# Basics of Algebra and Analysis For Computer Science

Jean Gallier
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@saul.cis.upenn.edu



September 5, 2001

# Contents

# Chapter 1

# Introduction

# Chapter 2

# Linear Algebra

## 2.1 Groups, Rings, and Fields

In this chapter, the basic algebraic structures (groups, rings, fields, vector spaces) are reviewed, with a major emphasis on vector spaces. Basic notions of linear algebra such as vector spaces, subspaces, linear combinations, linear independence, bases, quotient spaces, linear maps, matrices, change of bases, direct sums, linear forms, dual spaces, hyperplanes, transpose of a linear maps, are reviewed.

The set $\mathbb{R}$ of real numbers has two operations $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ (addition) and $* : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ (multiplication) satisfying properties that make $\mathbb{R}$ into an abelian group under $+$, and $\mathbb{R} - \{0\} = \mathbb{R}^*$ into an abelian group under $*$. Recall the definition of a group.

**Definition 2.1** A *group* is a set $G$ equipped with an operation $\cdot : G \times G \to G$ having the following properties: $\cdot$ is associative, has an identity element $e \in G$, and every element in $G$ is invertible (w.r.t. $\cdot$). More explicitly, this means that the following equations hold for all $a, b, c \in G$:

(G1) $a \cdot (b \cdot c) = (a \cdot b) \cdot c.$ (associativity);

(G2) $a \cdot e = e \cdot a = a.$ (identity);

(G3) For every $a \in G$, there is some $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$ (inverse).

A group $G$ is *abelian* (or *commutative*) if

$$a \cdot b = b \cdot a$$

for all $a, b \in G$.

A set $M$ together with an operation $\cdot : M \times M \to M$ and an element $e$ satisfying only conditions (G1) and (G2) is called a *monoid*. For example, the set $\mathbb{N} = \{0, 1, \ldots, n \ldots\}$ of natural numbers is a (commutative) monoid. However, it is not a group. Some examples of groups are given below.

**Example 2.1**

1. The set $\mathbb{Z} = \{\ldots, -n, \ldots, -1, 0, 1, \ldots, n \ldots\}$ of integers is a group under addition, with identity element 0. However, $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ is not a group under multiplication.

2. The set $\mathbb{Q}$ of rational numbers is a group under addition, with identity element 0. The set $\mathbb{Q}^* = \mathbb{Q} - \{0\}$ is also a group under multiplication, with identity element 1.

3. Similarly, the sets $\mathbb{R}$ of real numbers and $\mathbb{C}$ of complex numbers are groups under addition (with identity element 0), and $\mathbb{R}^* = \mathbb{R} - \{0\}$ and $\mathbb{C}^* = \mathbb{C} - \{0\}$ are groups under multiplication (with identity element 1).

4. The sets $\mathbb{R}^n$ and $\mathbb{C}^n$ of $n$-tuples of real or complex numbers are groups under componentwise addition:

$$(x_1, \ldots, x_n) + (y_1, \cdots, y_n) = (x_1 + y_n, \ldots, x_n + y_n),$$

with identity element $(0, \ldots, 0)$. All these groups are abelian.

5. Given any nonempty set $S$, the set of bijections $f \colon S \to S$, also called *permutations of* $S$, is a group under function composition (i.e., the multiplication of $f$ and $g$ is the composition $g \circ f$), with identity element the identity function $\mathrm{id}_S$. This group is not abelian as soon as $S$ has more than two elements.

6. The set of $n \times n$ matrices with real (or complex) coefficients is a group under addition of matrices, with identity element the null matrix. It is denoted by $\mathrm{M}_n(\mathbb{R})$ (or $\mathrm{M}_n(\mathbb{C})$).

7. The set $\mathbb{R}[X]$ of polynomials in one variable with real coefficients is a group under addition of polynomials.

8. The set of $n \times n$ invertible matrices with real (or complex) coefficients is a group under matrix multiplication, with identity element the identity matrix $I_n$. This group is called the *general linear group* and is usually denoted by $\mathbf{GL}(n, \mathbb{R})$ (or $\mathbf{GL}(n, \mathbb{C})$).

9. The set of $n \times n$ invertible matrices with real (or complex) coefficients and determinant $+1$ is a group under matrix multiplication, with identity element the identity matrix $I_n$. This group is called the *special linear group* and is usually denoted by $\mathbf{SL}(n, \mathbb{R})$ (or $\mathbf{SL}(n, \mathbb{C})$).

10. The set of $n \times n$ invertible matrices with real coefficients such that $RR^\top = I_n$ and of determinant $+1$ is a group called the *orthogonal group* and is usually denoted by $\mathbf{SO}(n)$ (where $R^\top$ is the *transpose* of the matrix $R$, i.e., the rows of $R^\top$ are the columns of $R$). It corresponds to the rotations in $\mathbb{R}^n$.

11. Given an open interval $]a, b[$, the set $\mathcal{C}(]a, b[)$ of continuous functions $f\colon ]a, b[\to \mathbb{R}$ is a group under the operation $f + g$ defined such that

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in ]a, b[$.

The groups $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, $\mathbb{C}$, $\mathbb{R}[X]$, and $\mathrm{M}_n(\mathbb{R})$ are more than an abelian groups, they are also commutative rings. Furthermore, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields. We now introduce rings and fields.

**Definition 2.2** A *ring* is a set $A$ equipped with two operations $+\colon A \times A \to A$ (called *addition*) and $*\colon A \times A \to A$ (called *multiplication*) having the following properties:

(R1) $A$ is an abelian group w.r.t. $+$;

(R2) $*$ is associative and has an identity element $1 \in A$;

(R3) $*$ is distributive w.r.t. $+$.

The identity element for addition is denoted 0, and the additive inverse of $a \in A$ is denoted by $-a$. More explicitly, the axioms of a ring are the following equations which hold for all $a, b, c \in A$:

$$a + (b + c) = (a + b) + c \qquad \text{(associativity of } +) \tag{2.1}$$
$$a + b = b + a \qquad \text{(commutativity of } +) \tag{2.2}$$
$$a + 0 = 0 + a = a \qquad \text{(zero)} \tag{2.3}$$
$$a + (-a) = (-a) + a = 0 \qquad \text{(additive inverse)} \tag{2.4}$$
$$a * (b * c) = (a * b) * c \qquad \text{(associativity of } *) \tag{2.5}$$
$$a * 1 = 1 * a = a \qquad \text{(identity for } *) \tag{2.6}$$
$$(a + b) * c = (a * c) + (b * c) \qquad \text{(distributivity)} \tag{2.7}$$
$$a * (b + c) = (a * b) + (a * c) \qquad \text{(distributivity)} \tag{2.8}$$

The ring $A$ is *commutative* if

$$a * b = b * a$$

for all $a, b \in A$.

From (2.7) and (2.8), we easily obtain

$$a * 0 = 0 * a = 0 \tag{2.9}$$
$$a * (-b) = (-a) * b = -(a * b). \tag{2.10}$$

Note that (2.9) implies that if $1 = 0$, then $a = 0$ for all $a \in A$, and thus, $A = \{0\}$. The ring $A = \{0\}$ is called the *trivial ring*. A ring for which $1 \neq 0$ is called *nontrivial*. The multiplication $a * b$ of two elements $a, b \in A$ is often denoted by $ab$.

**Example 2.2**

1. The additive groups $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, are commutative rings.

2. The group $\mathbb{R}[X]$ of polynomials in one variable with real coefficients is a ring under multiplication of polynomials. It is a commutative ring.

3. The group of $n \times n$ matrices $\mathrm{M}_n(\mathbb{R})$ is a ring under matrix multiplication. However, it is not a commutative ring.

4. The group $\mathcal{C}(]a, b[)$ of continuous functions $f: ]a, b[ \to \mathbb{R}$ is a ring under the operation $f \cdot g$ defined such that
$$(f \cdot g)(x) = f(x)g(x)$$
for all $x \in ]a, b[$.

When $ab = 0$ for $a \neq 0$ and $b \neq 0$, we say that $a$ is a *zero divisor* (and so is $b$). A ring $A$ is an *entire ring* (or an *integral domain*) if $0 \neq 1$, $A$ is commutative, and $ab = 0$ implies that $a = 0$ or $b = 0$, for all $a, b \in A$. In other words, an integral domain is a nontrivial commutative ring with no zero divisors.

**Example 2.3**

1. The rings $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, are integral domains.

2. The ring $\mathbb{R}[X]$ of polynomials in one variable with real coefficients is an integral domain.

3. The ring of $n \times n$ matrices $\mathrm{M}_n(\mathbb{R})$ is not an integral domain. It has zero divisors (see Example 2.12).

A homomorphism between rings is a mapping preserving addition, multiplications (and 0 and 1).

**Definition 2.3** Given two rings $A$ and $B$, a *homomorphism between $A$ and $B$* is a function $h: A \to B$ satisfying the following conditions for all $x, y \in A$:

$$\begin{aligned} h(x + y) &= h(x) + h(y) \\ h(xy) &= h(x)h(y) \\ h(0) &= 0 \\ h(1) &= 1. \end{aligned}$$

Actually, because $B$ is a group under addition, $h(0) = 0$ follows from

$$h(x + y) = h(x) + h(y).$$

**Example 2.4**

1. If $A$ is a ring, for any integer $n \in \mathbb{Z}$, for any $a \in A$, we define $n \cdot a$ by

$$n \cdot a = \underbrace{a + \cdots + a}_{n}$$

if $n \geq 0$ (with $0 \cdot a = 0$) and

$$n \cdot a = -(-n) \cdot a$$

if $n < 0$. Then, the map $h\colon \mathbb{Z} \to A$ given by

$$h(n) = n \cdot 1_A$$

is a ring homomorphism (where $1_A$ is the multiplicative identity of $A$).

