) = n \Longleftrightarrow \det(A) \neq 0\,.$$

Definition 3.6.11 If $A \in M_n(K)$ and $\det(A) \neq 0$, then the system (S) is called a *Cramer system.*

Theorem 3.6.12 *A Cramer system $Ax = b$ has a unique solution. More precisely, its unique solution $(x_1, \ldots, x_n)$ is computed by*

$$x_i = \det(A)^{-1} \cdot d_i,$$

where d_i is the determinant obtained from $\det(A)$ by replacing its i^{th} column by the column b for every $i \in \{1, \ldots, n\}$.

Proof. The matrix of a Cramer system is an invertible matrix $A \in M_n(K)$. Then we deduce that $x = A^{-1}b$ is the unique solution. Moreover, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A^{-1} \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \det(A)^{-1} \cdot A^* \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \det(A)^{-1} \cdot \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}.$$

Hence $x_i = \det(A)^{-1} \cdot d_i$ for every $i \in \{1, \ldots, n\}$. □

Corollary 3.6.13 *A homogeneous Cramer system has only the zero solution.*

Let us now give two classical compatibility theorems.

Theorem 3.6.14 (Kronecker-Capelli Theorem) *The system (S) is compatible if and only if* $\operatorname{rank}(\bar{A}) = \operatorname{rank}(A)$.

Proof. Let $(e_1, \ldots, e_n)$ be the canonical basis of the canonical vector space K^n over K and denote by $a^1, \ldots, a^n$ the columns of the matrix A. Then we have

$$\begin{aligned}
(S) \text{ is compatible} &\iff \exists x^0 \in K^n : \ f_A(x^0) = b \\
&\iff b \in \operatorname{Im} f_A \\
&\iff b \in f_A(\langle e_1, \ldots, e_n \rangle) \\
&\iff b \in \langle f_A(e_1), \ldots, f_A(e_n) \rangle \\
&\iff b \in \langle a^1, \ldots, a^n \rangle \\
&\iff \langle a^1, \ldots, a^n, b \rangle = \langle a^1, \ldots, a^n \rangle \\
&\iff \dim\langle a^1, \ldots, a^n, b \rangle = \dim\langle a^1, \ldots, a^n \rangle \\
&\iff \operatorname{rank}(\bar{A}) = \operatorname{rank}(A)\,,
\end{aligned}$$

which proves the result. □

Definition 3.6.15 A minor d_p of the matrix A is called a *principal determinant* if $d_p \neq 0$ and d_p has the order $\text{rank}(A)$.
We call *characteristic determinants associated to a principal determinant* d_p of A the minors of the augmented matrix $\bar{A}$ obtained by completing the matrix of d_p with a column containing the corresponding constants b_i and a row containing the corresponding elements of a row of $\bar{A}$.

Now we give the second compatibility theorem.

Theorem 3.6.16 (Rouché Theorem) *The system (S) is compatible if and only if all the characteristic determinants associated to a principal determinant are zero.*

Proof. $\boxed{\Longrightarrow}$ Suppose that the system (S) is compatible. Then by Theorem 3.6.14, $\text{rank}(\bar{A}) = \text{rank}(A)$. Denote this rank by r. Then there exists a principal determinant d_p of order r. Since $r = \text{rank}(A)$, any determinant of order $r+1$ is zero and consequently any characteristic determinant associated to d_p is zero.

$\boxed{\Longleftarrow}$ Suppose that all the characteristic determinants associated to a principal determinant are zero. Denote $r = \text{rank}(A)$. Then $r \leq \text{rank}(\bar{A})$ and there exists a non-zero minor, actually a principal determinant, d_r of A. But d_r is also a minor of $\bar{A}$ of order r.

Now let d_{r+1} be a minor of $\bar{A}$ of order $r+1$. We have two possibilities, namely either d_{r+1} is a minor of $\bar{A}$ or d_{r+1} is just a minor of A. In the first case, d_{r+1} is a characteristic determinant associated to the principal determinant d_r, hence $d_{r+1} = 0$ by hypothesis. In the second case, we have $d_{r+1} = 0$, since $\text{rank}(A) = r$.

Thus, $\text{rank}(\bar{A}) = r = \text{rank}(A)$. Now by Theorem 3.6.14, (S) is compatible. □

3.7 Gauss method

In this section we briefly present a very useful practical method to solve linear systems of equations, called the *Gauss method* (or *Gaussian elimination*).

In the sequel, suppose that $m \leq n$, that is, we talk about systems with less equations than unknowns. In fact, this is the interesting case.

The **Gauss method** consists of the following steps:

(1) *Write the augmented matrix $\bar{A}$ of the system (S).*

(2) *Apply elementary operations on rows for $\bar{A}$ to get to an echelon form A'.*

(3) *Use the Kronecker-Capelli Theorem to decide if the system is compatible or not.*

(4) *If compatible, write and solve the system corresponding to the echelon form, starting with the last equation.*

Remark 3.7.1 (1) Actually, the Gauss method simulates working with equations. When we apply an elementary operation on the rows of $\bar{A}$, say multiply a row by a scalar and

add it to another row, in fact we multiply an equation by a scalar and add it to another equation. That is why it is important to apply elementary operations only on rows, in order not to interchange the order of the unknowns.

(2) The initial system and the system corresponding to the echelon form are equivalent, that is, they have the same solutions. The great advantage is that the last system can be easily solved, starting with the last equation.

(3) The Gauss method includes checking compatibility, done by the Kronecker-Capelli Theorem.

(4) If the system is compatible, we have a principal determinant of order $r = \text{rank}(\bar{A}) = \text{rank}(A)$ and it is possible to continue the procedure on the matrix A' to get to a diagonal form having r elements on the principal diagonal and all the other elements zero. Then, when writing the equivalent system, in fact we directly get the solution. This completion of the Gauss method is called the *Gauss-Jordan method*.

Example 3.7.2 (a) Consider the system

$$\begin{cases} x + y - z = 2 \\ 3x + 2y - 2z = 6 \\ -x + y + z = 0 \end{cases}$$

with real coefficients. Then its augmented matrix is

$$\bar{A} = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 3 & 2 & -2 & 6 \\ -1 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 2 & 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 \end{pmatrix}.$$

Since $\text{rank}(\bar{A}) = 3 = \text{rank}(A)$, the system is determinate compatible. The equivalent system is

$$\begin{cases} x + y - z = 2 \\ -y + z = 0 \\ 2z = 2. \end{cases}$$

We immediately get the solution $x = 2$, $y = 1$, $z = 1$.

We could have got to the same solution by continuing with the Gauss-Jordan method. Indeed,

$$\bar{A} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 2 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

whence we immediately read the solution $x = 2$, $y = 1$, $z = 1$.

(b) Consider the system

$$\begin{cases} x + y + z = 0 \\ x + 4y + 10z = 3 \\ 2x + 3y + 5z = 1 \end{cases}$$

with real coefficients. Then its augmented matrix is

$$\bar{A} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 4 & 10 & 3 \\ 2 & 3 & 5 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 3 & 9 & 3 \\ 0 & 1 & 3 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 1 & 3 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $\operatorname{rank}(\bar{A}) = 2 = \operatorname{rank}(A)$, the system is non-determinate compatible. The equivalent system is

$$\begin{cases} x + y + z = 0 \\ \quad y + 3z = 1. \end{cases}$$

Then x and y are principal unknowns and z is a secondary unknown. We immediately get the solution

$$\begin{cases} x = 2z - 1 \\ y = 1 - 3z \\ z \in \mathbb{R}. \end{cases}$$

We could have got to the same solution by continuing with the Gauss-Jordan method. Indeed,

$$\bar{A} \sim \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -2 & -1 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The equivalent system is

$$\begin{cases} x - 2z = -1 \\ y + 3z = 1 \end{cases}$$

whence we get the solution

$$\begin{cases} x = 2z - 1 \\ y = 1 - 3z \\ z \in \mathbb{R}. \end{cases}$$

(c) Consider the system

$$\begin{cases} x + y + z = 3 \\ x - y + z = 1 \\ -2x + y - 2z = -3 \\ x + z = 4 \end{cases}$$

with real coefficients. Then its augmented matrix is

$$\bar{A} = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & -1 & 1 & 1 \\ -2 & 1 & -2 & -3 \\ 1 & 0 & 1 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -2 & 0 & -2 \\ 0 & 3 & 0 & 3 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 3 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Since $\operatorname{rank}(\bar{A}) = 3$ and $\operatorname{rank}(A) = 2$, the system is not compatible.

Remark 3.7.3 Following [20], let us analyze how many operations are required to solve a linear system of equations $Ax = b$ with $A \in M_n(K)$ invertible by the Gauss method. We assume that the operations of interchanging rows are negligible.

Let us first compute the cost of the reduction to a triangular echelon form having all elements on the principal diagonal equal to 1. We may assume that $a_{11} \neq 0$. We reduce a_{11} to 1 by dividing the first row of A by a_{11}. Then we produce zeros on the first column under the $(1,1)$-entry. For each of the $n-1$ rows we need n multiplications and n additions. Adding the corresponding operations for b, we have 1 more division, $n-1$ more multiplications and $n-1$ more additions. After finishing working with the first row, we move on to the second row, and so on til we get the required triangular form with elements 1 on the principal diagonal. Counting up the operations we have

- $n + (n-1) + \cdots + 1$ divisions on A and n divisions on b;
- $n(n-1) + (n-1)(n-2) + \cdots + 2 \cdot 1$ multiplications on A and $(n-1) + \cdots + 1$ multiplications on b;
- $n(n-1) + (n-1)(n-2) + \cdots + 2 \cdot 1$ additions on A and $(n-1) + \cdots + 1$ additions on b.

So far we have

- $\frac{n(n+1)}{2} + n$ divisions;
- $\frac{n^3-n}{3} + \frac{n(n-1)}{2}$ multiplications;
- $\frac{n^3-n}{3} + \frac{n(n-1)}{2}$ additions.

Now let us compute the cost of substitutions in the reduced triangular system. From the last equation we already have the unknown x_n. For the substitution on the previous but last equation to find x_{n-1} we need 1 multiplication and 1 addition. Continuing the procedure, for the first equation to find x_1 we need $n-1$ multiplications and $n-1$ additions. Counting up the operations, we have

- $(n-1) + \cdots + 1 = \frac{n(n-1)}{2}$ multiplications;
- $(n-1) + \cdots + 1 = \frac{n(n-1)}{2}$ additions.

Adding up the numbers of operations from the above two stages, it turns out that one needs:

(1) $\frac{n(n+1)}{2} + n$ divisions;

(2) $\frac{n^3-n}{3} + n(n-1)$ multiplications;

(3) $\frac{n^3-n}{3} + n(n-1)$ additions.

Hence the order of magnitude is $\frac{2}{3}n^3$ operations.

Extra: LU decomposition and Gauss method

We have already seen that LU decomposition may be used for speeding up computations of determinants of square matrices and inverses of invertible matrices. Now let us assume that there is an LU decomposition $A = L \cdot U \in M_n(K)$. Following [21], a system

$$Ax = b$$

of linear equations may be split into two triangular systems

$$Lc = b, \quad Ux = c.$$

Then $Ax = LUx = Lc = b$. The algorithm for solving the initial system needs two functions:

1. **Decompose**: given A, find L and U.
2. **Solve**: given L, U and b, find x (by the Gauss method).

If A does not need to be modified, then **Decompose** needs:

(1) $\frac{n(n-1)}{2} + n$ divisions;

(2) $\frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6}$ multiplications;

(3) $\frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6}$ additions.

Solve finds solution for the two triangular systems in $\frac{n^2}{2}$ steps each, that is, n^2 steps overall.
Hence solving a linear system of equations by this approach, the order of magnitude is $\frac{2}{3}n^3$ operations, as in the case of the usual Gauss method. Usually, the practical advantage of using LU decomposition is that one may have the same matrix A (for which an LU decomposition is performed only once) and several different new b's for which one only uses the function **Solve**.

Extra: Simple authentication scheme

Let us consider the following simple authentication scheme from cryptography, following [11]. We denote by E the canonical basis of the canonical vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$.

- The password is a vector $v = (x_1, \ldots, x_n) \in \mathbb{Z}_2^n$.
- As a challenge, Computer sends a random vector $u = (u_1, \ldots, u_n) \in \mathbb{Z}_2^n$.
- As the response, Human sends back the dot-product vector

$$u \cdot v = u_1x_1 + \cdots + u_nx_n \in \mathbb{Z}_2.$$

- The challenge-response interaction is repeated until Computer is convinced that Human knows password v.

Eve eavesdrops and learns m pairs (a_1, b_1), ..., (a_m, b_m) such that each b_i is the correct response to challenge a_i. For every $i \in \{1, \ldots, m\}$, denote $a_i = (a_{i1}, \ldots, a_{in})$.
Then the password $v = (x_1, \ldots, x_n)$ is a solution of the linear system of equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

Once the rank of the matrix of the system reaches n, the solution is unique, and Eve can use the Gauss method to find it, obtaining the password.

3.8 Eigenvectors and eigenvalues

The study of endomorphisms of vector spaces also makes use of vectors whose images are just scalar multiples of themselves, in other words vectors that are "stretched" by an endomorphism. They are the subject of the present section.

Definition 3.8.1 Let $f \in \mathrm{End}_K(V)$. A non-zero vector $v \in V$ is called an *eigenvector of* f if there exists $\lambda \in K$ such that $f(v) = \lambda \cdot v$. Here λ is called an *eigenvalue of* f.

Remark 3.8.2 Clearly, each eigenvector has a unique corresponding eigenvalue. But different eigenvectors may have the same corresponding eigenvalue.

For $f \in \mathrm{End}_K(V)$, denote

$$V(\lambda) = \{v \in V \mid f(v) = \lambda v\},$$

that is, the set consisting of the zero vector and the eigenvectors of f with eigenvalue λ.

Theorem 3.8.3 *Let $f \in \mathrm{End}_K(V)$ and let λ be an eigenvalue of f. Then $V(\lambda)$ is a subspace of V.*

Proof. Clearly, $0 \in V(\lambda)$, hence $V(\lambda) \neq \emptyset$. Now let $k_1, k_2 \in K$ and $v_1, v_2 \in V(\lambda)$. Then we have $f(v_1) = \lambda v_1$ and $f(v_2) = \lambda v_2$. It follows that

$$\begin{aligned} f(k_1v_1 + k_2v_2) &= k_1f(v_1) + k_2f(v_2) \\ &= k_1(\lambda v_1) + k_2(\lambda v_2) \\ &= (k_1\lambda)v_1 + (k_2\lambda)v_2 \\ &= \lambda(k_1v_1 + k_2v_2). \end{aligned}$$

Hence, $k_1v_1 + k_2v_2 \in V(\lambda)$ and consequently, $V(\lambda)$ is a subspace of V. □

Definition 3.8.4 Let $f \in \mathrm{End}_K(V)$ and let λ be an eigenvalue of f. Then $V(\lambda)$ is called the *eigenspace* (or the *characteristic subspace*) of λ with respect to f.

Theorem 3.8.5 *Let $f \in \mathrm{End}_K(V)$ and let λ be an eigenvalue of f. Then:*
(i) $\forall k \in \mathbb{N}$, λ^k is an eigenvalue of $f^k = \underbrace{f \circ \cdots \circ f}_{k \text{ times}}$.
(ii) If $f \in \mathrm{Aut}_K(V)$, then $\forall k \in \mathbb{Z}$, λ^k is an eigenvalue of f^k.

Proof. (i) Trivially, $\lambda^0 = 1$ is an eigenvalue of $f^0 = 1_V$. Let us now consider $k \in \mathbb{N}^*$. We are going to prove the property by induction on k. Clearly, $\lambda^1 = \lambda$ is an eigenvalue of $f^1 = f$. Assume that the property holds for k and prove it for $k+1$. Since λ^k is an eigenvalue of f^k, $\exists v \in V$, $v \neq 0$, such that $f^k(v) = \lambda^k v$, whence it follows that

$$f^{k+1}(v) = f(f^k(v)) = f(\lambda^k v) = \lambda^k f(v) = \lambda^k(\lambda v) = \lambda^{k+1} v\,.$$

Therefore, we have proved that λ^k is an eigenvalue of f^k, $\forall k \in \mathbb{N}^*$.

(ii) In order to show the property for negative integers, it is enough to prove it for $k = -1$ and then apply the first part. We have

$$f(v) = \lambda v \Longrightarrow f^{-1}(f(v)) = f^{-1}(\lambda v) \Longrightarrow v = \lambda f^{-1}(v) \Longrightarrow f^{-1}(v) = \lambda^{-1} v\,.$$

This shows the result. □

The next theorem offers the essence of the practical method to determine eigenvalues and eigenvectors.

Theorem 3.8.6 *Let V be a vector space over K, B a basis of V and $f \in \mathrm{End}_K(V)$ with the matrix $[f]_B = A = (a_{ij}) \in M_n(K)$. Then $\lambda \in K$ is an eigenvalue of f if and only if*

$$\det(A - \lambda \cdot I_n) = 0 \tag{1}$$

Proof. The element $\lambda \in K$ is an eigenvalue of f if and only if there exists a non-zero $v \in V$ such that $f(v) = \lambda v$. Consider $[v]_B = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Then it follows that

$$\begin{aligned}
&f(v) = \lambda v \iff f(v) - \lambda v = 0 \iff (f - \lambda \cdot 1_V)(v) = 0 \iff [(f - \lambda \cdot 1_V)(v)]_B = [0]_B \\
&\iff [f - \lambda \cdot 1_V]_B \cdot [v]_B = [0]_B \iff ([f]_B - \lambda \cdot [1_V]_B) \cdot [v]_B = [0]_B \\
&\iff (A - \lambda \cdot I_n) \cdot [v]_B = [0]_B \\
&\iff \begin{pmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - \lambda \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \\
&\iff \begin{cases} (a_{11} - \lambda)x_1 + a_{12}x_2 \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \cdots + a_{2n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n = 0. \end{cases}
\end{aligned}$$

Then λ is an eigenvalue of f if and only if the final system (S) of linear equations has a non-zero solution if and only if its determinant $\det(A - \lambda \cdot I_n)$ is zero. □

Definition 3.8.7 The equality (1) is called the *characteristic equation* and the system (S) is called the *characteristic system*. The determinant $\det(A - \lambda I_n)$ may be seen as a polynomial $p_A(\lambda)$ in λ and it is called the *characteristic polynomial of* f with respect to A (or the *characteristic polynomial* of A).

Now a question arises naturally: if we take another basis B' of V and use the matrix $[f]_{B'}$, do we get the same eigenvalues and eigenvectors of f? We will show that the answer is positive. But first we introduce the following notion.

Definition 3.8.8 Let $A, B \in M_n(K)$. Then A is called *similar to* B if there exists an invertible matrix $P \in M_n(K)$ such that $B = P^{-1}AP$.

Lemma 3.8.9 *The similarity relation is an equivalence relation on* $M_n(K)$.

Proof. For every $A \in M_n(K)$, we have $A = I_n^{-1}AI_n$, hence A is similar to itself.

Let $A, B, C \in M_n(K)$ be such that A is similar to B and B is similar to C. Then $B = P^{-1}AP$ and $C = Q^{-1}BQ$ for some invertible matrices $P, Q \in M_n(K)$. It follows that $C = (PQ)^{-1}A(PQ)$, where PQ is invertible. Hence A is similar to C.

Let $A, B \in M_n(K)$ be such that A is similar to B. Then $B = P^{-1}AP$ for some invertible matrix $P \in M_n(K)$. It follows that $A = PBP^{-1}$, and thus $A = Q^{-1}BQ$, where $Q = P^{-1} \in M_n(K)$ is invertible. Hence B is similar to A.

Hence the relation of similarity on $M_n(K)$ is reflexive, transitive and symmetric, and thus, it is an equivalence relation. □

Remark 3.8.10 Let V be a vector space over K with bases B, B', and let $f \in \mathrm{End}_K(V)$. By Corollary 3.5.9, we have $[f]_{B'} = T_{BB'}^{-1} \cdot [f]_B \cdot T_{BB'}$. This shows that $[f]_{B'}$ and $[f]_B$ are similar. Hence the matrices of an endomorphism in different bases are similar matrices.

Theorem 3.8.11 *Let* $A, B \in M_n(K)$ *be similar. Then* $p_A(\lambda) = p_B(\lambda)$.

Proof. Since $A, B \in M_n(K)$ are similar, we have $B = P^{-1}AP$ for some invertible matrix $P \in M_n(K)$. Then

$$\begin{aligned} p_B(\lambda) &= \det(B - \lambda I_n) = \det(P^{-1}AP - \lambda I_n P^{-1}P) = \det(P^{-1}(A - \lambda I_n)P) \\ &= \det(P^{-1}) \cdot \det(A - \lambda I_n) \cdot \det(P) = \det(A - \lambda I_n) = p_A(\lambda)\,, \end{aligned}$$

which proves the result. □

Corollary 3.8.12 *Let* V *be a vector space over* K, B *and* B' *bases of* V *and* $f \in \mathrm{End}_K(V)$ *with the matrices* $[f]_B = A \in M_n(K)$ *and* $[f]_{B'} = A' \in M_n(K)$. *Then* $p_A(\lambda) = p_{A'}(\lambda)$.

Proof. This follows by Remark 3.8.10 and Theorem 3.8.11. □

Remark 3.8.13 (1) Therefore, the eigenvalues and the eigenvectors *do not depend* on the basis chosen for writing the matrix of the endomorphism. Of course, the matrices might be different, but in the end we get the same characteristic polynomial. Consequently, we can say that the eigenvalues of an endomorphism (or simply, of a matrix) are just the roots in K of its unique characteristic polynomial.

(2) If V is a vector space over K with $\dim V = n$ and $f \in \mathrm{End}_K(V)$, then the degree of the characteristic polynomial of f is n, hence f may have at most n eigenvalues. If $K = \mathbb{C}$, then by the Fundamental Theorem of Algebra f has exactly n eigenvalues, not necessarily distinct.

(3) A non-zero vector $v \in K^n$ is an eigenvector of a matrix $A \in M_n(K)$ if and only if there exists $\lambda \in K$ such that $A[v]_E = \lambda[v]_E$, where E is the canonical basis of the canonical vector space K^n over K. In this case, λ is an eigenvalue of A.

Example 3.8.14 Let $f \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^3)$ be defined by

$$f(x, y, z) = (2x, y + 2z, -y + 4z)\,,\ \forall (x, y, z) \in \mathbb{R}^3\,.$$

We write its matrix in the simplest basis, namely in the canonical basis E of $\mathbb{R}^3$. Then

$$[f]_E = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 4 \end{pmatrix}\,.$$

The characteristic polynomial is $p(\lambda) = -(\lambda-2)^2(\lambda-3)$, so the eigenvalues are $\lambda_1 = \lambda_2 = 2$ and $\lambda_3 = 3$.

Let us take first $\lambda_1 = \lambda_2 = 2$. An eigenvector (x_1, x_2, x_3) is a non-zero solution of the characteristic system

$$\begin{pmatrix} 2-\lambda_1 & 0 & 0 \\ 0 & 1-\lambda_1 & 2 \\ 0 & -1 & 4-\lambda_1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\,,$$

that is,

$$\begin{cases} -x_2 + 2x_3 = 0 \\ -x_2 + 2x_3 = 0 \end{cases}\,.$$

Then $x_2 = 2x_3$ and $x_1, x_3 \in \mathbb{R}$, whence

$$V(2) = \{(x_1, 2x_3, x_3) \mid x_1, x_3 \in \mathbb{R}\} = \langle (1, 0, 0), (0, 2, 1) \rangle.$$

Any non-zero vector in $V(2)$ is an eigenvector of f with the associated eigenvalue $\lambda_1 = \lambda_2 = 2$.

Consider now $\lambda_3 = 3$. The corresponding characteristic system is

$$\begin{pmatrix} 2-\lambda_3 & 0 & 0 \\ 0 & 1-\lambda_3 & 2 \\ 0 & -1 & 4-\lambda_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

that is,

$$\begin{cases} -x_1 = 0 \\ -2x_2 + 2x_3 = 0 \\ -x_2 + x_3 = 0 \end{cases}.$$

We get the solution $x_1 = 0$, $x_2 = x_3$ and $x_3 \in \mathbb{R}$. Then

$$V(3) = \{(0, x_3, x_3) \mid x_3 \in \mathbb{R}\} = \langle (0, 1, 1) \rangle.$$

Any non-zero vector in $V(3)$ is an eigenvector of f with the associated eigenvalue $\lambda_3 = 3$.

Extra: PageRank

PageRank is a number assigned by Google to each web page. Pages with higher rank come higher in search results. We describe a simplified version, following [22].

- Consider pages $S_1, \ldots, S_n$, with some links between them. A link from S_j to S_i is a vote by S_j that S_i is important.
- Links from important pages should count for more (because the probability of visiting S_i will clearly increase); links from pages with many links should count for less (because that will decrease the probability that we click the one that leads to S_i).
- We want rankings $r_1, \ldots, r_n \geq 0$, normalized so that $\sum_{i=1}^{n} r_i = 1$.
- Say S_j links to N_j different pages, and assume $N_j > 0$. We use the rule: a link from S_j to S_i contributes $\frac{r_j}{N_j}$ to r_i.
- Thus, for every $i \in \{1, \ldots, n\}$, the following consistency condition should be satisfied:
$$r_i = \sum_{j \in J_i} \frac{r_j}{N_j},$$
where $J_i = \{j \in \{1, \ldots, n\} \mid \text{page } S_j \text{ links to page } S_i\}$.
- Define the matrix $P = (p_{ij}) \in M_n(\mathbb{R})$ by
$$p_{ij} = \begin{cases} \frac{1}{N_j} & \text{if there is a link from } S_j \text{ to } S_i \\ 0 & \text{otherwise.} \end{cases}$$
- Hence, for every $i \in \{1, \ldots, n\}$, the consistency condition becomes:
$$r_i = \sum_{j \in J_i} p_{ij} r_j.$$
- But this is equivalent to the matrix equation $Pr = r$, and thus r is an eigenvector of the matrix P with eigenvalue 1.

3.9 Cayley-Hamilton Theorem

We begin with a result on sums and products of eigenvalues. For a matrix $A \in M_n(K)$, we denote by $\mathrm{Tr}(A)$ the *trace* of A, that is, the sum of the elements of the principal diagonal of A.