2. Given any real $\lambda \in \mathbb{R}$, the evaluation map $\eta_\lambda\colon \mathbb{R}[X] \to \mathbb{R}$ defined by

$$\eta_\lambda(f(X)) = f(\lambda)$$

for every polynomial $f(X) \in \mathbb{R}[X]$ is a ring homomorphism.

A field is a commutative ring $K$ for which $A - \{0\}$ is a group under multiplication.

**Definition 2.4** A set $K$ is a *field* if it is a ring and the following properties hold:

(F1) $0 \neq 1$;

(F2) $K^* = K - \{0\}$ is a group w.r.t. $*$ (i.e., every $a \neq 0$ has an inverse w.r.t. $*$);

(F3) $*$ is commutative.

If $*$ is not commutative but (F1) and (F2) hold, we say that we have a *skew field* (or *noncommutative field*).

Note that we are assuming that the operation $*$ of a field is commutative. This convention is not universally adopted, but since $*$ will be commutative for most fields we will encounter, we may as well include this condition in the definition.

**Example 2.5**

1. The rings $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are fields.
2. The set of (formal) fractions $f(X)/g(X)$ of polynomials $f(X), g(X) \in \mathbb{R}[X]$, where $g(X)$ is not the null polynomial, is a field.
3. The ring $\mathcal{C}^*(]a, b[)$ of continuous functions $f\colon ]a, b[ \to \mathbb{R}$ such that $f(x) \neq 0$ for all $x \in ]a, b[$ is a field.

## 2.2 Vector Spaces

For every $n \geq 1$, let $\mathbb{R}^n$ be the set of $n$-tuples $x = (x_1, \ldots, x_n)$. Addition can be extended to $\mathbb{R}^n$ as follows:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n).$$

We can also define an operation $\cdot \colon \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ as follows:

$$\lambda \cdot (x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

The resulting algebraic structure has some interesting properties, those of a vector space.

**Definition 2.5** Given a field $K$, a *vector space over $K$* (or *$K$-vector space*) is a set $E$ (of vectors) together with two operations $+\colon E \times E \to E$ (called *vector addition*),[1] and $\cdot\colon K \times E \to E$ (called *scalar multiplication*) satisfying the following conditions for all $\alpha, \beta \in K$ and all $u, v \in E$;

(V0) $E$ is an abelian group w.r.t. $+$, with identity element $0$;

(V1) $\alpha \cdot (u + v) = (\alpha \cdot u) + (\alpha \cdot v)$;

(V2) $(\alpha + \beta) \cdot u = (\alpha \cdot u) + (\beta \cdot u)$;

(V3) $(\alpha * \beta) \cdot u = \alpha \cdot (\beta \cdot u)$;

(V4) $1 \cdot u = u$.

Given $\alpha \in K$ and $v \in E$, the element $\alpha \cdot v$ is also denoted by $\alpha v$. The field $K$ is often called the field of scalars.

Unless specified otherwise or unless we are dealing with several different fields, in the rest of this chapter, we assume that all $K$-vector spaces are defined with respect to a fixed field $K$. Thus, we will refer to a $K$-vector space simply as a vector space. In most cases, the field $K$ will be the field $\mathbb{R}$ of reals.

From (V0), a vector space always contains the null vector $0$, and thus is nonempty. From (V1), we get $\alpha \cdot 0 = 0$, and $\alpha \cdot (-v) = -(\alpha \cdot v)$. From (V2), we get $0 \cdot v = 0$, and $(-\alpha) \cdot v = -(\alpha \cdot v)$. The field $K$ itself can be viewed as a vector space over itself, addition of vectors being addition in the field, and multiplication by a scalar being multiplication in the field.

**Example 2.6**

1. The fields $\mathbb{R}$ and $\mathbb{C}$ are vector spaces over $\mathbb{R}$.

[1]The symbol $+$ is overloaded, since it denotes both addition in the field $K$ and addition of vectors in $E$. However, if we write elements of $E$ as vectors, i.e., of the form $u$, it will always be clear from the type of the arguments which $+$ is intended.

2. The groups $\mathbb{R}^n$ and $\mathbb{C}^n$ are vector spaces over $\mathbb{R}$, and $\mathbb{C}^n$ is a vector space over $\mathbb{C}$.

3. The ring $\mathbb{R}[X]$ of polynomials is a vector space over $\mathbb{R}$, and $\mathbb{C}[X]$ is a vector space over $\mathbb{R}$ and $\mathbb{C}$. The ring of $n \times n$ matrices $\mathrm{M}_n(\mathbb{R})$ is a vector space over $\mathbb{R}$.

4. The ring $\mathcal{C}^*(]a,b[)$ of continuous functions $f\colon ]a,b[ \to \mathbb{R}$ is a vector space over $\mathbb{R}$.

Let $E$ be a vector space. We would like to define the important notions of linear combination and linear independence. These notions can be defined for sets of vectors in $E$, but it will turn out to be more convenient to define them for families $(v_i)_{i\in I}$, where $I$ is any arbitrary index set.

## 2.3 Linear Independence, Subspaces

Given a set $A$, a family $(a_i)_{i\in I}$ of elements of $A$ is simply a function $a\colon I \to A$.

**Remark:** When considering a family $(a_i)_{i\in I}$, there is no reason to assume that $I$ is ordered. The crucial point is that every element of the family is uniquely indexed by an element of $I$. Thus, unless specified otherwise, we do not assume that the elements of an index set are ordered.

If $A$ is an abelian group (usually, when $A$ is a ring or a vector space) with identity 0, we say that a family $(a_i)_{i\in I}$ has *finite support* if $a_i = 0$ for all $i \in I - J$, where $J$ is a finite subset of $I$ (the support of the family).

We can deal with an arbitrary set $X$ by viewing it as the family $(X_x)_{x\in X}$ corresponding to the identity function $\mathrm{id}\colon X \to X$. We agree that when $I = \emptyset$, $(a_i)_{i\in I} = \emptyset$. A family $(a_i)_{i\in I}$ is finite if $I$ is finite.

Given two disjoint sets $I$ and $J$, the union of two families $(u_i)_{i\in I}$ and $(v_j)_{j\in J}$, denoted as $(u_i)_{i\in I} \cup (v_j)_{j\in J}$, is the family $(w_k)_{k\in(I\cup J)}$ defined such that $w_k = u_k$ if $k \in I$, and $w_k = v_k$ if $k \in J$. Given a family $(u_i)_{i\in I}$ and any element $v$, we denote by $(u_i)_{i\in I} \cup_k (v)$ the family $(w_i)_{i\in I\cup\{k\}}$ defined such that, $w_i = u_i$ if $i \in I$, and $w_k = v$, where $k$ is any index such that $k \notin I$. Given a family $(u_i)_{i\in I}$, a subfamily of $(u_i)_{i\in I}$ is a family $(u_j)_{j\in J}$ where $J$ is any subset of $I$.

In this chapter, unless specified otherwise, it is assumed that all families of scalars have finite support.

**Definition 2.6** Let $E$ be a vector space. A vector $v \in E$ is a *linear combination of a family $(u_i)_{i\in I}$ of elements of $E$* if there is a family $(\lambda_i)_{i\in I}$ of scalars in $K$ such that

$$v = \sum_{i\in I} \lambda_i u_i.$$

When $I = \emptyset$, we stipulate that $v = 0$. We say that a family $(u_i)_{i\in I}$ is *linearly independent* if for every family $(\lambda_i)_{i\in I}$ of scalars in $K$,

$$\sum_{i\in I} \lambda_i u_i = 0 \quad \text{implies that} \quad \lambda_i = 0 \text{ for all } i \in I.$$

Equivalently, a family $(u_i)_{i\in I}$ is *linearly dependent* if there is some family $(\lambda_i)_{i\in I}$ of scalars in $K$ such that

$$\sum_{i\in I} \lambda_i u_i = 0 \quad \text{and} \quad \lambda_j \neq 0 \text{ for some } j \in I.$$

We agree that when $I = \emptyset$, the family $\emptyset$ is linearly independent.

A family $(u_i)_{i\in I}$ is linearly dependent iff some $u_j$ in the family can be expressed as a linear combination of the other vectors in the family. Indeed, there is some family $(\lambda_i)_{i\in I}$ of scalars in $K$ such that

$$\sum_{i\in I} \lambda_i u_i = 0 \quad \text{and} \quad \lambda_j \neq 0 \text{ for some } j \in I,$$

which implies that

$$u_j = \sum_{i\in(I-\{j\})} -\lambda_j^{-1}\lambda_i u_i.$$

When $I$ is nonempty, if the family $(u_i)_{i\in I}$ is linearly independent, note that $u_i \neq 0$ for all $i \in I$, since otherwise we would have $\sum_{i\in I} \lambda_i u_i = 0$ with some $\lambda_i \neq 0$, since $0u_i = 0$.

**Example 2.7**

1. Any two disinct scalars $\lambda, \mu \neq 0$ in $K$ are linearly independent.
2. In $\mathbb{R}^3$, the vectors $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ are linearly independent.
3. In $\mathbb{R}^4$, the vectors $(1, 1, 1, 1)$, $(0, 1, 1, 1)$, $(0, 0, 1, 1)$, and $(0, 0, 0, 1)$ are linearly independent.
4. In $\mathbb{R}^2$, the vectors $u = (1, 1)$, $v = (0, 1)$ and $w = (2, 3)$ are linearly dependent, since

$$w = 2u + v.$$

Note that a family $(u_i)_{i\in I}$ is linearly independent iff $(u_j)_{j\in J}$ is linearly independent for every finite subset $J$ of $I$ (even when $I = \emptyset$). Indeed, when $\sum_{i\in I} \lambda_i u_i = 0$, the family $(\lambda_i)_{i\in I}$ of scalars in $K$ has finite support, and thus $\sum_{i\in I} \lambda_i u_i = 0$ really means that $\sum_{j\in J} \lambda_j u_j = 0$ for a finite subset $J$ of $I$. When $I$ is finite, we often assume that it is the set $I = \{1, 2, \ldots, n\}$. In this case, we denote the family $(u_i)_{i\in I}$ as $(u_1, \ldots, u_n)$.