Theorem 3.9.1 *Let $A \in M_n(K)$ having eigenvalues $\lambda_1, \ldots, \lambda_n$. Then:*
(i) $\lambda_1 + \cdots + \lambda_n = \mathrm{Tr}(A)$.
(ii) $\lambda_1 \cdots \lambda_n = \det(A)$.

Proof. We have

$$p_A(\lambda) = \det(A - \lambda I_n) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} - \lambda & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} - \lambda \end{vmatrix}.$$

If A_{ij} denotes the matrix obtained from A by crossing out row i and column j, then by the cofactor expansion along the first column of the matrix, we have

$$p_A(\lambda) = (a_{11} - \lambda)\det(A_{11}) + \sum_{i=2}^{n} a_{i1}(-1)^{i+1}\det(A_{i1}),$$

where A_{i1} does not contain the diagonal elements $a_{11} - \lambda$ and $a_{ii} - \lambda$ for $i \in \{2, \ldots, n\}$. By expanding A_{11} as above and continuing the proccess, we eventually see that

$$(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$$

is the only term in the expansion of $p_A(\lambda)$ involving more than $n - 2$ elements of the principal diagonal. It follows that the leading coefficient of λ^n in $p_A(\lambda)$ is $(-1)^n$. Hence we have

$$p_A(\lambda) = (-1)^n(\lambda - \lambda_1) \cdots (\lambda - \lambda_n) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda).$$

It follows that

$$\lambda_1 \cdots \lambda_n = p_A(0) = \det(A).$$

From the expression $(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$, the coefficient of $(-\lambda)^{n-1}$ is $\sum_{i=1}^{n} a_{ii}$. On the other hand, from $p_A(\lambda) = (\lambda_1 - \lambda) \cdots (\lambda_n - \lambda)$ the same coeffcient is $\sum_{i=1}^{n} \lambda_i$. It follows that

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} = \mathrm{Tr}(A),$$

which finishes the proof. □

Corollary 3.9.2 *Let $A \in M_2(K)$. Then the characteristic polynomial of A is*

$$p_A(\lambda) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A).$$

The following famous theorem involves the characteristic polynomial.

Theorem 3.9.3 (Cayley-Hamilton Theorem) *Every matrix $A \in M_n(K)$ is a root of its characteristic polynomial.*

Proof. We have $p_A(\lambda) = \det(A - \lambda I_n)$, say

$$p_A(\lambda) = \alpha_n \lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_0 \,.$$

Then

$$p_A(A) = \alpha_n A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_0 I_n \,.$$

Denote by $(A - \lambda I_n)^*$ the adjugate matrix of $A - \lambda I_n$. Then

$$(A - \lambda I_n)(A - \lambda I_n)^* = \det(A - \lambda I_n) \cdot I_n \,,$$

hence

$$(A - \lambda I_n)(A - \lambda I_n)^* = p_A(\lambda) \cdot I_n \,.$$

But we can write

$$(A - \lambda I_n)^* = A_{n-1}\lambda^{n-1} + \cdots + A_1\lambda + A_0$$

for some matrices $A_i \in M_n(K)$. Then we have the following equality:

$$(A - \lambda I_n)(A_{n-1}\lambda^{n-1} + \cdots + A_1\lambda + A_0) = p_A(\lambda) \cdot I_n \,.$$

Identifying the entries of the matrices from the left hand side and from the right hand side, we get

$$\begin{cases} -A_{n-1} & = \alpha_n I_n \\ AA_{n-1} - A_{n-2} & = \alpha_{n-1} I_n \\ \cdots\cdots & \cdots\cdot \\ AA_1 - A_0 & = \alpha_1 I_n \\ AA_0 & = \alpha_0 I_n \end{cases}$$

Now multiply the first n equalities respectively by $A^n, A^{n-1}, \ldots, A$ and then add all the equalities to obtain

$$p_A(A) = \alpha_n A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_0 I_n = 0_n \,.$$

Hence, A is a root of its characteristic polynomial. □

Corollary 3.9.4 *Let $A \in M_2(K)$. Then:*

$$A^2 - \mathrm{Tr}(A) \cdot A + \det(A) \cdot I_2 = 0_2.$$

Proof. We have seen that the characteristic polynomial of A is

$$p_A(\lambda) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det(A).$$

Then use Theorem 3.9.3. □

Cayley-Hamilton Theorem may be used for computing the inverse or powers of a matrix.

Example 3.9.5 Let

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{R}).$$

Then $\det(A) = 2 \neq 0$, hence A is invertible. Its characteristic polynomial is

$$p_A(\lambda) = \det \begin{pmatrix} 2-\lambda & 0 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 1 & 1-\lambda \end{pmatrix} = -\lambda^3 + 4\lambda^2 - 5\lambda + 2\,.$$

By Theorem 3.9.3, we have

$$A^3 - 4A^2 + 5A - 2I_3 = 0_3\,.$$

It follows that

$$A\left[\frac{1}{2}(A^2 - 4A + 5I_3)\right] = \left[\frac{1}{2}(A^2 - 4A + 5I_3)\right] A = I_3,$$

whence

$$A^{-1} = \frac{1}{2}(A^2 - 4A + 5I_3) = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -2 & 2 \end{pmatrix}\,.$$

For $k \geq 3$, the powers A^k can be computed using the recurrence relation given by Theorem 3.9.3, namely

$$A^k = 4A^{k-1} - 5A^{k-2} + 2A^{k-3}\,.$$

3.10 Diagonalization

The theory of eigenvectors and eigenvalues of an endomorphism is important for finding a basis in which the matrix of the endomorphism has a "nicer" form, if possible a diagonal one.

Definition 3.10.1 A matrix $D \in M_n(K)$ is called *diagonal* if it is of the form

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & d_n \end{pmatrix}.$$

In this case, we denote $D = \operatorname{diag}(d_1, \dots, d_n)$.

Remark 3.10.2 For a diagonal matrix one may easily compute powers. More generally, if we may represent a matrix A as $A = PDP^{-1}$ for some diagonal matrix D and invertible matrix P, then we may easily compute powers

$$A^k = (PDP^{-1})^k = PD^kP^{-1}.$$

This makes a big difference in practice. For instance, let A be a 25×25-matrix for which we need to compute A^{100}. Each multiplication by A requires about $\frac{25^3}{3}$ operations, which is approximately 5200. Hence the total number of operations is $99 \cdot 5200$, which is about 515600 operations. On the other hand, by using the above representation we have $A^k = PD^kP^{-1}$. A multiplication of a 25×25-matrix by a diagonal matrix needs about $25^2 = 625$ operations, while the multiplication of two 25×25-matrices needs about $\frac{25^3}{3}$ operations. Summing up, one uses only about 5800 operations. Of course, in this case one should also find a representation $A = PDP^{-1}$, which adds operations, but it is still much more effective over all.

Definition 3.10.3 A matrix $A \in M_n(K)$ is called *diagonalizable* if A is similar to a diagonal matrix, that is, there exists a diagonal matrix $D \in M_n(K)$ and an invertible matrix $P \in M_n(K)$ such that $A = PDP^{-1}$.
Let $f \in \mathrm{End}_K(V)$. Then f is called *diagonalizable* if there exists a basis B' of V such that $[f]_{B'}$ is diagonal.

Remark 3.10.4 Let V be a vector space over K with a basis B, and let $f \in \mathrm{End}_K(V)$. If f is diagonalizable, then there exists a basis B' of V such that $[f]_{B'}$ is diagonal. Since $T_{BB'}$ is invertible, $[f]_{B'}$ is diagonal and $[f]_B = T_{BB'} \cdot [f]_{B'} \cdot T_{BB'}^{-1}$, it follows that f is diagonalizable if and only if the matrix $[f]_B$ is diagonalizable.

Theorem 3.10.5 *A matrix $A \in M_n(K)$ is diagonalizable if and only if A has n linearly independent eigenvectors.*

Proof. $\boxed{\Longrightarrow}$ Assume that $A \in M_n(K)$ is diagonalizable, say $A = PDP^{-1}$ for some $D, P \in M_n(K)$ with D diagonal and P invertible. Then we have $AP = PD$. If we denote the columns of P by $v^1, \ldots, v^n$ and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, it follows that

$$[Av^1 \ldots Av^n] = AP = DP = [\lambda_1 v^1 \ldots \lambda_n v^n],$$

that is, $Av^i = \lambda_i v^i$ for every $i \in \{1, \ldots, n\}$. Since P is invertible, $v^1, \ldots, v^n$ are linearly independent, and clearly non-zero. Hence $v^1, \ldots, v^n$ are linearly independent eigenvectors of A.

$\boxed{\Longleftarrow}$ Assume that A has n linearly independent eigenvectors, say $v^1, \ldots, v^n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Hence we have $Av^i = \lambda_i v^i$ for every $i \in \{1, \ldots, n\}$. Now choose $P = [v^1 \ldots v^n] \in M_n(K)$ and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \in M_n(K)$. Then we have $AP = PD$. Since the columns $v^1, \ldots, v^n$ of P are linearly independent, P is invertible. This implies that $A = PDP^{-1}$. □

Remark 3.10.6 By Theorem 3.10.5 and its proof, $A = PDP^{-1}$ for some $D, P \in M_n(K)$ with D diagonal and P invertible if and only if the columns of P are n linearly independent vectors of A. In this case the elements on the principal diagonal of D are the eigenvalues of A corresponding to the eigenvectors given by the columns of P.

Example 3.10.7 Let

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 4 \end{pmatrix} \in M_3(\mathbb{R}).$$

In Example 3.8.14, we have computed its eigenvalues $\lambda_1 = \lambda_2 = 2$ and $\lambda_3 = 3$ with corresponding eigenspaces $V(\lambda_1) = \langle (1,0,0),(0,2,1)\rangle$ and $V(\lambda_3) = \langle (0,1,1)\rangle$. The 3 generating vectors of $V(\lambda_1)$ and $V(\lambda_3)$ respectively are eigenvectors which are linearly independent, because the determinant consisting of their components is non-zero. Hence A is diagonalizable by Theorem 3.10.5. Moreover, we have $A = PDP^{-1}$, where

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

We may compute some power $A^{10} = PD^{10}P^{-1}$ as follows:

$$A^{10} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2^{10} & 0 & 0 \\ 0 & 2^{10} & 0 \\ 0 & 0 & 3^{10} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 1024 & 0 & 0 \\ 0 & -57001 & 116050 \\ 0 & -58025 & 117074 \end{pmatrix}.$$

Example 3.10.8 Let

$$A = \begin{pmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{pmatrix} \in M_3(\mathbb{R}).$$

Its characteristic polynomial is $p_A(\lambda) = -(\lambda-1)(\lambda+2)^2$, hence its eignevalues are $\lambda_1 = 1$ and $\lambda_2 = \lambda_3 = -2$. One computes $V(\lambda_1) = \langle (1,-1,1)\rangle$ and $V(\lambda_2) = \langle (1,-1,0)\rangle$. Note that A does not have 3 linearly independent eigenvectors, hence A is not diagonalizable by Theorem 3.10.5.

Theorem 3.10.9 *Let $f \in \mathrm{End}_K(V)$ and let $v_1, \ldots, v_n$ be eigenvectors with distinct corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $v_1, \ldots, v_n$ are linearly independent.*

Proof. We prove the result by induction on n. It is clearly true for $n = 1$, because a single eigenvector, which is not zero, is linearly independent. Now suppose that the property holds for n and prove it for $n+1$. Let $k_1, \ldots, k_{n+1} \in K$ be such that

$$k_1 v_1 + \cdots + k_{n+1} v_{n+1} = 0.$$

Apply f to get

$$k_1 f(v_1) + \cdots + k_{n+1} f(v_{n+1}) = 0,$$

and further

$$\lambda_1 k_1 v_1 + \cdots + \lambda_{n+1} k_{n+1} v_{n+1} = 0\,.$$

Now multiply the first equality by $-\lambda_{n+1}$ and add it to the last equality. Then

$$(\lambda_1 - \lambda_{n+1}) k_1 v_1 + \cdots + (\lambda_n - \lambda_{n+1}) k_n v_n = 0$$

By the induction hypothesis, the vectors $v_1, \ldots, v_n$ are linearly independent, whence it follows that $k_1 = \cdots = k_n = 0$. But then $k_{n+1} v_{n+1} = 0$, hence $k_{n+1} = 0$. Consequently, the vectors $v_1, \ldots, v_n$ are linearly independent. □

Theorem 3.10.10 *Let $A \in M_n(K)$ having n distinct eigenvalues. Then A is diagonalizable.*
Moreover, if $A = [f]_B$ for some $f \in \mathrm{End}_K(V)$ and basis B of V, then A is similar to a matrix

$$A' = [f]_{B'} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix},$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A, and B' is the basis of the corresponding eigenvectors.

Proof. This follows by Theorems 3.10.5 and 3.10.9, and Remark 3.10.6. □

We will give a more general result than Theorem 3.10.10. But first let us introduce the following definition.

Definition 3.10.11 Let $f \in \mathrm{End}_K(V)$ and let λ be a root of the characteristic polynomial of f. Then:
(i) The multiplicity of the root λ is called the *algebraic multiplicity* of λ, and is denoted by $a(\lambda)$.
(ii) $\dim V(\lambda)$ is called the *geometric multiplicity* of λ, and is denoted by $g(\lambda)$.

Theorem 3.10.12 *Let $f \in \mathrm{End}_K(V)$ and let λ_0 be a root of the characteristic polynomial of f. Then $g(\lambda_0) \leq a(\lambda_0)$.*

Proof. Let $B = (v_1, \dots, v_m)$ be a basis of the eigenspace $V(\lambda_0)$. Hence $g(\lambda_0) = m$. Then we have $f(v_j) = \lambda_0 v_j$ for every $j \in \{1, \dots, m\}$. Now complete B to a basis of V, say $B' = (v_1, \dots, v_m, v'_{k+1}, \dots, v'_n)$. Let $A_1 = \lambda_0 I_m \in M_m(K)$. Then we have

$$[f]_{B'} = \begin{pmatrix} A_1 & A_2 \\ 0_{n-m,m} & A_3 \end{pmatrix}$$

for some suitable submatrices $A_2 \in M_{m,n-m}(K)$ and $A_3 \in M_{n-m}(K)$. It follows that

$$\begin{aligned} p_f(\lambda) &= \det[([f]_{B'}) - \lambda I_n] \\ &= \det \begin{pmatrix} A_1 - \lambda I_m & A_2 \\ 0_{n-m,m} & A_3 - \lambda_{n-m} I_{n-m} \end{pmatrix} \\ &= \det(A_1 - \lambda I_m) \det(A_3 - \lambda_{n-m} I_{n-m}) \\ &= (\lambda_0 - \lambda)^m \det(A_3 - \lambda_{n-m} I_{n-m}). \end{aligned}$$

This shows that $g(\lambda_0) = m \leq a(\lambda_0)$. □

Theorem 3.10.13 *Let $f \in \mathrm{End}_K(V)$. Then f is diagonalizable if and only if*
(i) the characteristic polynomial of f has all the roots in K.
(ii) for every eigenvalue λ of f, $g(\lambda) = a(\lambda)$.

Proof. $\boxed{\Longrightarrow}$ Assume that f is diagonalizable. Then there exists a basis $B = (v_1, \ldots, v_n)$ of V such that

$$[f]_B = \operatorname{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{n_1}, \ldots, \underbrace{\lambda_m, \ldots, \lambda_m}_{n_m})$$

with distinct entries, $m \leq n$, $n_1, \ldots, n_m \in \{1, \ldots, n\}$ and $n_1 + \cdots + n_m = n$. Then we have

$$p_f(\lambda) = (\lambda_1 - \lambda)^{n_1} \cdots (\lambda_m - \lambda)^{n_m}.$$

Hence $p_f(\lambda)$ has all its n roots in K.

By Remark 3.10.6, $\lambda_1, \ldots, \lambda_m$ are eigenvalues corresponding to the eigenvectors $v_1, \ldots, v_n$ of f. For every $i \in \{1, \ldots, m\}$, $V(\lambda_i)$ contains n_i linearly independent vectors, namely

$$v_{n_1+\cdots+n_{i-1}+1}, \ldots, v_{n_1+\cdots+n_{i-1}+n_i}.$$

Then by Steinitz Theorem, we have $a(\lambda_i) = n_i \leq \dim V(\lambda_i) = g(\lambda_i)$. But we also have $g(\lambda_i) \leq a(\lambda_i)$ by Theorem 3.10.12. Thus, $g(\lambda_i) = a(\lambda_i)$ for every $i \in \{1, \ldots, m\}$.

$\boxed{\Longleftarrow}$ Assume that conditions (i) and (ii) hold. Let $\lambda_1, \ldots, \lambda_m$ be the distinct roots of $p_f(\lambda)$ with orders of multiplicities $n_1, \ldots, n_m \in \{1, \ldots, n\}$ respectively, where $m \leq n$ and $n_1 + \cdots + n_m = n$. For every $i \in \{1, \ldots, m\}$, we have $g(\lambda_i) = a(\lambda_i) = n_i$, hence $V(\lambda_i)$ has a basis $(v_{i1}, \ldots, v_{in_i})$. We prove that

$$B = (v_{11}, \ldots, v_{1n_1}, \ldots, v_{m1}, \ldots, v_{mn_m})$$

is linearly independent. Note that B has $n_1 + \cdots + n_m = n$ vectors. Let

$$\sum_{j_1=1}^{n_1} a_{1j_1} v_{1j_1} + \cdots + \sum_{j_m=1}^{n_m} a_{mj_m} v_{mj_m} = 0$$

for some $a_{ij_i} \in K$, $i \in \{1, \ldots, m\}$. Note that the sums are from $V(\lambda_1), \ldots, V(\lambda_m)$ respectively. By Theorem 3.10.10, eigenvectors corresponding to distinct eigenvalues are linearly independent. Hence we must have

$$\sum_{j_1=1}^{n_1} a_{1j_1} v_{1j_1} = \cdots = \sum_{j_m=1}^{n_m} a_{mj_m} v_{mj_m} = 0.$$

Since each $(v_{i1}, \ldots, v_{in_i})$ is a basis of $V(\lambda_i)$, it follows that $a_{i1} = \cdots = a_{in_i} = 0$ for every $i \in \{1, \ldots, m\}$. This shows that the n vectors of B are linearly independent eigenvectors. Hence f is diagonalizable by Theorem 3.10.5. □

Remark 3.10.14 One may use Theorem 3.10.13 instead of Theorem 3.10.5 in order to see that the matrix from Example 3.10.7 is diagonalizable and the matrix from Example 3.10.8 is not diagonalizable.

Extra: Singular value decomposition

Definition 3.10.15 A matrix $Q \in M_n(\mathbb{R})$ is called *orthogonal* if $Q^T \cdot Q = Q \cdot Q^T = I_n$. A matrix $A \in M_{m,n}(\mathbb{R})$ has a *singular value decomposition (SVD)* if it may be written

as $A = U\Sigma V^T$ for some diagonal matrix Σ and orthogonal matrices $U \in M_m(\mathbb{R})$ and $V \in M_n(\mathbb{R})$.

Theorem 3.10.16 *Every matrix $A \in M_{m,n}(\mathbb{R})$ has a singular value decomposition (not necessarily unique), where the values on the diagonal of Σ (called* singular values*) are the square roots of the non-zero eigenvalues of both $A \cdot A^T$ and $A \cdot A^T$ (usually in decreasing order), the columns of U are eigenvectors of $A \cdot A^T$, and the columns of V are eigenvectors of $A^T \cdot A$.*

Example 3.10.17 Let

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \in M_3(\mathbb{R}).$$

The matrix

$$A \cdot A^T = A^T \cdot A = A^2 = \begin{pmatrix} 2 & 0 & 2 \\ 0 & 1 & 0 \\ 2 & 0 & 2 \end{pmatrix}$$

has eigenvalues 4, 1 and 0 with corresponding eigenspaces $V(4) = \langle (1,0,1) \rangle$, $V(1) = \langle (0,1,0) \rangle$ and $V(0) = \langle (-1,0,1) \rangle$ respectively. Eventually, a singular value decomposition of A is

$$A = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix} \cdot \begin{pmatrix} \boxed{2} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{pmatrix}^T,$$

where the singular values are the values 2 and 1 from the middle matrix.

We illustrate SVD with an application to image processing. Suppose that one needs to transmit an image having (say) $1000 \times 1000 = 1000000$ pixels, the color of each pixel being given by a certain number. Instead of sending a whole 1000×1000-matrix A associated to the image, one may use the SVD, say $A = U\Sigma V^T$, in order to keep the most important information from the image. Thus, one may keep only the largest (say) 100 singular values from Σ, and send only the corresponding 50 columns of U and V. All other 900 columns involved in the multiplication $U\Sigma V^T$ use small singular values, and may be ignored. In this case, one only sends 100 times 2000 numbers from the SVD given by $U\Sigma V^T$, that is, 200000 numbers. The image quality becomes better as more singular values are taken into consideration.

Chapter 3 quiz

Decide whether the following statements are **true** or **false**.

1. Let V be a vector space over K and $n \in \mathbb{N}^*$. Then every elementary operation $V^n \to V^n$ is an isomorphism of vector spaces over K.

2. Let V be a vector space over K and let X, X', X'' be lists of vectors in the vector space V^n over K, where $n \in \mathbb{N}^*$. If X is equivalent to X' and X' is equivalent to X'', then X is equivalent to X''.

3. Elementary operations on rows of a square matrix $A \in M_n(K)$ do not change the determinant of the matrix.

4. The echelon form of any non-zero matrix $A \in M_n(K)$ is unique.

5. An elementary matrix $E \in M_n(K)$ is a matrix that differs from the identity matrix by one elementary row operation.

6. Every non-zero square matrix $A \in M_n(K)$ is equivalent to an identity matrix.

7. Let $A \in M_n(K)$ be an invertible matrix viewed as a list of row-vectors $(a_1, \ldots, a_n)$. Then $dim\langle a_1, \ldots, a_n\rangle = n$.

8. Let $f : V \to V$ be a K-linear map such that its matrix $[f]_B$ in a basis B of V is an elementary matrix. Then f is an automorphism.

9. Let V be a vector space over K, and let $T_{BB'} = (t_{ij}) \in M_n(K)$ be the change matrix from the basis $B = (v_1, \ldots, v_n)$ to the basis $B' = (v'_1, \ldots, v'_n)$ of V. Then for every $j \in \{1, \ldots, n\}$, we have

$$v'_j = \sum_{i=1}^{n} t_{ij} v_i.$$

10. Let $f : V \to V'$ be a K-linear map with $[f]_{BB'} = (a_{ij}) \in M_{m,n}(K)$, where $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_m)$ are bases of V and V' respectively. Then for every $i \in \{1, \ldots, m\}$, we have

$$f(v_j) = \sum_{j=1}^{n} a_{ij} v'_i.$$

11. Let $f : V \to V'$ be a K-linear map, $B = (v_1, \ldots, v_n)$ a basis of V, $B' = (v'_1, \ldots, v'_m)$ a basis of V' and $v \in V$. Then

$$[f(v)]_{B'} = [v]_B \cdot [f]_{BB'}.$$

12. Let $f \in \mathrm{End}_K(V)$ and let $T_{BB'}, T_{B'B} \in M_n(K)$ be the change matrices from the basis B to the basis B' of V and from B' to B respectively. Then

$$T_{B'B} \cdot [f]_{B'} = [f]_B \cdot T_{BB'}.$$

13. Let $f, g : V \to V'$ be K-linear maps, let B and B' be bases of V and V' respectively, and let $\alpha, \beta \in K$. Then

$$[\alpha f + \beta g]_{BB'} = \alpha \cdot [f]_{BB'} + \beta \cdot [g]_{BB'}.$$

14. Every homogeneous linear system of equations over K has a non-zero solution.

15. A linear system of equations over K with fewer equations than unknowns must have infinitely many solutions.

16. Every homogeneous linear system of equations with matrix $A \in M_{m,n}(K)$ and $\mathrm{rank}(A) = n$ has only the zero solution.

17. If a linear system of equations $A \cdot X = B$ over K has a unique solution, then A must be a square matrix.

18. A linear system of equations over K having matrix A and augmented matrix $\bar{A}$ is compatible if and only if $\text{rank}(\bar{A}) \leq \text{rank}(A)$.

19. Every linear system of equations over K having more equations than unknowns is not compatible.

20. An eigenvector may have more than one associated eigenvalue.

21. A matrix $A \in M_n(K)$ has the same eigenvalues as its transpose.

22. A matrix $A \in M_n(K)$ has the same eigenvectors as its transpose.

23. Two different matrices $A, B \in M_n(K)$ must have different characteristic polynomials.

24. For every field K, every matrix $A \in M_n(K)$ has n eigenvalues in K, considering possibly multiple ones.

25. Let V be a vector space over K with $dim\, V = n$ and $f \in \text{End}_K(V)$. If the characteristic polynomial of f has n distinct roots, then there is a basis B of V such that $[f]_B$ is diagonal.

Chapter 3 projects

Use a programming language of your choice and implement the following projects.