The notion of a subspace of a vector space is defined as follows.

**Definition 2.7** Given a vector space $E$, a subset $F$ of $E$ is a *linear subspace* (or *subspace*) of $E$ if $F$ is nonempty and $\lambda u + \mu v \in F$ for all $u, v \in F$, and all $\lambda, \mu \in K$.

It is easy to see that a subspace $F$ of $E$ is indeed a vector space, since the restriction of $+ \colon E \times E \to E$ to $F \times F$ is indeed a function $+ \colon F \times F \to F$, and the restriction of $\cdot \colon K \times E \to E$ to $K \times F$ is indeed a function $\cdot \colon K \times F \to F$. It is also easy to see that a subspace $F$ of $E$ is closed under arbitrary linear combinations of vectors from $F$, and that any intersection of subspaces is a subspace. Letting $\lambda = \mu = 0$, we see that every subspace contains the vector 0. The subspace $\{0\}$ will be denoted by $(0)$, or even 0 (with a mild abuse of notation).

**Example 2.8**

1. In $\mathbb{R}^2$, the set of vectors $u = (x, y)$ such that
$$x + y - 1 = 0$$
is a subspace.

2. In $\mathbb{R}^3$, the set of vectors $u = (x, y, z)$ such that
$$x + y + z - 1 = 0$$
is a subspace.

3. For any $n \geq 0$, the set of polynomials $f(X) \in \mathbb{R}[X]$ of degree at most $n$ is a subspace of $\mathbb{R}[X]$.

4. The set of upper triangular $n \times n$ matrices is a subspace of the space of $n \times n$ matrices.

## 2.4 Bases of a Vector Space

Given a vector space $E$, given a family $(v_i)_{i \in I}$, the subset $V$ of $E$ consisting of the null vector 0 and of all linear combinations of $(v_i)_{i \in I}$ is easily seen to be a subspace of $E$. Subspaces having such a "generating family" play an important role, and motivate the following definition.

**Definition 2.8** Given a vector space $E$ and a subspace $V$ of $E$, a family $(v_i)_{i \in I}$ of vectors $v_i \in V$ *spans* $V$ or *generates* $V$ if for every $v \in V$, there is some family $(\lambda_i)_{i \in I}$ of scalars in $K$ such that
$$v = \sum_{i \in I} \lambda_i v_i.$$
We also say that the elements of $(v_i)_{i \in I}$ are *generators* of $V$ and that $V$ is *spanned by* $(v_i)_{i \in I}$, or *generated by* $(v_i)_{i \in I}$. If a subspace $V$ of $E$ is generated by a finite family $(v_i)_{i \in I}$, we say that $V$ is *finitely generated*. A family $(u_i)_{i \in I}$ that spans $V$ and is linearly independent is called a *basis* of $V$.

**Example 2.9**

1. In $\mathbb{R}^3$, the vectors $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ form a basis.

2. In the subspace of polynomials in $\mathbb{R}[X]$ of degree at most $n$, the polynomials $1, X, X^2$, $\ldots, X^n$ form a basis.

3. The polynomials $\binom{n}{k}(1-X)^k X^{n-k}$ for $k = 0, \ldots, n$, also form a basis of that space.

It is a standard result of linear algebra that every vector space $E$ has a basis, and that for any two bases $(u_i)_{i\in I}$ and $(v_j)_{j\in J}$, $I$ and $J$ have the same cardinality. In particular, if $E$ has a finite basis of $n$ elements, every basis of $E$ has $n$ elements, and the integer $n$ is called the *dimension* of the vector space $E$. We prove the above result in the case where a vector space is generated by a finite family. We begin with a crucial lemma.

**Lemma 2.1** *Given a linearly independent family $(u_i)_{i\in I}$ of elements of a vector space $E$, if $v \in E$ is not a linear combination of $(u_i)_{i\in I}$, then the family $(u_i)_{i\in I} \cup_k (v)$ obtained by adding $v$ to the family $(u_i)_{i\in I}$ is linearly independent (where $k \notin I$).*

*Proof*. Assume that $\mu v + \sum_{i\in I} \lambda_i u_i = 0$, for any family $(\lambda_i)_{i\in I}$ of scalars in $K$. If $\mu \neq 0$, then $\mu$ has an inverse (because $K$ is a field), and thus we have $v = -\sum_{i\in I}(\mu^{-1}\lambda_i)u_i$, showing that $v$ is a linear combination of $(u_i)_{i\in I}$ and contradicting the hypothesis. Thus, $\mu = 0$. But then, we have $\sum_{i\in I} \lambda_i u_i = 0$, and since the family $(u_i)_{i\in I}$ is linearly independent, we have $\lambda_i = 0$ for all $i \in I$. $\square$

Note that the proof of the above lemma holds for arbitrary vector spaces, not only for finitely generated vector spaces. The next theorem also holds in general, but the proof is more sophisticated for vector spaces that do not have a finite set of generators (it uses Zorn's lemma). Thus, we only prove the theorem for finitely generated vector spaces.

**Theorem 2.2** *Given any finite family $S = (u_i)_{i\in I}$ generating a vector space $E$ and any linearly independent subfamily $L = (u_j)_{j\in J}$ of $S$ (where $J \subseteq I$), there is a basis $B$ of $E$ such that $L \subseteq B \subseteq S$.*

*Proof*. Consider the set of linearly independent families $B$ such that $L \subseteq B \subseteq S$. Since this set is nonempty and finite, it has some maximal element, say $B = (u_h)_{h\in H}$. We claim that $B$ generates $E$. Indeed, if $B$ does not generate $E$, then there is some $u_p \in S$ that is not a linear combination of vectors in $B$ (since $S$ generates $E$), with $p \notin H$. Then, by Lemma 2.1, the family $B' = (u_h)_{h\in H\cup\{p\}}$ is linearly independent, and since $L \subseteq B \subset B' \subseteq S$, this contradicts the maximality of $B$. Thus, $B$ is a basis of $E$ such that $L \subseteq B \subseteq S$. $\square$

Theorem 2.2 also holds for vector spaces that are not finitely generated. In this case, the problem is to guarantee the existence of a maximal linearly independent family $B$ such that $L \subseteq B \subseteq S$. The existence of such a maximal family can be shown using Zorn's lemma.

Readers who want to learn more about Zorn's lemma and the proof of Theorem 2.2 should consult either Lang [18], Appendix 2, §2 (pp. 878-884) and Chapter III, §5 (pp. 139-140), or Artin [1], Appendix §1 (pp. 588-589).

The following proposition giving useful properties characterizing a basis is an immediate consequence of Theorem 2.2.

**Proposition 2.3** *Given a vector space $E$, for any family $B = (v_i)_{i\in I}$ of vectors of $E$, the following properties are equivalent:*

*(1) $B$ is a basis of $E$.*

*(2) $B$ is a maximal linearly independent family of $E$.*

*(3) $B$ is a minimal generating family of $E$.*

The following *replacement proposition* shows the relationship between finite linearly independent families and finite families of generators of a vector space.

**Proposition 2.4** *Given a vector space $E$, let $(u_i)_{i\in I}$ be any finite linearly independent family in $E$, where $|I| = m$, and let $(v_j)_{j\in J}$ be any finite family such that every $u_i$ is a linear combination of $(v_j)_{j\in J}$, where $|J| = n$. Then, there exists a set $L$ and an injection $\rho\colon L \to J$ such that $L \cap I = \emptyset$, $|L| = n - m$, and the families $(u_i)_{i\in I} \cup (v_{\rho(l)})_{l\in L}$ and $(v_j)_{j\in J}$ generate the same subspace of $E$. In particular, $m \leq n$.*

*Proof*. We proceed by induction on $|I| = m$. When $m = 0$, the family $(u_i)_{i\in I}$ is empty, and the proposition holds trivially with $L = J$ ($\rho$ is the identity). Assume $|I| = m+1$. Consider the linearly independent family $(u_i)_{i\in(I-\{p\})}$, where $p$ is any member of $I$. By the induction hypothesis, there exists a set $L$ and an injection $\rho\colon L \to J$ such that $L \cap (I - \{p\}) = \emptyset$, $|L| = n-m$, and the families $(u_i)_{i\in(I-\{p\})} \cup (v_{\rho(l)})_{l\in L}$ and $(v_j)_{j\in J}$ generate the same subspace of $E$. If $p \in L$, we can replace $L$ by $(L - \{p\}) \cup \{p'\}$ where $p'$ does not belong to $I \cup L$, and replace $\rho$ by the injection $\rho'$ which agrees with $\rho$ on $L - \{p\}$ and such that $\rho'(p') = \rho(p)$. Thus, we can always assume that $L \cap I = \emptyset$. Since $u_p$ is a linear combination of $(v_j)_{j\in J}$ and the families $(u_i)_{i\in(I-\{p\})} \cup (v_{\rho(l)})_{l\in L}$ and $(v_j)_{j\in J}$ generate the same subspace of $E$, $u_p$ is a linear combination of $(u_i)_{i\in(I-\{p\})} \cup (v_{\rho(l)})_{l\in L}$. Let

$$u_p = \sum_{i\in(I-\{p\})} \lambda_i u_i + \sum_{l\in L} \lambda_l v_{\rho(l)}. \tag{1}$$

If $\lambda_l = 0$ for all $l \in L$, we have

$$\sum_{i\in(I-\{p\})} \lambda_i u_i - u_p = 0,$$

contradicting the fact that $(u_i)_{i\in I}$ is linearly independent. Thus, $\lambda_l \neq 0$ for some $l \in L$, say $l = q$. Since $\lambda_q \neq 0$, we have

$$v_{\rho(q)} = \sum_{i\in(I-\{p\})} (-\lambda_q^{-1}\lambda_i)u_i + \lambda_q^{-1}u_p + \sum_{l\in(L-\{q\})} (-\lambda_q^{-1}\lambda_l)v_{\rho(l)}. \tag{2}$$

We claim that the families $(u_i)_{i\in(I-\{p\})} \cup (v_{\rho(l)})_{l\in L}$ and $(u_i)_{i\in I} \cup (v_{\rho(l)})_{l\in(L-\{q\})}$ generate the same subset of $E$. Indeed, the second family is obtained from the first by replacing $v_{\rho(q)}$ by $u_p$, and vice-versa, and $u_p$ is a linear combination of $(u_i)_{i\in(I-\{p\})} \cup (v_{\rho(l)})_{l\in L}$, by (1), and $v_{\rho(q)}$ is a linear combination of $(u_i)_{i\in I} \cup (v_{\rho(l)})_{l\in(L-\{q\})}$, by (2). Thus, the families $(u_i)_{i\in I} \cup (v_{\rho(l)})_{l\in(L-\{q\})}$ and $(v_j)_{j\in J}$ generate the same subspace of $E$, and the proposition holds for $L - \{q\}$ and the restriction of the injection $\rho\colon L \to J$ to $L - \{q\}$, since $L \cap I = \emptyset$ and $|L| = n - m$ imply that $(L - \{q\}) \cap I = \emptyset$ and $|L - \{q\}| = n - (m + 1)$. $\square$

Actually, one can prove that Proposition 2.4 implies Theorem 2.2 when the vector space is finitely generated. Putting Theorem 2.2 and Proposition 2.4 together, we obtain the following fundamental theorem.