Project 3.1

- *Input*: natural numbers $m, n \geq 2$
- *Output*:
  1. the number of matrices $A \in M_{m,n}(\mathbb{Z}_2)$ in *reduced* (that is, each column containing a leading 1 has zeros everywhere else) echelon form
  2. the matrices $A \in M_{m,n}(\mathbb{Z}_2)$ in reduced echelon form (for $2 \leq m, n \leq 5$)

Example:

- *Input*: $m = 2$, $n = 3$
- *Output*:
  1. the number of matrices $A \in M_{2,3}(\mathbb{Z}_2)$ in reduced echelon form is 15
  2. the matrices $A \in M_{2,3}(\mathbb{Z}_2)$ in reduced echelon form are (the leading 1's are framed):

$$\begin{pmatrix} \boxed{1} & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} \boxed{1} & 0 & 1 \\ 0 & \boxed{1} & 1 \end{pmatrix} \quad \begin{pmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 1 \end{pmatrix} \quad \begin{pmatrix} \boxed{1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & \boxed{1} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\begin{pmatrix} \boxed{1} & 1 & 0 \\ 0 & 0 & \boxed{1} \end{pmatrix} \quad \begin{pmatrix} \boxed{1} & 0 & 1 \\ 0 & \boxed{1} & 0 \end{pmatrix} \quad \begin{pmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & \boxed{1} & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & \boxed{1} \\ 0 & 0 & 0 \end{pmatrix}$$
$$\begin{pmatrix} \boxed{1} & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} \boxed{1} & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} \end{pmatrix} \quad \begin{pmatrix} 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Project 3.2

- *Input*: natural numbers $m, n \geq 2$ and a matrix $A \in M_{m,n}(\mathbb{R})$
- *Output*:
  1. the dimension of $\operatorname{Ker} f$ and $\operatorname{Im} f$, where $f : \mathbb{R}^n \to \mathbb{R}^m$ is an $\mathbb{R}$-linear map having the matrix A in the pair of canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$
  2. a basis for the above $\operatorname{Ker} f$ and $\operatorname{Im} f$

Example:

- *Input*: $m = 2$, $n = 3$, $A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \in M_{2,3}(\mathbb{R})$
- *Output*:
  1. $\dim \operatorname{Ker} f = 1$ and $\dim \operatorname{Im} f = 2$
  2. a basis for $\operatorname{Ker} f$ is $((0, 0, 1))$ and a basis for $\operatorname{Im} f$ is $((0, -1), (1, 0))$

Chapter 4

Introduction to coding theory

As an application of Linear Algebra we have studied so far, in this final chapter we present a brief introduction to Coding Theory. We put emphasis on linear codes, which is a large class of codes used in real life applications. We mainly follow [8] for our presentation.

4.1 Coding theory

Coding Theory is one of the most important research fields relating Algebra and Computer Science. Its primary objective is to ensure accurate transmissions through open communication channels. There is always some "noise" on communication channels, that is, the data may be altered during the transmission process. That is why it is important to have good systems to discover errors and to correct them. This is achieved by Coding Theory through *error-detecting codes* and *error-correcting codes*. They are widely used in real life, in various places, ranging from satellite and space transmissions to credit cards or storage devices (CD's, DVD's, Blu-ray discs etc.). In case of the latter ones the errors may appear during the read/write processes. We should note that different codes may be suitable for different applications.

The starting points of Coding Theory may be considered the works of Shannon on Information Theory (1948) and Hamming on Error-Correcting Codes (1950). When talking about codes, one may distinguish two large main classes, namely *source coding*, which refer to data compression, and *channel coding*, which refer to error-detecting and error-correcting codes. In this chapter we ony refer to the latter class.

Suppose that we have a communication channel whose probability of a correct transmission is p. The probability of t errors in a message of length m is

$$C_m^t p^{m-t}(1-p)^t.$$

For instance, for $p = 0.99$ and $m = 50$, we have the following table:

t	Probability of t errors
0	0.605
1	0.3056
2	0.0756
3	0.0122
4	0.00145

These probabilities decrease if m is small enough, more precisely when $m < \frac{p}{1-p}$. Hence we should not expect too many errors during a transmission. But still they happen, and should be detected and corrected.

Let us begin with an example of a widely used simple code.

Example 4.1.1 *EAN-13 International Article Number* is a sequence of 13 digits $a_1, a_2, \ldots, a_{13}$ that identifies a product. Digit a_{13} is a check digit that is computed as

$$a_{13} = 10 - (a_1 + 3a_2 + a_3 + 3a_4 + \cdots + a_{11} + 3a_{12}) \bmod 10,$$

where $a \bmod n$ is the remainder of the division of an integer a to a non-zero natural number n.

Digits are written in binary; black bars for 1, white bars for 0.

In particular, one has ISBN (International Standard Book Number), UPC (Universal Product Code) etc.

The general scheme of error-correcting (detecting) codes is the following one.

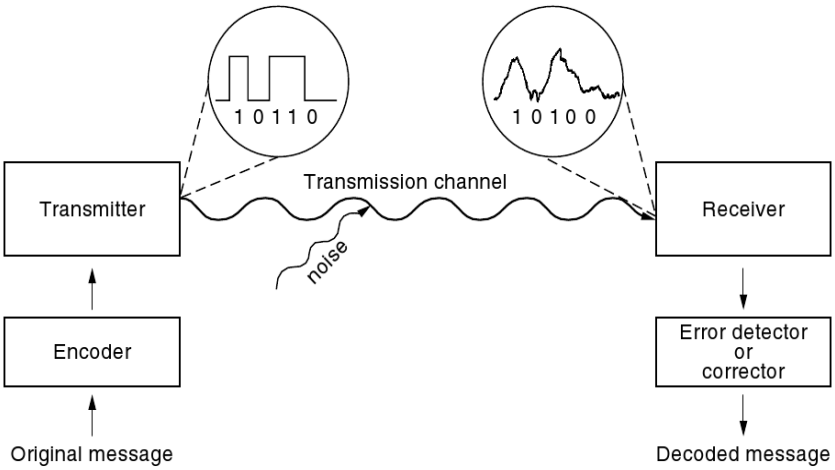


Figure 4.1: General scheme of error-correcting (detecting) codes [8].

We shall use the following general setting:

- We discuss *binary codes*, that is, codes over the field $\mathbb{Z}_2$. The elements $\widehat{0}$ and $\widehat{1}$ of $\mathbb{Z}_2$ are simply denoted by 0 and 1 respectively. In general, one may use codes over finite fields.

- We consider *symmetric channels*: the probability of 1 being changed into 0 is the same as that of 0 being changed into 1.

- We assume that the number of errors is less than the number of correctly transmitted bits.

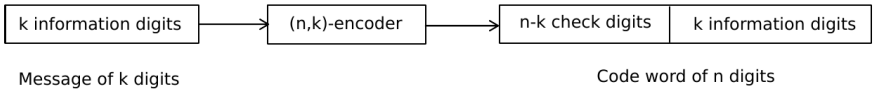


Figure 4.2: General scheme of an (n, k)-code.

Definition 4.1.2 Let $k, n \in \mathbb{N}^*$ with $k < n$. An (n, k)-*code* is described by the above general scheme.
The ratio $\frac{k}{n}$ is called the *information rate* of the (n, k)-code.

Remark 4.1.3 Note that there are 2^k possible messages, and thus 2^k code words. Also, there are 2^n possible received words.

On one hand, one is looking to send as much information as possible in one transmission. On the other hand, one needs to use enough check digits in order to allow error-detection and error-correction. Now we may state the Coding Theory Problem, which is the main challenge in Coding Theory.

Coding Theory Problem: Find the right balance between k and $n - k$.

Next we present two examples of simple codes.

Example 4.1.4 (a) *The* $(3, 2)$*-parity check code.*

- The check digit is the sum modulo 2 of the message digits.
- The encoding process goes as follows:

Message	Code word
00	000
01	101
10	110
11	011

  How many errors can this code detect/correct? It can detect one error, but it cannot correct any error.

- The decoding process goes as follows:

Received words	101	111	100	000	110
Parity check	passes	fails	fails	passes	passes
Decoded words	01	-	-	00	10

- The information rate is $\frac{2}{3}$.

(b) *The* $(3, 1)$*-repeating code.*

- The two check digits repeat the message digit.
- The encoding process goes as follows:

Message	Code word
0	000
1	111

  How many errors can this code detect/correct? It can detect up to two errors, and it can correct one error.

- The decoding process goes as follows:

Received words	111	010	011	000
Decoded words	1	0	1	0

- The information rate is $\frac{1}{3}$.

4.2 Hamming distance

Definition 4.2.1 The *Hamming distance* between two words of the same length is the number of positions in which they difer.
The *Hamming distance* of a code is the minimum Hamming distance between code words.
The *Hamming weight* of a word is the number of its non-zero bits.

We denote the Hamming distance between two words u, v of the same length by $d(u, v)$, and the Hamming weight of a word u by $w(u)$.

Remark 4.2.2 For every $u, v \in \mathbb{Z}_2^n$, we have $w(u - v) = d(u, v)$.

Example 4.2.3 (a) We have $d(101, 100) = 1$, $d(110, 001) = 3$ and $d(101, 011) = 2$.

(b) We have $w(101) = 2$ and $w(100) = 1$.

Theorem 4.2.4 *The Hamming distance is a* metric *on the set $\mathbb{Z}_2^n$ of words of length n, that is, the following properties hold for every $u, v, t \in \mathbb{Z}_2^n$:*
(i) $d(u, v) = d(v, u)$.
(ii) $d(u, v) \geq 0$ with equality if and only if $u = v$.
(iii) $d(u, v) + d(v, t) \geq d(u, t)$.

Proof. (i), (ii) These are clear.

(iii) For a word $w \in \mathbb{Z}_2^n$ denote by w_i its digits for $i \in \{1, \ldots, n\}$. Notice that if u_i and t_i are different, then either u_i and v_i are different, or v_i and t_i are different. □

Corollary 4.2.5 *The Hamming distance $d_{\min}$ of a code is the minimum weight $w_{\min}$ of a non-zero code word.*

Proof. If c is a code word of minimum weight, then $w(c) = d(0, c)$. Since 0 is a code word, we must have $d_{\min} \leq w_{\min}$.

If c_1, c_2 are code words at the minimum Hamming distance, then $d(c_1, c_2) = w(c_1 - c_2)$. Since $c_1 - c_2$ is a code word, we must have $w_{\min} \leq d_{\min}$. □

Remark 4.2.6 (1) The problem of finding the Hamming distance of a code is difficult in general. We will see in a forthcoming result how it can be computed by using the parity check matrix of an (n, k)-code.

(2) In practice, (n, k)-codes may be described more precisely as (n, k, d)-codes, where d is the Hamming distance of the code.

There is a geometric interpretation of the Hamming distance, as follows:

- In an (n, k)-code, the 2^n received words can be thought of as placed at the vertices of an n-dimensional cube with unit sides.
- The Hamming distance between two words is the shortest distance between their corresponding vertices along the edges of the n-cube.
- The 2^k code words form a subset of the 2^n vertices, and the code has better error-correcting and error-detecting capabilities the farther apart these code words are.
- Cube representations of the $(3, 2)$-parity check and $(3, 1)$-repeating codes:

Figure 4.3: Cube representations of the $(3, 2)$-parity check and $(3, 1)$-repeating codes [8].

Next we present two results on error detection/correction capabilities of codes.

Theorem 4.2.7 *A code detects all sets of t or fewer errors if and only if the Hamming distance of the code is at least $t + 1$.*

Proof. Denote by r the number of ocurred errors when a code word u is transmitted. Then the received word v is at Hamming distance $d(v, u) = r$ from u. These errors are detected if and only if v is not another code word. Hence all sets of t or fewer errors in the code word u will be detected if and only if the Hamming distance of u from all the other code words is at least $t + 1$. □

Theorem 4.2.8 *A code is capable of correcting all sets of t or fewer errors if and only if the Hamming distance of the code is at least $2t + 1$.*

Proof. $\boxed{\Longrightarrow}$ Assume that the code is capable of correcting all sets of t or fewer errors. Let u_1 and u_2 be code words at Hamming distance $d(u_1, u_2) \leq 2t$. Then there exists a received word v such that $d(v, u_1) \leq t$ and $d(v, u_2) \leq t$. It follows that the received word v could have been obtained from u_1 or u_2 with t or fewer errors, and hence would not be correctly decoded in both these situations.

$\boxed{\Longleftarrow}$ Assume that the minimum Hamming distance between code words is at least $2t + 1$. Any code whose code words are at least $2t + 1$ apart is capable of correcting up to t errors. Note that this can be achieved in decoding by choosing the code word that is closest to each received word. □

We summarize the above results in the following table.

Code	**Hamming distance of the code**	**No. of detectable errors**	**No. of correctable errors**	**Information rate**
(n, k)-code	d	$d - 1$	$\leq \frac{d-1}{2}$	$\frac{k}{n}$
$(3, 2)$-parity check code	2	1	0	$\frac{2}{3}$
$(3, 1)$-repeating code	3	2	1	$\frac{1}{3}$

The Hamming distance may be used to describe the following **naive method for error-correcting and decoding**.

(1) *Given a received word, compute all Hamming distances to the code words.*

(2) *The code word closest to the received word will be assumed to be the most likely transmitted word.*

In practice one uses large codes, which make this method inefficient. In what follows we show how Linear Algebra may be used to develop a practical method for error-correcting and decoding.

4.3 Code representations

Throughout this section and the next ones, we have $k, n \in \mathbb{N}^*$ with $k < n$. We discuss two possible representations for the words of an (n, k)-code, namely the polynomial and the matrix representations.

Definition 4.3.1 The *polynomial representation* of a binary n-digit word $a_0 a_1 \dots a_{n-1}$ is the polynomial

$$a_0 + a_1 X + \cdots + a_{n-1} X^{n-1} \in \mathbb{Z}_2[X].$$

Definition 4.3.2 Let $p \in \mathbb{Z}_2[X]$ be of degree $n-k$.
The *polynomial code generated by* p is an (n,k)-code whose code words are those polynomials of degree less than n which are divisible by p.
Then the polynomial p is called the *generator* of the code.

We need the following result on polynomials over a field. The division of polynomials over arbitrary fields works in the same way as the usual division of polynomials over numerical fields.

Theorem 4.3.3 (Division Algorithm for polynomials) *Let K be a field, and let $f, g \in K[X]$ with $g \neq 0$. Then there are unique $q, r \in K[X]$ such that*

$$f = g \cdot q + r \quad \text{degree}(r) < \text{degree}(g).$$

Next we describe the **encoding process** for a polynomial (n,k)-code generated by a polynomial $p \in \mathbb{Z}_2[X]$ of degree $n-k$.