**Theorem 2.5** *Let $E$ be a finitely generated vector space. Any family $(u_i)_{i\in I}$ generating $E$ contains a subfamily $(u_j)_{j\in J}$ which is a basis of $E$. Furthermore, for every two bases $(u_i)_{i\in I}$ and $(v_j)_{j\in J}$ of $E$, we have $|I| = |J| = n$.*

*Proof*. The first part follows immediately by applying Theorem 2.2 with $L = \emptyset$ and $S = (u_i)_{i\in I}$. Assume that $(u_i)_{i\in I}$ and $(v_j)_{j\in J}$ are bases of $E$. Since $(u_i)_{i\in I}$ is linearly independent and $(v_j)_{j\in J}$ spans $E$, proposition 2.4 implies that $|I| \leq |J|$. A symmetric argument yields $|J| \leq |I|$. $\square$

**Remark:** Theorem 2.5 also holds for vector spaces that are not finitely generated. This can be shown as follows. Let $(u_i)_{i\in I}$ be a basis of $E$, let $(v_j)_{j\in J}$ be a generating family of $E$, and assume that $I$ is infinite. For every $j \in J$, let $L_j \subseteq I$ be the finite set

$$L_j = \{i \in I \mid v_j = \sum_{i\in I} v_i u_i,\ v_i \neq 0\}.$$

Let $L = \bigcup_{j\in J} L_j$. Since $(u_i)_{i\in I}$ is a basis of $E$, we must have $I = L$, since otherwise $(u_i)_{i\in L}$ would be another basis of $E$, and by Lemma 2.1, this would contradict the fact that $(u_i)_{i\in I}$ is linearly independent. Furthermore, $J$ must be infinite, since otherwise, because the $L_j$ are finite, $I$ would be finite. But then, since $I = \bigcup_{j\in J} L_j$ with $J$ infinite and the $L_j$ finite, by a standard result of set theory, $|I| \leq |J|$. If $(v_j)_{j\in J}$ is also a basis, by a symmetric argument, we obtain $|J| \leq |I|$, and thus, $|I| = |J|$ for any two bases $(u_i)_{i\in I}$ and $(v_j)_{j\in J}$ of $E$.

When $|I|$ is infinite, we say that $E$ is of infinite dimension, the dimension $|I|$ being a cardinal number which depends only on the vector space $E$. The dimension of a vector space $E$ is denoted by $\dim(E)$. Clearly, if the field $K$ itself is viewed as a vector space, then every family $(a)$ where $a \in K$ and $a \neq 0$ is a basis. Thus $\dim(K) = 1$.

Let $(u_i)_{i\in I}$ be a basis of a vector space $E$. For any vector $v \in E$, since the family $(u_i)_{i\in I}$ generates $E$, there is a family $(\lambda_i)_{i\in I}$ of scalars in $K$, such that

$$v = \sum_{i\in I} \lambda_i u_i.$$

A very important fact is that the family $(\lambda_i)_{i\in I}$ is **unique**.

**Proposition 2.6** *Given a vector space $E$, let $(u_i)_{i\in I}$ be a family of vectors in $E$. Let $v \in E$, and assume that $v = \sum_{i\in I} \lambda_i u_i$. Then, the family $(\lambda_i)_{i\in I}$ of scalars such that $v = \sum_{i\in I} \lambda_i u_i$ is unique iff $(u_i)_{i\in I}$ is linearly independent.*

*Proof*. First, assume that $(u_i)_{i\in I}$ is linearly independent. If $(\mu_i)_{i\in I}$ is another family of scalars in $K$ such that $v = \sum_{i\in I} \mu_i u_i$, then we have

$$\sum_{i\in I} (\lambda_i - \mu_i) u_i = 0,$$

and since $(u_i)_{i\in I}$ is linearly independent, we must have $\lambda_i - \mu_i = 0$ for all $i \in I$, that is, $\lambda_i = \mu_i$ for all $i \in I$. The converse is shown by contradiction. If $(u_i)_{i\in I}$ was linearly dependent, there would be a family $(\mu_i)_{i\in I}$ of scalars not all null such that

$$\sum_{i\in I} \mu_i u_i = 0$$

and $\mu_j \neq 0$ for some $j \in I$. But then,

$$v = \sum_{i\in I} \lambda_i u_i + 0 = \sum_{i\in I} \lambda_i u_i + \sum_{i\in I} \mu_i u_i = \sum_{i\in I} (\lambda_i + \mu_i) u_i,$$

with $\lambda_j \neq \lambda_j + \mu_j$ since $\mu_j \neq 0$, contradicting the assumption that $(\lambda_i)_{i\in I}$ is the unique family such that $v = \sum_{i\in I} \lambda_i u_i$. $\square$

If $(u_i)_{i\in I}$ is a basis of a vector space $E$, for any vector $v \in E$, if $(v_i)_{i\in I}$ is the unique family of scalars in $K$ such that

$$v = \sum_{i\in I} v_i u_i,$$

each $v_i$ is called the *component (or coordinate) of index $i$ of $v$ with respect to the basis $(u_i)_{i\in I}$*.

Given a field $K$ and any (nonempty) set $I$, we can form a vector space $K^{(I)}$ which, in some sense, is the standard vector space of dimension $|I|$.

**Definition 2.9** Given a field $K$ and any (nonempty) set $I$, let $K^{(I)}$ be the subset of the cartesian product $K^I$ consisting of all families $(\lambda_i)_{i\in I}$ with finite support of scalars in $K$.[2] We define addition and multiplication by a scalar as follows:

$$(\lambda_i)_{i\in I} + (\mu_i)_{i\in I} = (\lambda_i + \mu_i)_{i\in I},$$

[2]Where $K^I$ denotes the set of all functions from $I$ to $K$.

and

$$\lambda \cdot (\mu_i)_{i\in I} = (\lambda\mu_i)_{i\in I}.$$

It is immediately verified that addition and multiplication by a scalar are well defined. Thus, $K^{(I)}$ is a vector space. Furthermore, because families with finite support are considered, the family $(e_i)_{i\in I}$ of vectors $e_i$, defined such that $(e_i)_j = 0$ if $j \neq i$ and $(e_i)_i = 1$, is clearly a basis of the vector space $K^{(I)}$. When $I = \{1, \ldots, n\}$, we denote $K^{(I)}$ by $K^n$. The function $\iota: I \to K^{(I)}$, such that $\iota(i) = e_i$ for every $i \in I$, is clearly an injection.

When $I$ is a finite set, $K^{(I)} = K^I$, but this is false when $I$ is infinite. In fact, $\dim(K^{(I)}) = |I|$, but $\dim(K^I)$ is strictly greater when $I$ is infinite.

Many interesting mathematical structures are vector spaces. A very important example is the set of linear maps between two vector spaces to be defined in the next section. Here is an example that will prepare us for the vector space of linear maps.

**Example 2.10** Let $X$ be any nonempty set and let $E$ be a vector space. The set of all functions $f: X \to E$ can be made into a vector space as follows: Given any two functions $f: X \to E$ and $g: X \to E$, let $(f + g): X \to E$ be defined such that

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in X$, and for every $\lambda \in K$, let $\lambda f: X \to E$ be defined such that

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in X$. The axioms of a vector space are easily verified. Now, let $E = K$, and let $I$ be the set of all nonempty subsets of $X$. For every $S \in I$, let $f_S: X \to E$ be the function such that $f_S(x) = 1$ iff $x \in S$, and $f_S(x) = 0$ iff $x \notin S$. We leave as an exercise to show that $(f_S)_{S\in I}$ is linearly independent.

## 2.5 Linear Maps

A function between two vector spaces that preserves the vector space structure is called a homomorphism of vector spaces, or linear map. Linear maps formalize the concept of linearity of a function. In the rest of this section, we assume that all vector spaces are over a given field $K$ (say $\mathbb{R}$).

**Definition 2.10** Given two vector spaces $E$ and $F$, a *linear map* between $E$ and $F$ is a function $f: E \to F$ satisfying the following two conditions:

$$\begin{aligned} f(x + y) &= f(x) + f(y) && \text{for all } x, y \in E; \\ f(\lambda x) &= \lambda f(x) && \text{for all } \lambda \in K,\ x \in E. \end{aligned}$$

Setting $x = y = 0$ in the first identity, we get $f(0) = 0$. The basic property of linear maps is that they transform linear combinations into linear combinations. Given a family $(u_i)_{i\in I}$ of vectors in $E$, given any family $(\lambda_i)_{i\in I}$ of scalars in $K$, we have

$$f(\sum_{i\in I}\lambda_i u_i) = \sum_{i\in I}\lambda_i f(u_i).$$

The above identity is shown by induction on the size of the support of the family $(\lambda_i u_i)_{i\in I}$, using the properties of Definition 2.10.