(1) *A message of length k is represented by a polynomial $m \in \mathbb{Z}_2[X]$ of degree less than k.*

(2) *Since the message is stored in the right hand side of a word, the message digits are carried by the higher-order coefficients of a polynomial. So we consider $m \cdot X^{n-k}$.*

(3) *To encode the message polynomial m we first use the Division Algorithm for polynomials in order to find unique $q, r \in \mathbb{Z}_2[X]$ such that*

$$m \cdot X^{n-k} = q \cdot p + r\,, \quad \text{degree}(r) < \text{degree}(p) = n-k.$$

(4) *The code polynomial is*

$$v = r + m \cdot X^{n-k}.$$

The check digits of the message are carried by r.

Theorem 4.3.4 *With the above notation, the code polynomial v is divisible by p.*

Proof. We have

$$v = r + m \cdot X^{n-k} = r + q \cdot p + r = q \cdot p,$$

because $r \in \mathbb{Z}_2[X]$, and thus $r + r = 0$. □

Example 4.3.5 *Let*

$$p = 1 + X^2 + X^3 + X^4 \in \mathbb{Z}_2[X]$$

be the generator polynomial of a $(7,3)$-code. Let us encode the message 101.

Note that $n = 7$ and $k = 3$. The encoding process goes as follows:

$$\begin{aligned}\text{message } 101 &\rightsquigarrow m = 1 \cdot 1 + 0 \cdot X + 1 \cdot X^2 = 1 + X^2 \\ &\rightsquigarrow mX^{n-k} = (1 + X^2) \cdot X^4 = X^4 + X^6 \\ &\rightsquigarrow r = mX^{n-k} \bmod p = (X^4 + X^6) \bmod p = 1 + X \\ &\rightsquigarrow v = r + mX^{n-k} = 1 + X + X^4 + X^6 \\ &\rightsquigarrow \text{code word } \boxed{1100}\boxed{101}\end{aligned}$$

Example 4.3.6 *If the generator polynomial of a* $(6, 3)$*-code is*

$$p = 1 + X + X^3 \in \mathbb{Z}_2[X],$$

test whether the following received words contain detectable errors: 100011, 100110.

One checks whether the received words are code words, that is, their associated polynomials are divisible by p.

Example 4.3.7 *Write down all the code words for the* $(6, 3)$*-code generated by the polynomial*

$$p = 1 + X + X^3 \in \mathbb{Z}_2[X].$$

Note that $n = 6$, $k = 3$, and we have $2^k = 8$ code words.
By applying the above algorithm for encoding, we obtain the following table:

message	code word
000	000000
001	111001
010	011010
011	100011
100	110100
101	001101
110	101110
111	010111

For instance, the code word associated to the message 110 is 101110, because we have:

$$\begin{aligned}110 &\rightsquigarrow m = 1 + X \\ &\rightsquigarrow mX^{n-k} = X^3 + X^4 \\ &\rightsquigarrow r = mX^{n-k} \bmod p = (X^3 + X^4) \bmod p = 1 + X^2 \\ &\rightsquigarrow v = r + mX^{n-k} = 1 + X^2 + X^3 + X^4 \\ &\rightsquigarrow \boxed{101}\boxed{110}\end{aligned}$$

Definition 4.3.8 The *matrix representation* of a binary n-digit word $a_0a_1 \dots a_{n-1}$ is a one-column matrix

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} \in M_{n,1}(\mathbb{Z}_2).$$

For an (n,k)-code, we see the 2^k possible messages as the elements of the vector space $\mathbb{Z}_2^k$ over $\mathbb{Z}_2$, and the 2^n possible received words as the elements of the vector space $\mathbb{Z}_2^n$ over $\mathbb{Z}_2$.

Definition 4.3.9 An *encoder* of an (n,k)-code is an injective function

$$\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$$

or equivalently,

$$\gamma : M_{k,1}(\mathbb{Z}_2) \to M_{n,1}(\mathbb{Z}_2).$$

An (n,k)-code is called *linear* if its encoder is a linear map.

Let us note that some of the most important codes used in practice are linear. For instance: the *Reed-Solomon code*, used for CD's, DVD's, Blu-ray discs etc.

A large class of linear codes is given in the following theorem.

Theorem 4.3.10 *Any (n,k)-code generated by a polynomial of degree $n-k$ is linear.*

Proof. Let $p \in \mathbb{Z}_2[X]$ be the generator polynomial of an $(n.k)$-code. We have seen that we encode the message $m \in \mathbb{Z}_2[X]$ as

$$v = r + m \cdot X^{n-k},$$

where $r \in \mathbb{Z}_2[X]$ is the remainder of the division of m by p, that is, $m \bmod p$.

Hence the encoder $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ associates to the k-tuple of the coefficients of m the n-tuple of the coefficients of $v = r + m \cdot X^{n-k}$.

Next we show that γ is a linear map, that is,

$$\gamma(k_1 m_1 + k_2 m_2) = k_1\gamma(m_1) + k_2\gamma(m_2), \quad \forall k_1, k_2 \in \mathbb{Z}_2, \forall m_1, m_2 \in \mathbb{Z}_2^k.$$

Since 0 and 1 are the only scalars, it is enough to show that

$$\gamma(m_1 + m_2) = \gamma(m_1) + \gamma(m_2), \quad \forall m_1, m_2 \in \mathbb{Z}_2^k.$$

Let $m_1, m_2 \in \mathbb{Z}_2^k$. We have

$$\gamma(m_1) = v_1 = r_1 + m_1 \cdot X^{n-k},$$
$$\gamma(m_2) = v_2 = r_2 + m_2 \cdot X^{n-k}$$

for some unique $r_1, r_2 \in \mathbb{Z}_2[X]$ with $\text{degree}(r_1) < n-k$ and $\text{degree}(r_1) < n-k$. Then

$$v_1 + v_2 = r_1 + r_2 + (m_1 + m_2) \cdot X^{n-k},$$

where $\text{degree}(r_1 + r_2) < n-k$. Hence $r_1 + r_2$ is the remainder of the division of $(m_1 + m_2) \cdot X^{n-k}$ by p. Thus we must have $\gamma(m_1 + m_2) = v_1 + v_2$. This implies that

$$\gamma(m_1 + m_2) = \gamma(m_1) + \gamma(m_2),$$

as required. □

4.4 Generator matrix and parity check matrix

From now on we will discuss only linear codes. In this section we see that every linear (n, k)-code is completely determined by some associated matrix.

Definition 4.4.1 Consider a linear (n, k)-code with encoder $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$. Let E, E' be the canonical bases of the $\mathbb{Z}_2$-vector spaces $\mathbb{Z}_2^k$ and $\mathbb{Z}_2^n$ respectively. Then the matrix

$$G = [\gamma]_{EE'} \in M_{n,k}(K)$$

is called the *generator matrix* of the code.

Now let us see how one can use the generator matrix to encode a message. Recall that a message $m \in \mathbb{Z}_2^k$ encodes as $\gamma(m)$. But for $m \in \mathbb{Z}_2^k$, we have

$$[\gamma(m)]_{E'} = [\gamma]_{EE'} \cdot [m]_E.$$

Hence a message $m \in M_{k,1}(\mathbb{Z}_2)$ encodes as $G \cdot m$.

We immediately deduce the following result.

Theorem 4.4.2 *Consider a linear (n, k)-code with encoder $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ and generator matrix $G \in M_{n,k}(K)$. Then:*
(i) The code words of the linear (n, k)-code are the vectors in the subspace $\operatorname{Im}\gamma$ of $\mathbb{Z}_2^n$. Hence a binary linear (n, k)-code means a k-dimensional subspace of the vector space $\mathbb{Z}_2^n$.
(ii) The columns of G form a basis of this subspace, and thus a vector is a code vector if and only if it is a (unique) linear combination of the columns of G.

Remark 4.4.3 A code word contains the message digits on the last k positions. Hence the generator matrix G of a linear (n, k)-code is always of the form

$$G = \begin{pmatrix} P \\ I_k \end{pmatrix} \in M_{n,k}(\mathbb{Z}_2),$$

where $P \in M_{n-k,k}(\mathbb{Z}_2)$ and $I_k \in M_k(\mathbb{Z}_2)$ is the identity matrix.

Next we define another matrix related to the generator matrix of the code, which will be useful for obtaining a more efficient practical characterization of received vectors that are code vectors.

Definition 4.4.4 With the above notation, the matrix

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} \in M_{n-k,n}(\mathbb{Z}_2)$$

is called the *parity check matrix* of the code.

Theorem 4.4.5 *Consider a linear (n,k)-code with generator and parity check matrices*

$$G = \begin{pmatrix} P \\ I_k \end{pmatrix} \in M_{n,k}(\mathbb{Z}_2) \text{ and } H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} \in M_{n-k,n}(\mathbb{Z}_2)$$

respectively, where $P \in M_{n-k,k}(\mathbb{Z}_2)$. Let

$$\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n \text{ and } \eta : \mathbb{Z}_2^n \to \mathbb{Z}_2^{n-k}$$

be the $\mathbb{Z}_2$-linear maps corresponding to G and H respectively. Then:
(i) $V = \operatorname{Im}\gamma = \operatorname{Ker}\eta$.
(ii) A received vector $u \in M_{n,1}(\mathbb{Z}_2)$ (or $u \in \mathbb{Z}_2^n$) is a code vector if and only if $H \cdot u = 0$.

Proof. (i) Let E, E', E'' be the canonical bases of the $\mathbb{Z}_2$-vector spaces $\mathbb{Z}_2^k$, $\mathbb{Z}_2^n$ and $\mathbb{Z}_2^{n-k}$ respectively. Then we have

$$[\eta \circ \gamma]_{EE'} = [\eta]_{E'E''}[\gamma]_{EE'} = HG = (I_{n-k}|P)\begin{pmatrix} P \\ I_k \end{pmatrix} = I_{n-k}P + PI_k = P + P = 0,$$

hence $\eta \circ \gamma = 0$. For every $v \in \operatorname{Im}\gamma$, we may write $v = \gamma(v_0)$ for some $v_0 \in \mathbb{Z}_2^k$, whence we deduce that

$$\eta(v) = \eta(\gamma(v_0)) = (\eta \circ \gamma)(v_0) = 0.$$

Hence $\operatorname{Im}\gamma \subseteq \operatorname{Ker}\eta$.

On the other hand, the first $n-k$ columns of H consist of the vectors of the basis E'' of $\mathbb{Z}_2^{n-k}$, hence $\operatorname{Im}\eta$ generates $\mathbb{Z}_2^{n-k}$ and has 2^{n-k} elements. By the Isomorphism Theorem for the $\mathbb{Z}_2$-linear map $\eta : \mathbb{Z}_2^n \to \mathbb{Z}_2^{n-k}$ we have

$$\mathbb{Z}_2^n/\operatorname{Ker}\eta \cong \operatorname{Im}\eta.$$

It follows that

$$|\operatorname{Ker}\eta| = \frac{|\mathbb{Z}_2^n|}{|\operatorname{Im}\eta|} = \frac{2^n}{2^{n-k}} = 2^k.$$

Hence $\operatorname{Im}\gamma$ is a subset of $\operatorname{Ker}\eta$, and both sets have 2^k elements. This implies that $\operatorname{Im}\gamma = \operatorname{Ker}\eta$.

(ii) A received vector $u \in \mathbb{Z}_2^n$ (or $u \in M_{n,1}(\mathbb{Z}_2)$) is a code vector if and only if $u \in \operatorname{Im}\gamma = \operatorname{Ker}\eta$ if and only if $\eta(u) = 0$ if and only if $H \cdot u = [\eta]_{E'E''} \cdot u = 0$. □

Corollary 4.4.6 *The Hamming distance d of a linear (n,k)-code is equal to the smallest number of linearly dependent columns of its parity check matrix.*

Proof. We have seen that a received vector u is a code vector if and only if $H \cdot u = 0$.

Suppose that d columns of H are linearly dependent. Let u be a vector having bits 1 in the positions corresponding to those columns and bits 0 elsewhere. Then $w(u) = d$.

Now suppose that u is a code word with $w(u) < d$. Then the bits 1 from u will give a set of $w(u)$ linearly dependent columns of H.

Hence d is equal to the smallest number of linearly dependent columns of H. □

Remark 4.4.7 Let us note some particular cases. If H has a zero column, then $d = 1$. If H has two equal columns, then $d \leq 2$. If all columns of H are distinct and non-zero, then $d \geq 3$.

Example 4.4.8 *Determine the generator matrix and the parity check matrix of the* $(3, 2)$*-parity check code, and characterize the code vectors.*

Note that $n = 3$ and $k = 2$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, that is, $\gamma : \mathbb{Z}_2^2 \to \mathbb{Z}_2^3$. The encoding of v is $\gamma(v)$.

- The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2^2$ and $\mathbb{Z}_2^3$ respectively.