**Example 2.11**

1. The map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined such that

$$\begin{aligned} x' &= x - y \\ y' &= x + y \end{aligned}$$

   is a linear map.

2. The map $D\colon \mathbb{R}[X] \to \mathbb{R}[X]$ defined such that

$$D(f(X)) = f'(X),$$

   where $f'(X)$ is the derivative of the polynomial $f(X)$, is a linear map

Given a linear map $f\colon E \to F$, we define its *image (or range)* $\operatorname{Im} f = f(E)$, as the set

$$\operatorname{Im} f = \{y \in F \mid f(x) = y,\ \text{for some } x \in E\},$$

and its *Kernel (or nullspace)* $\operatorname{Ker} f = f^{-1}(0)$, as the set

$$\operatorname{Ker} f = \{x \in E \mid f(x) = 0\}.$$

**Proposition 2.7** *Given a linear map $f\colon E \to F$, the set $\operatorname{Im} f$ is a subspace of $F$ and the set $\operatorname{Ker} f$ is a subspace of $E$. The linear map $f\colon E \to F$ is injective iff $\operatorname{Ker} f = 0$ (where $0$ is the trivial subspace $\{0\}$).*

*Proof*. Given any $x, y \in \operatorname{Im} f$, there are some $u, v \in E$ such that $x = f(u)$ and $y = f(v)$, and for all $\lambda, \mu \in K$, we have

$$f(\lambda u + \mu v) = \lambda f(u) + \mu f(v) = \lambda x + \mu y,$$

and thus, $\lambda x + \mu y \in \operatorname{Im} f$, showing that $\operatorname{Im} f$ is a subspace of $F$. Given any $x, y \in \operatorname{Ker} f$, we have $f(x) = 0$ and $f(y) = 0$, and thus,

$$f(\lambda x + \mu y) = \lambda f(x) + \mu f(y) = 0,$$

that is, $\lambda x + \mu y \in \operatorname{Ker} f$, showing that $\operatorname{Ker} f$ is a subspace of $E$. Note that $f(x) = f(y)$ iff $f(x - y) = 0$. Thus, $f$ is injective iff $\operatorname{Ker} f = 0$. $\square$

A fundamental property of bases in a vector space is that they allow the definition of linear maps as unique homomorphic extensions, as shown in the following proposition.

**Proposition 2.8** *Given any two vector spaces $E$ and $F$, given any basis $(u_i)_{i\in I}$ of $E$, given any other family of vectors $(v_i)_{i\in I}$ in $F$, there is a unique linear map $f\colon E \to F$ such that $f(u_i) = v_i$ for all $i \in I$. Furthermore, $f$ is injective iff $(v_i)_{i\in I}$ is linearly independent, and $f$ is surjective iff $(v_i)_{i\in I}$ generates $F$.*

*Proof*. If such a linear map $f\colon E \to F$ exists, since $(u_i)_{i\in I}$ is a basis of $E$, every vector $x \in E$ can written uniquely as a linear combination

$$x = \sum_{i\in I} x_i u_i,$$

and by linearity, we must have

$$f(x) = \sum_{i\in I} x_i f(u_i) = \sum_{i\in I} x_i v_i.$$

Define the function $f\colon E \to F$, by letting

$$f(x) = \sum_{i\in I} x_i v_i$$

for every $x = \sum_{i\in I} x_i u_i$. It is easy to verify that $f$ is indeed linear, it is unique by the previous reasoning, and obviously, $f(u_i) = v_i$.

Now, assume that $f$ is injective. Let $(\lambda_i)_{i\in I}$ be any family of scalars, and assume that

$$\sum_{i\in I} \lambda_i v_i = 0.$$

Since $v_i = f(u_i)$ for every $i \in I$, we have

$$f(\sum_{i\in I} \lambda_i u_i) = \sum_{i\in I} \lambda_i f(u_i) = \sum_{i\in I} \lambda_i v_i = 0.$$

Since $f$ is injective iff $\mathrm{Ker}\, f = 0$, we have

$$\sum_{i\in I} \lambda_i u_i = 0,$$

and since $(u_i)_{i\in I}$ is a basis, we have $\lambda_i = 0$ for all $i \in I$, and $(v_i)_{i\in I}$ is linearly independent. Conversely, assume that $(v_i)_{i\in I}$ is linearly independent. If

$$f(\sum_{i\in I} \lambda_i u_i) = 0,$$

then

$$\sum_{i\in I} \lambda_i v_i = \sum_{i\in I} \lambda_i f(u_i) = f(\sum_{i\in I} \lambda_i u_i) = 0,$$

and $\lambda_i = 0$ for all $i \in I$, since $(v_i)_{i\in I}$ is linearly independent. Since $(u_i)_{i\in I}$ is a basis of $E$, we just showed that $\mathrm{Ker}\, f = 0$, and $f$ is injective. The part where $f$ is surjective is left as a simple exercise. $\square$

By the second part of Proposition 2.8, an injective linear map $f\colon E \to F$ sends a basis $(u_i)_{i\in I}$ to a linearly independent family $(f(u_i))_{i\in I}$ of $F$, which is also a basis when $f$ is bijective. Also, when $E$ and $F$ have the same finite dimension $n$, $(u_i)_{i\in I}$ is a basis of $E$, and $f\colon E \to F$ is injective, then $(f(u_i))_{i\in I}$ is a basis of $F$ (by Proposition 2.3).

We can now show that the vector space $K^{(I)}$ of Definition 2.9 has a universal property that amounts to saying that $K^{(I)}$ is the vector space freely generated by $I$. Recall that $\iota\colon I \to K^{(I)}$, such that $\iota(i) = e_i$ for every $i \in I$, is an injection from $I$ to $K^{(I)}$.

**Proposition 2.9** *Given any set $I$, for any vector space $F$, and for any function $f\colon I \to F$, there is a unique linear map $\overline{f}\colon K^{(I)} \to F$, such that*

$$f = \overline{f} \circ \iota,$$

*as in the following diagram:*

$$\begin{tikzcd} I \arrow[r, "\iota"] \arrow[dr, "f"'] & K^{(I)} \arrow[d, "\overline{f}"] \\ & F \end{tikzcd}$$

*Proof*. If such a linear map $\overline{f}\colon K^{(I)} \to F$ exists, since $f = \overline{f} \circ \iota$, we must have

$$f(i) = \overline{f}(\iota(i)) = \overline{f}(e_i),$$

for every $i \in I$. However, the family $(e_i)_{i\in I}$ is a basis of $K^{(I)}$, and $(f(i))_{i\in I}$ is a family of vectors in $F$, and by Proposition 2.8, there is a unique linear map $\overline{f}\colon K^{(I)} \to F$ such that $\overline{f}(e_i) = f(i)$ for every $i \in I$, which proves the existence and uniqueness of a linear map $\overline{f}$ such that $f = \overline{f} \circ \iota$. $\square$

The following simple proposition is also useful.

**Proposition 2.10** *Given any two vector spaces $E$ and $F$, with $F$ nontrivial, given any family $(u_i)_{i\in I}$ of vectors in $E$, the following properties hold:*

*(1) The family $(u_i)_{i\in I}$ generates $E$ iff for every family of vectors $(v_i)_{i\in I}$ in $F$, there is at most one linear map $f\colon E \to F$ such that $f(u_i) = v_i$ for all $i \in I$.*

*(2) The family $(u_i)_{i\in I}$ is linearly independent iff for every family of vectors $(v_i)_{i\in I}$ in $F$, there is some linear map $f\colon E \to F$ such that $f(u_i) = v_i$ for all $i \in I$.*

*Proof*. (1) If there is any linear map $f: E \to F$ such that $f(u_i) = v_i$ for all $i \in I$, since $(u_i)_{i \in I}$ generates $E$, every vector $x \in E$ can be written as some linear combination

$$x = \sum_{i \in I} x_i u_i,$$

and by linearity, we must have

$$f(x) = \sum_{i \in I} x_i f(u_i) = \sum_{i \in I} x_i v_i.$$

This shows that $f$ is unique if it exists. Conversely, assume that $(u_i)_{i \in I}$ does not generate $E$. Since $F$ is nontrivial, there is some some vector $y \in F$ such that $y \neq 0$. Since $(u_i)_{i \in I}$ does not generate $E$, there is some vector $w \in E$ that is not in the subspace generated by $(u_i)_{i \in I}$. By Theorem 2.2, there is a linearly independent subfamily $(u_i)_{i \in I_0}$ of $(u_i)_{i \in I}$ generating the same subspace. Since by hypothesis, $w \in E$ is not in the subspace generated by $(u_i)_{i \in I_0}$, by Lemma 2.1 and by Theorem 2.2 again, there is a basis $(e_j)_{j \in I_0 \cup J}$ of $E$, such that $e_i = u_i$, for all $i \in I_0$, and $w = e_{j_0}$, for some $j_0 \in J$. Letting $(v_i)_{i \in I}$ be the family in $F$ such that $v_i = 0$ for all $i \in I$, defining $f: E \to F$ to be the constant linear map with value 0, we have a linear map such that $f(u_i) = 0$ for all $i \in I$. By Proposition 2.8, there is a unique linear map $g: E \to F$ such that $g(w) = y$, and $g(e_j) = 0$, for all $j \in (I_0 \cup J) - \{j_0\}$. By definition of the basis $(e_j)_{j \in I_0 \cup J}$ of $E$, we have, $g(u_i) = 0$ for all $i \in I$, and since $f \neq g$, this contradicts the fact that there is at most one such map.