  We have

  $$e_1 = (1, 0) \rightsquigarrow 10 \rightsquigarrow \boxed{1}\boxed{10} \rightsquigarrow (1, 1, 0) = \gamma(e_1),$$
  $$e_2 = (0, 1) \rightsquigarrow 01 \rightsquigarrow \boxed{1}\boxed{01} \rightsquigarrow (1, 0, 1) = \gamma(e_2).$$

  Hence

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

- The parity check matrix is

  $$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_1 & P \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

- $(u_1, u_2, u_3) \in \mathbb{Z}_2^3$ is a code word $\Leftrightarrow H \cdot [u]_{E'} = [0]_{E'} \Leftrightarrow u_1 + u_2 + u_3 = 0 \Leftrightarrow u_1 = u_2 + u_3$.

Example 4.4.9 *Determine the generator matrix and the parity check matrix of the* $(3, 1)$*-repeating code, and characterize the code vectors.*

Note that $n = 3$ and $k = 1$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, that is, $\gamma : \mathbb{Z}_2 \to \mathbb{Z}_2^3$. The encoding of v is $\gamma(v)$.

- The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2$ and $\mathbb{Z}_2^3$ respectively.

  We have

  $$e_1 = 1 \rightsquigarrow \boxed{11}\boxed{1} \rightsquigarrow (1, 1, 1) = \gamma(e_1).$$

  Hence

  $$G = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} P \\ I_1 \end{pmatrix} = \begin{pmatrix} P \\ I_k \end{pmatrix}.$$

- The parity check matrix is

  $$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_2 & P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

- $(u_1, u_2, u_3) \in \mathbb{Z}_2^3$ is a code word $\Leftrightarrow H \cdot [u]_{E'} = [0]_{E'} \Leftrightarrow u_1 + u_3 = 0$ and $u_2 + u_3 = 0$ $\Leftrightarrow u_1 = u_2 = u_3$.

Example 4.4.10 *Determine the generator matrix and the parity check matrix of the* $(6,3)$*-code generated by the polynomial*

$$p = 1 + X + X^3 \in \mathbb{Z}_2[X],$$

and characterize the code vectors.

Note that $n = 6$ and $k = 3$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, that is, $\gamma : \mathbb{Z}_2^3 \to \mathbb{Z}_2^6$. The encoding of v is $\gamma(v)$.

- The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2$ and $\mathbb{Z}_2^3$ respectively. We have

$$\begin{aligned} e_1 = (1,0,0) &\rightsquigarrow 100 \\ &\rightsquigarrow m = 1 \\ &\rightsquigarrow m \cdot X^{n-k} = X^3 \\ &\rightsquigarrow r = m \cdot X^{n-k} \bmod p = X^3 \bmod p = 1 + X \\ &\rightsquigarrow v = r + m \cdot X^{n-k} = 1 + X + X^3 \\ &\rightsquigarrow \boxed{110}\boxed{100} \\ &\rightsquigarrow (1,1,0,1,0,0) = \gamma(e_1). \end{aligned}$$

  Similarly, we obtain:

$$e_2 = (0,1,0) \rightsquigarrow (0,1,1,0,1,0) = \gamma(e_2),$$

$$e_3 = (0,0,1) \rightsquigarrow (1,1,1,0,0,1) = \gamma(e_3).$$

- Hence the generator matrix is

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} P \\ I_3 \end{pmatrix} = \begin{pmatrix} P \\ I_k \end{pmatrix}.$$

- The parity check matrix is

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_3 & P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

- $(u_1, u_2, u_3, u_4, u_5, u_6) \in \mathbb{Z}_2^6$ is a code word $\Leftrightarrow H \cdot [u]_{E'} = [0]_{E'} \Leftrightarrow$

$$\Leftrightarrow \begin{cases} u_1 + u_4 + u_6 = 0 \\ u_2 + u_4 + u_5 + u_6 = 0 \\ u_3 + u_5 + u_6 = 0 \end{cases} \Leftrightarrow \begin{cases} u_1 = u_4 + u_6 \\ u_2 = u_4 + u_5 + u_6 \\ u_3 = u_5 + u_6 \end{cases}.$$

Example 4.4.11 *Determine the Hamming distance of the* $(6,3)$*-code with parity check matrix*

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

All columns of H are distinct and non-zero, hence we have $d \geq 3$. Note that the sum of columns 1, 2 and 4 of H is the zero vector, hence $u = 110100$ is a code word of weight 3. This shows that $d = 3$. Hence this code may detect 2 errors, and correct 1 error.

4.5 Error-correcting and decoding

In this section we describe a method to correct errors and decode messages encoded with a linear (n,k)-code. Consider a linear (n,k)-code with generator and parity check matrices

$$G = \begin{pmatrix} P \\ I_k \end{pmatrix} \in M_{n,k}(\mathbb{Z}_2), \text{ and } H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} \in M_{n-k,n}(\mathbb{Z}_2)$$

respectively, where $P \in M_{n-k,k}(\mathbb{Z}_2)$. Let $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ and $\eta : \mathbb{Z}_2^n \to \mathbb{Z}_2^{n-k}$ be the $\mathbb{Z}_2$-linear maps corresponding to G and H respectively. Denote by $V = \operatorname{Im}\gamma = \operatorname{Ker}\eta$ the subspace of code vectors.

Let us start with a code vector $v \in V = \operatorname{Im}\gamma \leq \mathbb{Z}_2^n$, and assume that an error $e \in \mathbb{Z}_2^n$ occurs during transmission. Then the received vector is

$$u = v + e \in \mathbb{Z}_2^n.$$

The receiver determines the most likely transmitted vector by finding the most likely error pattern (called the *coset leader*)

$$e = u - v = u + v \in u + V.$$

The coset leader will usually be the coset containing the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's in the error pattern are bunched together as much as possible.

Definition 4.5.1 With the above notation,

$$\eta(u) \in \operatorname{Im}\eta \leq \mathbb{Z}_2^{n-k} \quad (\text{or } H \cdot u \in M_{n-k,1}(\mathbb{Z}_2))$$

is called the *syndrome* of u.

Remark 4.5.2 The number of syndromes for an (n,k)-code is

$$|\operatorname{Im}\eta| = |\mathbb{Z}_2^n/V| = |\mathbb{Z}_2^n|/|V| = \frac{2^n}{2^k} = 2^{n-k}.$$

Theorem 4.5.3 *Two vectors are in the same coset of* $\mathbb{Z}_2^n$ *by* V *if and only if they have the same syndrome.*

Proof. Let $u_1, u_2 \in \mathbb{Z}_2^n$. Then we have

$$u_1 + V = u_2 + V \iff u_1 - u_2 \in V \iff H(u_1 - u_2) = 0 \iff Hu_1 = Hu_2,$$

which shows the conclusion. □

Now we may describe a **general method for decoding** as follows:

(1) *Calculate the syndrome of the received word.*

(2) *Find the coset leader of the coset corresponding to the syndrome.*

(3) *Subtract the coset leader from the received word to obtain the most likely transmitted word.*

(4) *Drop the check digits to obtain the most likely message.*

Example 4.5.4 *Construct a table of coset leaders and syndromes for the* $(6, 3)$*-code with parity check matrix*

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

and then decode the received words 011100 *and* 100011.

We have an (n, k)-code, where $n = 6$ and $k = 3$. The number of syndromes is $2^{n-k} = 2^3 = 8$.

We write down all possible syndromes in a table, and then we determine their corresponding coset leaders.

syndrome	**coset leader**
000	
001	
010	
011	
100	
101	
110	
111	

syndrome	**coset leader**
000	000000
001	001000
010	010000
011	000010
100	100000
101	000110
110	000100
111	000001

The coset leaders (the most likely errors) are chosen such that they contain the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's are bunched together as much as possible.

We first consider the coset leader with all bits 0, then coset leaders having only one bit 1, then two consecutive bits 1, then two bits 1 not necessarily consecutive etc., until we find all correspondences with the syndromes.

We use the general matrix equality:

$$[\text{syndrome}] = H \cdot [\text{vector}].$$

- The syndrome of $u = 000000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 100000$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 010000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 001000$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.
- The syndrome of $u = 000100$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.
- The syndrome of $u = 000010$ is $H \cdot [u] = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.
- The syndrome of $u = 000001$ is $H \cdot [u] = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Note that the syndrome of a received vector u having all bits 0 except for a bit 1 in position i is exactly the column i of the parity check matrix H.

For the remaining syndrome, namely 101, we try with 110000, 011000, 001100, 000110 or 000011. The correct one is $u = 000110$, because

$$H \cdot [u] = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Note that the syndrome of $u = 000110$, which has bits 1 only in positions 3 and 4, is the sum of the columns 3 and 4 of the parity check matrix H.

To decode $u_1 = 011100$, compute its syndrome

$$H \cdot [u_1] = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Its corresponding coset leader is $e_1 = 000110$. The most likely code vector is

$$v_1 = u_1 + e_1 = 011010.$$

Hence the most likely message is 010.

To decode $u_2 = 100011$, compute its syndrome

$$H \cdot [u_2] = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Its corresponding coset leader is $e_2 = 000000$. The most likely code vector is

$$v_2 = u_2 + e_2 = 100011.$$

Hence the most likely message is 011.

Example 4.5.5 *Construct a table of coset leaders and syndromes for the* $(7, 3)$*-code generated by the polynomial*

$$p = 1 + X + X^4 \in \mathbb{Z}_2[X].$$

We have an (n, k)-code, where $n = 7$ and $k = 3$. The encoder is a $\mathbb{Z}_2$-linear map $\gamma : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$, that is, $\gamma : \mathbb{Z}_2^3 \to \mathbb{Z}_2^7$. The encoding of v is $\gamma(v)$.

The generator matrix is $G = [\gamma]_{EE'}$, where E, E' are the canonical bases of $\mathbb{Z}_2$ and $\mathbb{Z}_2^3$ respectively. We have

$$\begin{aligned} e_1 = (1,0,0) &\rightsquigarrow 100 \\ &\rightsquigarrow m = 1 \\ &\rightsquigarrow m \cdot X^{n-k} = X^4 \\ &\rightsquigarrow r = m \cdot X^{n-k} \bmod p = X^4 \bmod p = 1 + X \\ &\rightsquigarrow v = r + m \cdot X^{n-k} = 1 + X + X^4 \\ &\rightsquigarrow \boxed{1100 \mid 100} \\ &\rightsquigarrow (1,1,0,0,1,0,0) = \gamma(e_1). \end{aligned}$$

Similarly, we obtain:

$$e_2 = (0,1,0) \rightsquigarrow (0,1,1,0,0,1,0) = \gamma(e_2),$$

$$e_3 = (0,0,1) \rightsquigarrow (0,0,1,1,0,0,1) = \gamma(e_3).$$

Hence the generator mtrix is

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} P \\ I_3 \end{pmatrix} = \begin{pmatrix} P \\ I_k \end{pmatrix}.$$

The parity check matrix is

$$H = \begin{pmatrix} I_{n-k} & P \end{pmatrix} = \begin{pmatrix} I_4 & P \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

The number of syndromes is $2^{n-k} = 2^4 = 16$. We write down all possible syndromes in a table, and then we determine their corresponding coset leaders.

The coset leaders (the most likely errors) are chosen such that they contain the smallest number of 1's. If two or more error patterns are equally likely, the coset leader is chosen such that the 1's are bunched together as much as possible.

We first consider the coset leader with all bits 0, then coset leaders having only one bit 1, then two consecutive bits 1, then two bits 1 not necessarily consecutive etc., until we find all correspondences with the syndromes.

We use the general matrix equality:

$$[\text{syndrome}] = H \cdot [\text{vector}].$$

After computations, we obtain the following table:

syndrome	**coset leader**
0000	0000000
0001	0001000
0010	0010000
0011	0000001
0100	0100000
0101	0000011
0110	0000010
0111	0100001

syndrome	**coset leader**
1000	1000000
1001	1001000
1010	0000110
1011	1000001
1100	0000100
1101	0001100
1110	0010100
1111	0000101

Chapter 4 quiz

Decide whether the following statements are **true** or **false**.

1. Let C_1 be an (n, k_1)-code and let C_2 be an (n, k_2)-code such that $k_1 < k_2$. Then the information rate of C_2 is greater than the information rate of C_1.

2. Let C_1 be an (n_1, k)-code and let C_2 be an (n_2, k)-code such that $n_1 < n_2$. Then the information rate of C_2 is greater than the information rate of C_1.

3. An (n, k)-code with minimum Hamming distance 4 may detect up to 4 errors.

4. An (n, k)-code with minimum Hamming distance 4 may correct up to 2 errors.

5. The polynomial code generated by $p \in \mathbb{Z}_2[X]$ with degree $n-k$ is an (n, k)-code whose code words are the polynomials of degree less than n which divide p.

6. The check digits of a message m encoded with an (n, k)-polynomial code generated by $p \in \mathbb{Z}_2[X]$ of degree $n-k$ are stored in the coefficients of the remainder of the division of mX^{n-k} by p.

7. The encoder of an (n, k)-code is an injective function $\gamma : \mathbb{Z}_2^n \to \mathbb{Z}_2^k$.

8. Every (n, k)-code is linear.

9. Every code vector of a linear (n, k)-code may be uniquely written as a linear combination of the columns of its generator matrix.

10. A received vector of a linear (n, k)-code is a code vector if and only if the product between the generator matrix of the code and the vector is equal to zero.

11. The generator matrix of a linear (n, k)-code contains a submatrix of rank k.

12. The parity check matrix of a linear (n, k)-code always contains the identity matrix of order k as a submatrix.

13. The generator matrix of a linear (n, k)-code is invertible.

14. The number of syndromes of an (n, k)-code is greater than the number of code words.

15. Different coset leaders of an (n, k)-code may have the same corresponding syndrome.

Chapter 4 project

Use a programming language of your choice and implement the following project.