(2) If the family $(u_i)_{i \in I}$ is linearly independent, then by Theorem 2.2, $(u_i)_{i \in I}$ can be extended to a basis of $E$, and the conclusion follows by Proposition 2.8. Conversely, assume that $(u_i)_{i \in I}$ is linearly dependent. Then, there is some family $(\lambda_i)_{i \in I}$ of scalars (not all zero) such that

$$\sum_{i \in I} \lambda_i u_i = 0.$$

By the assumption, for every $i \in I$, there is some linear map $f_i: E \to F$, such that $f_i(u_i) = y$, and $f_i(u_j) = 0$, for $j \in I - \{i\}$. Then, we would get

$$0 = f_i(\sum_{i \in I} \lambda_i u_i) = \sum_{i \in I} \lambda_i f_i(u_i) = \lambda_i y,$$

and since $y \neq 0$, this implies $\lambda_i = 0$, for every $i \in I$. Thus, $(u_i)_{i \in I}$ is linearly independent. □

Although in this book, we will not have many occasions to use quotient spaces, they are fundamental in algebra. The next section may be omitted until needed.

## 2.6 Quotient Spaces

Let $E$ be a vector space, and let $M$ be any subspace of $E$. The subspace $M$ induces a relation $\equiv_M$ on $E$, defined as follows: For all $u, v \in E$,

$u \equiv_M v$ iff $u - v \in M$.

We have the following simple proposition.

**Proposition 2.11** *Given any vector space $E$ and any subspace $M$ of $E$, the relation $\equiv_M$ is an equivalence relation with the following two congruential properties:*

1. *If $u_1 \equiv_M v_1$ and $u_2 \equiv_M v_2$, then $u_1 + u_2 \equiv_M v_1 + v_2$, and*

2. *if $u \equiv_M v$, then $\lambda u \equiv_M \lambda v$.*

*Proof*. It is obvious that $\equiv_M$ is an equivalence relation. Note that $u_1 \equiv_M v_1$ and $u_2 \equiv_M v_2$ are equivalent to $u_1 - v_1 = w_1$ and $u_2 - v_2 = w_2$, with $w_1, w_2 \in M$, and thus,

$$(u_1 + u_2) - (v_1 + v_2) = w_1 + w_2,$$

and $w_1 + w_2 \in M$, since $M$ is a subspace of $E$. Thus, we have $u_1 + u_2 \equiv_M v_1 + v_2$. If $u - v = w$, with $w \in M$, then

$$\lambda u - \lambda v = \lambda w,$$

and $\lambda w \in M$, since $M$ is a subspace of $E$, and thus $\lambda u \equiv_M \lambda v$. $\square$

Proposition 2.11 shows that we can define addition and multiplication by a scalar on the set $E/M$ of equivalence classes of the equivalence relation $\equiv_M$.

**Definition 2.11** Given any vector space $E$ and any subspace $M$ of $E$, we define the following operations of addition and multiplication by a scalar on the set $E/M$ of equivalence classes of the equivalence relation $\equiv_M$ as follows: for any two equivalence classes $[u], [v] \in E/M$, we have

$$\begin{aligned} [u] + [v] &= [u + v], \\ \lambda[u] &= [\lambda u]. \end{aligned}$$

By Proposition 2.11, the above operations do not depend on the specific choice of representatives in the equivalence classes $[u], [v] \in E/M$. It is also immediate to verify that $E/M$ is a vector space. The function $\pi\colon E \to E/F$, defined such that $\pi(u) = [u]$ for every $u \in E$, is a surjective linear map called the *natural projection of $E$ onto $E/F$*. The vector space $E/M$ is called the *quotient space of $E$ by the subspace $M$*.

Given any linear map $f\colon E \to F$, we know that $\mathrm{Ker}\, f$ is a subspace of $E$, and it is immediately verified that $\mathrm{Im}\, f$ is isomorphic to the quotient space $E/\mathrm{Ker}\, f$.

## 2.7 Matrices

Proposition 2.8 shows that given two vector spaces $E$ and $F$ and a basis $(u_j)_{j\in J}$ of $E$, every linear map $f\colon E \to F$ is uniquely determined by the family $(f(u_j))_{j\in J}$ of the images under $f$ of the vectors in the basis $(u_j)_{j\in J}$. Thus, in particular, taking $F = K^{(J)}$, we get an isomorphism between any vector space $E$ of dimension $|J|$ and $K^{(J)}$. If $J = \{1, \ldots, n\}$, a vector space $E$ of dimension $n$ is isomorphic to the vector space $K^n$. If we also have a basis $(v_i)_{i\in I}$ of $F$, then every vector $f(u_j)$ can be written in a unique way as

$$f(u_j) = \sum_{i\in I} a_{i\,j} v_i,$$

where $j \in J$, for a family $(a_{i\,j})_{i\in I}$. Thus, with respect to the two bases $(u_j)_{j\in J}$ of $E$ and $(v_i)_{i\in I}$ of $F$, the linear map $f$ is completely determined by a possibly infinite "$I \times J$-matrix" $M(f) = (a_{i\,j})_{i\in I,\ j\in J}$.

**Remark:** Note that we intentionally assigned the index set $J$ to the basis $(u_j)_{j\in J}$ of $E$, and the index $I$ to the basis $(v_i)_{i\in I}$ of $F$, so that the rows of the matrix $M(f)$ associated with $f\colon E \to F$ are indexed by $I$, and the columns of the matrix $M(f)$ are indexed by $J$. Obviously, this causes a mildly unpleasant reversal. If we had considered the bases $(u_i)_{i\in I}$ of $E$ and $(v_j)_{j\in J}$ of $F$, we would obtain a $J \times I$-matrix $M(f) = (a_{j\,i})_{j\in J,\ i\in I}$. No matter what we do, there will be a reversal! We decided to stick to the bases $(u_j)_{j\in J}$ of $E$ and $(v_i)_{i\in I}$ of $F$, so that we get an $I \times J$-matrix $M(f)$, knowing that we may occasionally suffer from this decision!

When $I$ and $J$ are finite, and say, when $|I| = m$ and $|J| = n$, the linear map $f$ is determined by the matrix $M(f)$ whose entries in the $j$-th column are the components of the vector $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, that is, the matrix

$$M(f) = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix}$$

whose entry on row $i$ and column $j$ is $a_{i\,j}$ $(1 \leq i \leq m,\ 1 \leq j \leq n)$.

Given vector spaces $E$, $F$, and $G$, and linear maps $f\colon E \to F$ and $g\colon F \to G$, it is easily verified that the composition $g \circ f\colon E \to G$ of $f$ and $g$ is a linear map. A linear map $f\colon E \to F$ is an *isomorphism* iff there is a linear map $g\colon F \to E$, such that $g \circ f = \mathrm{id}_E$, and $f \circ g = \mathrm{id}_F$. It is immediately verified that such a map $g$ is unique. The map $g$ is called the *inverse* of $f$ and it is also denoted by $f^{-1}$.

One can verify that if $f\colon E \to F$ is a bijective linear map, then its inverse $f^{-1}\colon F \to E$ is also a linear map, and thus $f$ is an isomorphism. We leave as an easy exercise to show that if $E$ and $F$ are two vector spaces, $(u_i)_{i\in I}$ is a basis of $E$, and $f\colon E \to F$ is a linear map which is an isomorphism, then the family $(f(u_i))_{i\in I}$ is a basis of $F$.

The set of all linear maps between two vector spaces $E$ and $F$ is denoted as $\mathrm{L}(E;F)$. The set $\mathrm{L}(E;F)$ is a vector space under the operations defined at the end of Section 2.1, namely

$$(f+g)(x) = f(x) + g(x)$$

for all $x \in E$, and

$$(\lambda f)(x) = \lambda f(x)$$

for all $x \in E$. The point worth checking carefully is that $\lambda f$ is indeed a linear map, which uses the commutativity of $*$ in the field $K$. Indeed, we have

$$(\lambda f)(\mu x) = \lambda f(\mu x) = \lambda \mu f(x) = \mu \lambda f(x) = \mu (\lambda f)(x).$$

When $E$ and $F$ have finite dimensions, the vector space $\mathrm{L}(E;F)$ also has finite dimension, as we shall see shortly. When $E = F$, a linear map $f\colon E \to E$ is also called an *endomorphism*. It is also important to note that composition confers to $\mathrm{L}(E;E)$ a ring structure. Indeed, composition is an operation $\circ\colon \mathrm{L}(E;E) \times \mathrm{L}(E;E) \to \mathrm{L}(E;E)$, which is associative and has an identity $\mathrm{id}_E$, and the distributivity properties hold:

$$\begin{aligned}(g_1 + g_2) \circ f &= g_1 \circ f + g_2 \circ f;\\ g \circ (f_1 + f_2) &= g \circ f_1 + g \circ f_2.\end{aligned}$$

The ring $\mathrm{L}(E;E)$ is an example of a noncommutative ring. It is easily seen that the set of bijective linear maps $f\colon E \to E$ is a group under composition. It is called the *general linear group (of $E$)*, and it is denoted by $\mathbf{GL}(E)$, or when $E = K^n$, by $\mathbf{GL}(n, K)$, or even by $\mathbf{GL}(n)$.

We will now show that when $E$ and $F$ have finite dimension, linear maps can be very conveniently represented by matrices, and that composition of linear maps corresponds to matrix multiplication. We will follow rather closely an elegant presentation method due to Emil Artin.

Let $E$ and $F$ be two vector spaces, and assume that $E$ has a finite basis $(u_1, \ldots, u_n)$ and that $F$ has a finite basis $(v_1, \ldots, v_m)$. Recall that we have shown that every vector $x \in E$ can be written in a unique way as

$$x = x_1 u_1 + \cdots + x_n u_n,$$

and similarly every vector $y \in F$ can be written in a unique way as

$$y = y_1 v_1 + \cdots + y_m v_m.$$

Let $f\colon E \to F$ be a linear map between $E$ and $F$. Then, for every $x = x_1 u_1 + \cdots + x_n u_n$ in $E$, by linearity, we have

$$f(x) = x_1 f(u_1) + \cdots + x_n f(u_n).$$

Let
$$f(u_j) = a_{1\,j}v_1 + \cdots + a_{m\,j}v_m,$$
or more concisely,
$$f(u_j) = \sum_{i=1}^{m} a_{i\,j}v_i,$$
for every $j$, $1 \leq j \leq n$. Then, substituting the right-hand side of each $f(u_j)$ into the expression for $f(x)$, we get
$$f(x) = x_1(\sum_{i=1}^{m} a_{i\,1}v_i) + \cdots + x_n(\sum_{i=1}^{m} a_{i\,n}v_i),$$
which, by regrouping terms to obtain a linear combination of the $v_i$, yields
$$f(x) = (\sum_{j=1}^{n} a_{1\,j}x_j)v_1 + \cdots + (\sum_{j=1}^{n} a_{m\,j}x_j)v_m.$$
Thus, letting $f(x) = y = y_1v_1 + \cdots + y_mv_m$, we have
$$y_i = \sum_{j=1}^{n} a_{i\,j}x_j \tag{1}$$
for all $i$, $1 \leq i \leq m$. Let us now consider how the composition of linear maps is expressed in terms of bases.

Let $E$, $F$, and $G$, be three vectors spaces with respective bases $(u_1, \ldots, u_p)$ for $E$, $(v_1, \ldots, v_n)$ for $F$, and $(w_1, \ldots, w_m)$ for $G$. Let $g\colon E \to F$ and $f\colon F \to G$ be linear maps. As explained earlier, $g\colon E \to F$ is determined by the images of the basis vectors $u_j$, and $f\colon F \to G$ is determined by the images of the basis vectors $v_k$. We would like to understand how $f \circ g\colon E \to G$ is determined by the images of the basis vectors $u_j$.

**Remark:** Note that we are considering linear maps $g\colon E \to F$ and $f\colon F \to G$, instead of $f\colon E \to F$ and $g\colon F \to G$, which yields the composition $f \circ g\colon E \to G$ instead of $g \circ f\colon E \to G$. Our perhaps unusual choice is motivated by the fact that if $f$ is represented by a matrix $M(f) = (a_{i\,k})$ and $g$ is represented by a matrix $M(g) = (b_{k\,j})$, then $f \circ g\colon E \to G$ is represented by the product $AB$ of the matrices $A$ and $B$. If we had adopted the other choice where $f\colon E \to F$ and $g\colon F \to G$, then $g \circ f\colon E \to G$ would be represented by the product $BA$. Personally, we find it easier to remember the formula for the entry in row $i$ and column of $j$ of the product of two matrices when this product is written by $AB$, rather than $BA$. Obviously, this is a matter of taste! We will have to live with our perhaps unorthodox choice.

Thus, let
$$f(v_k) = \sum_{i=1}^{m} a_{i\,k}w_i,$$

for every $k$, $1 \leq k \leq n$, and let

$$g(u_j) = \sum_{k=1}^{n} b_{k\,j} v_k,$$

for every $j$, $1 \leq j \leq p$. By previous considerations, for every

$$x = x_1 u_1 + \cdots + x_p u_p,$$

letting $g(x) = y = y_1 v_1 + \cdots + y_n v_n$, we have

$$y_k = \sum_{j=1}^{p} b_{k\,j} x_j \tag{2}$$

for all $k$, $1 \leq k \leq n$, and for every

$$y = y_1 v_1 + \cdots + y_n v_n,$$

letting $f(y) = z = z_1 w_1 + \cdots + z_m w_m$, we have

$$z_i = \sum_{k=1}^{n} a_{i\,k} y_k \tag{3}$$

for all $i$, $1 \leq i \leq m$. Then, if $y = g(x)$ and $z = f(y)$, we have $z = f(g(x))$, and in view of (2) and (3), we have

$$\begin{aligned} z_i &= \sum_{k=1}^{n} a_{i\,k} (\sum_{j=1}^{p} b_{k\,j} x_j) \\ &= \sum_{k=1}^{n} \sum_{j=1}^{p} a_{i\,k} b_{k\,j} x_j \\ &= \sum_{j=1}^{p} \sum_{k=1}^{n} a_{i\,k} b_{k\,j} x_j \\ &= \sum_{j=1}^{p} (\sum_{k=1}^{n} a_{i\,k} b_{k\,j}) x_j. \end{aligned}$$

Thus, defining $c_{i\,j}$ such that

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

for $1 \leq i \leq m$, and $1 \leq j \leq p$, we have

$$z_i = \sum_{j=1}^{p} c_{i\,j} x_j \tag{4}$$

Identity (4) suggests defining a multiplication operation on matrices, and we proceed to do so. We have the following definitions.

**Definition 2.12** Given a field $K$, an *$m \times n$-matrix* is a family $(a_{i\,j})_{1 \le i \le m,\ 1 \le j \le n}$ of scalars in $K$, represented as an array

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \dots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \dots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \dots & a_{m\,n} \end{pmatrix}$$

In the special case where $m = 1$, we have a *row vector*, represented as

$$(a_{1\,1}, \dots, a_{1\,n})$$

and in the special case where $n = 1$, we have a *column vector*, represented as

$$\begin{pmatrix} a_{1\,1} \\ \vdots \\ a_{m\,1} \end{pmatrix}$$

In these last two cases, we usually omit the constant index 1 (first index in case of a row, second index in case of a column). The set of all $m \times n$-matrices is denoted by $\mathrm{M}_{m,n}(K)$ or $\mathrm{M}_{m,n}$. An $n \times n$-matrix is called a *square matrix of dimension $n$.* The set of all square square matrices of dimension $n$ is denoted by $\mathrm{M}_n(K)$, or $\mathrm{M}_n$.

**Remark:** As defined, a matrix $A = (a_{i\,j})_{1 \le i \le m,\ 1 \le j \le n}$ is a *family*, that is, a function from $\{1, 2, \dots, m\} \times \{1, 2, \dots, n\}$ to $K$. As such, there is no reason to assume an ordering on the indices. Thus, the matrix $A$ can be represented in many different ways as an array, by adopting different orders for the rows or the columns. However, it is customary (and usually convenient) to assume the natural ordering on the sets $\{1, 2, \dots, m\}$ and $\{1, 2, \dots, n\}$, and to represent $A$ as an array according to this ordering of the rows and columns.

We also define some operations on matrices as follows.

**Definition 2.13** Given two $m \times n$ matrices $A = (a_{i\,j})$ and $B = (b_{i\,j})$, we define their *sum* $A + B$ as the matrix $C = (c_{i\,j})$ such that $c_{i\,j} = a_{i\,j} + b_{i\,j}$. Given a scalar $\lambda \in K$, we define the matrix $\lambda A$ as the matrix $C = (c_{i\,j})$ such that $c_{i\,j} = \lambda a_{i\,j}$. Given an $m \times n$ matrices $A = (a_{i\,k})$ and an $n \times p$ matrices $B = (b_{k\,j})$, we define their *product* $AB$ as the $m \times p$ matrix $C = (c_{i\,j})$ such that

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

for $1 \le i \le m$, and $1 \le j \le p$. Given an $m \times n$ matrix $A = (a_{i\,j})$, its *transpose* $A^{\top} = (a^{\top}_{j\,i})$, is the $n \times m$-matrix such that $a^{\top}_{j\,i} = a_{i\,j}$, for all $i$, $1 \le i \le m$, and all $j$, $1 \le j \le n$. The transpose of a matrix $A$ is sometimes denoted by $A^t$, or even by ${}^tA$.

In the product $AB = C$ shown below

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix} \begin{pmatrix} b_{1\,1} & b_{1\,2} & \ldots & b_{1\,p} \\ b_{2\,1} & b_{2\,2} & \ldots & b_{2\,p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n\,1} & b_{n\,2} & \ldots & b_{n\,p} \end{pmatrix} = \begin{pmatrix} c_{1\,1} & c_{1\,2} & \ldots & c_{1\,p} \\ c_{2\,1} & c_{2\,2} & \ldots & c_{2\,p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m\,1} & c_{m\,2} & \ldots & c_{m\,p} \end{pmatrix}$$

note that the entry of index $i$ and $j$ of the matrix $AB$ is obtained by multiplying the matrices $A$ and $B$ can be identified with the product of the row matrix corresponding to the $i$-th row of $A$ with the column matrix corresponding to the $j$-column of $B$:

$$(a_{i\,1}, \ldots, a_{i\,n}) \begin{pmatrix} b_{1\,j} \\ \vdots \\ b_{n\,j} \end{pmatrix} = (\sum_{k=1}^{n} a_{i\,k} b_{k\,j})$$

The square matrix $I_n$ of dimension $n$ containing 1 on the diagonal and 0 everywhere else is called the *identity matrix*. It is denoted as

$$\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

Note that the transpose $A^\top$ of a matrix $A$ has the property that the $j$-th row of $A^\top$ is the $j$-th column of $A$. In other words, transposition exchanges the rows and the columns of a matrix. For every square matrix $A$ of dimension $n$, it is immediately verified that $AI_n = I_n A = A$. If a matrix $B$ such that $AB = BA = I_n$ exists, then it is unique, and it is called the *inverse* of $A$. The matrix $B$ is also denoted by $A^{-1}$.

Consider the $m \times n$-matrices $E_{i,j} = (e_{h\,k})$, defined such that $e_{i\,j} = 1$, and $e_{h\,k} = 0$, if $h \neq i$ or $k \neq j$. It is clear that every matrix $A = (a_{i\,j}) \in \mathrm{M}_{m,n}(K)$ can be written in a unique way as

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{i\,j} E_{i,j}.$$

Thus, the family $(E_{i,j})_{1 \le i \le m, 1 \le j \le n}$ is a basis of the vector space $\mathrm{M}_{m,n}(K)$, which has dimension $mn$.