Project 4.1

- *Input*: non-zero natural numbers k and n with $k < n$ and a polynomial $p \in \mathbb{Z}_2[X]$ with $\text{degree}(p) = n - k$
- *Output*:
  1. the generator matrix G and the parity check matrix H of the (n, k)-code generated by p
  2. a table of syndromes and coset leaders for the (n, k)-code generated by p

Example:

- *Input*: $k = 3$, $n = 6$, $p = 1 + X + X^3 \in \mathbb{Z}_2[X]$
- *Output*: the matrices G, H and the table of syndromes and coset leaders are:

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad H = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix},$$

syndrome	**coset leader**
000	000000
001	001000
010	010000
011	000010
100	100000
101	000110
110	000100
111	000001

.

Bibliography

[1] S. Axler, *Linear Algebra Done Right*, Springer, New York, 2015.

[2] N. Both, *Algebră*, Lito. Univ. "Babeş-Bolyai", Cluj-Napoca, 1996.

[3] N. Both, S. Crivei, *Culegere de probleme de algebră*, Lito. Univ. "Babeş-Bolyai", Cluj-Napoca, 1997.

[4] G. Călugăreanu, *Lecţii de algebră liniară*, Lito. Univ. "Babeş-Bolyai", Cluj-Napoca, 1995.

[5] S. Crivei, *Basic Abstract Algebra*, Editura Casa Cărţii de Ştiinţă, Cluj-Napoca, 2002.

[6] J. B. Fraleigh, *A First Course in Abstract Algebra*, Addison-Wesley, Reading, 1970.

[7] J. Gilbert, L. Gilbert, *Elements of Modern Algebra*, PWS-Kent, Boston, 1992.

[8] W. J. Gilbert, W. K. Nicholson, *Modern Algebra with Applications*, John Wiley, New Jersey, 2004.

[9] J. S. Golan, *The Linear Algebra a Beginning Graduate Student Ought to Know*, Springer, Dordrecht, 2007.

[10] I. D. Ion, N. Radu, *Algebră*, Editura Didactică şi Pedagogică, Bucureşti, 1991.

[11] P. N. Klein, *Coding the Matrix. Linear Algebra through Applications to Computer Science*, Newtonian Press, 2013.

[12] D. C. Lay, *Linear Algebra and Its Applications*, Addison-Wesley, Boston, 2012.

[13] S. J. Leon, *Linear Algebra with Applications*, Pearson, Harlow, 2015.

[14] R. Lidl, G. Pilz, *Applied Abstract Algebra*, Springer, New York, 1998.

[15] MacTutor History of Mathematics Archive.
`https://mathshistory.st-andrews.ac.uk`

[16] A. Mărcuş, *Lineáris algebra*, Editura Studium, Cluj-Napoca, 1998.

[17] C. Năstăsescu, C. Niţă, C. Vraciu, *Bazele algebrei*, Editura Academiei, Bucureşti, 1986.

[18] I. Purdea, C. Pelea, *Probleme de algebră*, Editura Eikon, Cluj-Napoca, 2008.

[19] I. Purdea, I. Pop, *Algebră*, Editura GIL, Zalău, 2003.

[20] L. Robbiano, *Linear Algebra for Everyone*, Springer, Milan, 2011.

[21] G. Strang, *Linear Algebra and its Applications*, Brooks/Cole, 1988.

[22] N. Strickland, *Linear Mathematics for Applications*.
`https://neilstrickland.github.io/linear_maths/notes/linear_maths.pdf`.

Historical notes

This addendum presents some historical notes on fundamental notions and scientists (mostly mathematicians) whose names appear in this book. Our main sources were [9, 15].

Some notions

- **Matrices**

  The present terminology of "matrix" was first given in 1848 by the British mathematician James Joseph Sylvester, who was one of the main researchers in the theory of matrices and determinants.

  The 19th century Irish mathematician and physicist William Rowan Hamilton was one of the creators of modern matrix theory, together with his fellow British mathematicians Arthur Cayley and James Joseph Sylvester.

  The 19th century French mathematician Jacques Philippe Marie Binet first defined matrix multiplication.

  The 19th century German mathematician Ferdinand Georg Frobenius was the first to consider the rank of a matrix.

- **Determinants**

  The 17th century German mathematician, philosopher and diplomat Gottfried Wilhelm von Leibniz first used determinants in his research work.

  The terminology of "determinant" in the present sense was given by 19th French mathematican Auguste Louis Cauchy, who had one of the important early works on determinants.

  The common properties of determinants were first studied by 19th century German mathematician Heinrich Ferdinand Scherk.

  A systematic theory of determinants was done by the 19th century French mathematician Augustin Louis Cauchy, based on previous works.

  Later important contributions were given by Arthur Cayley and James Joseph Sylvester.

- **Vector spaces**

  The theory of vector spaces emerged from the work of the Scottish physicist James Clerk Maxwell, the German high-school teacher Herman Günter Grassmann, and the French engineer Adhémar Jean Claude Barré de Saint-Venant.

The American engineer and physicist Josiah Willard Gibbs and the British engineer Oliver Heaviside developed the theory in the 1880s.

William Rowan Hamilton first used the terminology "vector" and "scalar" in an algebraic context.

The present final form of the definition of a vector space was given by the 19th century Italian mathematician Giuseppe Peano.

- **Basis and dimension**

  Giuseppe Peano first proved that every finitely generated vector space has a basis, the notion of dimension being implicit.

  The dimension of a vector space was systematically studied by the 20th century German mathematician Hermann Klaus Hugo Weyl.

- **Linear maps**

  Linear maps between vector space were studied by Giuseppe Peano in the finite-dimensional case, and by the 19th century Italian mathematician Salvadore Pincherle in the infinite-dimensional case.

  Arthur Cayley developed the theory of matrices and their relationship to linear maps in the 19th century.

- **Linear systems**

  The 17th century Japanese mathematician Takakazu Seki Kowa developed matrix-based solving methods based on old Chinese texts.

  The 18th century Swiss mathematician Gabriel Cramer was one of the first to study the use of determinants to solve linear systems of equations.

  The 19th centry French mathematician Edmond Nicholas Laguerre wrote a book on linear systems of equations in 1867.

  The 19th century German mathematician, physicist and astronomer Johann Carl Friedrich Gauss developed his solving method in connection with his work in astronomy in 1809.

  Augusta Ada King Lovelace wrote the first computer program to solve a linear system of equations (up to 10 linear equations in 10 unknowns) by Gaussian elimination, when working for Charles Babbage's mechanical computer in the 19th century.

- **Eigenvalues and eigenvectors**

  The terminology of "eigenvalue" and "eigenvector" was given by the 19th-20th century German mathematican David Hilbert. They were previously called characteristic values and characteristic vectors by Auguste Louis Cauchy, who also studied diagonalization of matrices and similar matrices.

  Cayley-Hamilton theorem was stated by Arthur Cayley, but only proved for 2×2-matrices and 3×3-matrices by himself and William Rowan Hamilton in the 1850-1860s. The general case of the theorem was proved by the German mathematician Ferdinand Georg Frobenius in 1878.

Some scientists

- **Niels Henrik Abel** (1802–1829), Norway
- **Charles Babbage** (1791–1871), U.K.
- **Jacques Philippe Marie Binet** (1786–1856), France
- **George Boole** (1815–1864), U.K.
- **Alfredo Capelli** (1855–1910), Italy
- **Augustin Louis Cauchy** (1789–1857), France
- **Arthur Cayley** (1821–1895), U.K.
- **Gabriel Cramer** (1704–1752), Switzerland
- **Ferdinand Georg Frobenius** (1849–1917), Germany
- **Johann Carl Friedrich Gauss** (1777–1855), Germany
- **Josiah Willard Gibbs** (1839–1903), U.S.A
- **Herman Günter Grassmann** (1809–1877), Germany
- **William Rowan Hamilton** (1805–1865), Ireland
- **Richard Wesley Hamming** (1915–1998), U.S.A.
- **Oliver Heaviside** (1850–1925), U.K.
- **David Hilbert** (1862–1943), Germany
- **Marie Ennemond Camille Jordan** (1838–1922), France
- **Felix Christian Klein** (1849–1925), Germany
- **Takakazu Seki Kowa** (1642–1708), Japan
- **Leopold Kronecker** (1823–1891), Germany
- **Edmond Nicholas Laguerre** (1834–1886), France
- **Pierre-Simon Laplace** (1749–1827), France
- **Gottfried Wilhelm von Leibniz** (1646–1716), Germany
- **Augusta Ada King Lovelace** (1815–1852), U.K.
- **James Clerk Maxwell** (1831–1879), U.K.
- **Giuseppe Peano** (1858–1932), Italy
- **Salvadore Pincherle** (1853–1936), Italy

- **Eugène Rouché** (1832–1910), France
- **Adhémar Jean Claude Barré de Saint-Venant** (1797–1886), France
- **Claude Elwood Shannon** (1916–2001), U.S.A.
- **Heinrich Ferdinand Scherk** (1798–1885), Germany
- **Ernst Steinitz** (1871–1928), Germany
- **James Joseph Sylvester** (1814–1897), U.K.
- **Hermann Klaus Hugo Weyl** (1885–1955), Germany

Computer Science topics using Linear Algebra

The Association for Computing Machinery (ACM) has developed the 2012 ACM Computing Classification System for the research topics in the field of Computer Science (`https://www.acm.org`) under the form of a multi-level tree. We mention some higher level branches of this tree in which Linear Algebra has important applications.

Networks

- Network architectures
  - Network design principles
- Network types
  - Public Internet

Theory of Computation

- Models of computation
  - Quantum computation theory
- Computational complexity and cryptography
  - Cryptographic protocols
- Randomness, geometry and discrete structures
  - Error-correcting codes
- Theory and algorithms for application domains
  - Machine learning theory

Mathematics of Computing

- Information theory
  - Coding theory
- Mathematical analysis
  - Mathematical optimization

Information Systems

- World Wide Web

  - Web searching and information discovery
- Information retrieval
  - Retrieval models and ranking

Security and Privacy

- Cryptography
  - Symmetric cryptography and hash functions
- Network security
  - Security protocols

Computing Methodologies

- Machine learning
  - Machine learning approaches
- Computer graphics
  - Image manipulation

Applied Computing

- Electronic commerce
  - Online banking
- Operations research
  - Decision analysis

English-Romanian selected notions dictionary

This brief dictionary contains the terminology for the notions that have a less usual or less obvious equivalent in Romanian. The parantheses are sometimes used to mention the syntagma or the context in which the notions appear.

abelian group grup abelian, grup comutativ
adjugate matrix matrice adjunctă
augmented matrix matrice extinsă

basis bază (a unui spaţiu vectorial)

change matrix matrice de trecere
coset leader eroare asociată (unui sindrom)

direct sum sumă directă (de subspaţii)
Division Algorithm teorema împărţirii cu rest
division ring corp necomutativ

echelon form formă eşalon (a unei matrice)
eigenspace subspaţiu propriu, subspaţiu caracteristic
eigenvalue valoare proprie
eigenvector vector propriu
endomorphism ring inelul endomorfismelor (unui grup abelian)
equivalence class clasă de echivalenţă
equivalence relation (relaţie de) echivalenţă
error-correcting code cod corector de erori
error-detecting code cod detector de erori

factor (spaţiu vectorial) cât, factor
field corp comutativ

generator matrix matrice generatoare (a unui cod)
graph grafic (al unei relaţii sau funcţii)
groupoid grupoid

homomorphism morfism, omomorfism

identity element element neutru
identity function funcţia identică

identity law proprietatea elementului neutru
inverses law proprietatea elementului inversabil

kernel nucleu (al unui morfism)

linearly (in)dependent (vectori) liniar (in)dependenţi
linear map aplicaţie liniară, transformare liniară
linear space spaţiu liniar, spaţiu vectorial
linear transformation transformare liniară, aplicaţie liniară

map funcţie, aplicaţie
mapping funcţie, aplicaţie
monoid monoid

null space spaţiul nul (al unei aplicaţii liniare)

parity check matrix matrice de verificare a parităţii (a unui cod)
permutation group grup de permutări
polynomial ring inel de polinoame
power set mulţimea părţilor (unei mulţimi)

quotient (mulţime, spaţiu vectorial) cât, factor

range space spaţiul imaginii (unei aplicaţii liniare)
relation class secţiune a unei relaţii
representative reprezentant (al unei clase de echivalenţă)
residue classes modulo n clase de resturi modulo n

semigroup semigrup
similar matrices matrice similare
skew field corp necomutativ
spanned (subspaţiu) generat
stable subset parte (submulţime) stabilă
subfield subcorp
subgroup subgrup
subring subinel
subspace subspaţiu
syndrome sindrom (al unui cod)
system of generators sistem de generatori (al unui spaţiu vectorial)

transition matrix matrice de trecere

unitary homomorphism morfism unital
unitary ring inel cu unitate, inel unitar

vector space spaţiu vectorial, spaţiu liniar

Index of extra material

Throughout the book we presented some extra material, which is indexed in the order of appearance as follows.

Index

ISBN: 978-606-37-1582-2