**Remark:** Definition 2.12 and Definition 2.13 also make perfect sense when $K$ is a (commutative) ring rather than a field. In this more general setting, the framework of vector spaces is too narrow, but we can consider structures over a commutative ring $A$ satisfying all the axioms of Definition 2.5. Such structures are called *modules*. The theory of modules is (much) more complicated than that of vector spaces. For example, modules do not always have a basis, and other properties holding for vector spaces usually fail for modules. When a module has a basis, it is called a *free module*. For example, when $A$ is a commutative ring, the structure $A^n$ is a module such that the vectors $e_i$, with $(e_i)_i = 1$ and $(e_i)_j = 0$ for $j \neq i$, form a basis of $A^n$. Many properties of vector spaces still hold for $A^n$. Thus, $A^n$ is a free module. As another example, when $A$ is a commutative ring, $\mathrm{M}_{m,n}(A)$ is a free module with basis $(E_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$. Polynomials over a commutative ring also form a free module of infinite dimension.

Square matrices provide a natural example of a noncommutative ring with zero divisors.

**Example 2.12** For example, letting $A, B$ be the $2 \times 2$-matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

then

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

and

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

We now formalize the representation of linear maps by matrices.

**Definition 2.14** Let $E$ and $F$ be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis for $E$, and $(v_1, \ldots, v_m)$ be a basis for $F$. Each vector $x \in E$ expressed in the basis $(u_1, \ldots, u_n)$ as $x = x_1 u_1 + \cdots + x_n u_n$ is represented by the column matrix

$$M(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and similarly for each vector $y \in F$ expressed in the basis $(v_1, \ldots, v_m)$.

Every linear map $f \colon E \to F$ is represented by the matrix $M(f) = (a_{i\,j})$, where $a_{i\,j}$ is the $i$-th component of the vector $f(u_j)$ over the basis $(v_1, \ldots, v_m)$, i.e., where

$$f(u_j) = \sum_{i=1}^{m} a_{i\,j} v_i,$$

for every $j$, $1 \leq j \leq n$. Explicitly, we have

$$M(f) = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \ldots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \ldots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m\,1} & a_{m\,2} & \ldots & a_{m\,n} \end{pmatrix}$$

The matrix $M(f)$ associated with the linear map $f\colon E \to F$ is called the *matrix of $f$ with respect to the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$*. When $E = F$ and the basis $(v_1, \ldots, v_m)$ is identical to the basis $(u_1, \ldots, u_n)$ of $E$, the matrix $M(f)$ associated with $f\colon E \to E$ (as above) is called the *matrix of $f$ with respect to the base $(u_1, \ldots, u_n)$*.

**Remark:** As in the remark after Definition 2.12, there is no reason to assume that the vectors in the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ are ordered in any particular way. However, it is often convenient to assume the natural ordering. When this is so, authors sometimes refer to the matrix $M(f)$ as the matrix of $f$ with respect to the *ordered bases* $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$.

Then, given a linear map $f\colon E \to F$ represented by the matrix $M(f) = (a_{i\,j})$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, by equations (1) and the definition of matrix multiplication, the equation $y = f(x)$ correspond to the matrix equation $M(y) = M(f)M(x)$, that is,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{1\,1} & \ldots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{m\,1} & \ldots & a_{m\,n} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

The function that associates to a linear map $f\colon E \to F$ the matrix $M(f)$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ has the property that matrix multiplication corresponds to composition of linear maps. The following proposition states the main properties of the mapping $f \mapsto M(f)$ between $\mathrm{L}(E; F)$ and $\mathrm{M}_{m,n}$. In short, it is an isomorphism of vector spaces.

**Proposition 2.12** *Given three vector spaces $E$, $F$, $G$, with respective bases $(u_1, \ldots, u_p)$, $(v_1, \ldots, v_n)$, and $(w_1, \ldots, w_m)$, the mapping $M\colon \mathrm{L}(E; F) \to \mathrm{M}_{n,p}$ that associates the matrix $M(g)$ to a linear map $g\colon E \to F$ satisfies the following properties for all $x \in E$, all $g, h\colon E \to F$, and all $f\colon F \to G$:*

$$\begin{aligned} M(g(x)) &= M(g)M(x) \\ M(g + h) &= M(g) + M(h) \\ M(\lambda g) &= \lambda M(g) \\ M(f \circ g) &= M(f)M(g). \end{aligned}$$

*Thus, $M\colon \mathrm{L}(E; F) \to \mathrm{M}_{n,p}$ is an isomorphism of vector spaces, and when $p = n$ and the basis $(v_1, \ldots, v_n)$ is identical to the basis $(u_1, \ldots, u_p)$, $M\colon \mathrm{L}(E; E) \to \mathrm{M}_n$ is an isomorphism of rings.*

*Proof*. That $M(g(x)) = M(g)M(x)$ was shown just before stating the proposition, using identity (1). The identities $M(g + h) = M(g) + M(h)$ and $M(\lambda g) = \lambda M(g)$ are straightforward, and $M(f \circ g) = M(f)M(g)$ follows from (4) and the definition of matrix multiplication.

The mapping $M\colon \mathrm{L}(E;F) \to \mathrm{M}_{n,p}$ is clearly injective, and since every matrix defines a linear map, it is also surjective, and thus bijective. In view of the above identities, it is an isomorphism (and similarly for $M\colon \mathrm{L}(E;E) \to \mathrm{M}_n$). $\square$

In view of Proposition 2.12, it seems preferable to represent vectors from a vector space of finite dimension as column vectors rather than row vectors. Thus, from now on, we will denote vectors of $\mathbb{R}^n$ (or more generally, of $K^n$) as columm vectors.

It is important to observe that the isomorphism $M\colon \mathrm{L}(E;F) \to \mathrm{M}_{n,p}$ given by Proposition 2.12 depends on the choice of the bases $(u_1, \ldots, u_p)$ and $(v_1, \ldots, v_n)$, and similarly for the isomorphism $M\colon \mathrm{L}(E;E) \to \mathrm{M}_n$, which depends on the choice of the basis $(u_1, \ldots, u_n)$. Thus, it would be useful to know how a change of basis affects the representation of a linear map $f\colon E \to F$ as a matrix. The following simple proposition is needed.

**Proposition 2.13** *Let $E$ be a vector space, and let $(u_1, \ldots, u_n)$ be a basis of $E$. For every family $(v_1, \ldots, v_n)$, let $P = (a_{i\,j})$ be the matrix defined such that $v_j = \sum_{i=1}^n a_{i\,j} u_i$. The matrix $P$ is invertible iff $(v_1, \ldots, v_n)$ is a basis of $E$.*

*Proof.* Note that we have $P = M(f)$, the matrix associated with the unique linear map $f\colon E \to E$ such that $f(u_i) = v_i$. By Proposition 2.8, $f$ is bijective iff $(v_1, \ldots, v_n)$ is a basis of $E$. Furthermore, it is obvious that the identity matrix $I_n$ is the matrix associated with the identity $\mathrm{id}\colon E \to E$ w.r.t. any basis. If $f$ is an isomorphism, then $f \circ f^{-1} = f^{-1} \circ f = \mathrm{id}$, and by Proposition 2.12, we get $M(f)M(f^{-1}) = M(f^{-1})M(f) = I_n$, showing that $P$ is invertible and that $M(f^{-1}) = P^{-1}$. $\square$

Proposition 2.13 suggests the following definition.

**Definition 2.15** Given a vector space $E$ of dimension $n$, for any two bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ of $E$, let $P = (a_{i\,j})$ be the invertible matrix defined such that

$$v_j = \sum_{i=1}^n a_{i\,j} u_i,$$

which is also the matrix of the identity $\mathrm{id}\colon E \to E$ with respect to the bases $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_n)$, *in that order* (indeed, we express each $\mathrm{id}(v_j) = v_j$ over the basis $(u_1, \ldots, u_n)$). The matrix $P$ is called the *change of basis matrix from* $(u_1, \ldots, u_n)$ *to* $(v_1, \ldots, v_n)$.

Clearly, the change of basis matrix from $(v_1, \ldots, v_n)$ to $(u_1, \ldots, u_n)$ is $P^{-1}$. Since $P = (a_{i,j})$ is the matrix of the identity $\mathrm{id}\colon E \to E$ with respect to the bases $(v_1, \ldots, v_n)$ and $(u_1, \ldots, u_n)$, given any vector $x \in E$, if $x = x_1 u_1 + \cdots + x_n u_n$ over the basis $(u_1, \ldots, u_n)$ and $x = x'_1 v_1 + \cdots + x'_n v_n$ over the basis $(v_1, \ldots, v_n)$, from Proposition 2.12, we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{1\,1} & \ldots & a_{1\,n} \\ \vdots & \ddots & \vdots \\ a_{n\,1} & \ldots & a_{n\,n} \end{pmatrix} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

showing that the *old* coordinates $(x_i)$ of $x$ (over $(u_1,\ldots,u_n)$) are expressed in terms of the *new* coordinates $(x'_i)$ of $x$ (over $(v_1,\ldots,v_n)$). Since the matrix $P$ expresses the *new* basis $(v_1,\ldots,v_n)$ in terms of the *old* basis $(u_1,\ldots,u_n)$, we observe that the coordinates $(x_i)$ of a vector $x$ vary in the *opposite direction* of the change of basis. For this reason, vectors are sometimes said to be *contravariant*. However, this expression does not make sense! Indeed, a vector in an intrinsic quantity that does not depend on a specific basis. What makes sense is that the *coordinates* of a vector vary in a contravariant fashion.

The effect of a change of bases on the representation of a linear map is described in the following proposition.

**Proposition 2.14** *Let $E$ and $F$ be vector spaces, let $(u_1,\ldots,u_n)$ and $(u'_1,\ldots,u'_n)$ be two bases of $E$, and let $(v_1,\ldots,v_m)$ and $(v'_1,\ldots,v'_m)$ be two bases of $F$. Let $P$ be the change of basis matrix from $(u_1,\ldots,u_n)$ to $(u'_1,\ldots,u'_n)$, and let $Q$ be the change of basis matrix from $(v_1,\ldots,v_m)$ to $(v'_1,\ldots,v'_m)$. For any linear map $f\colon E\to F$, let $M(f)$ be the matrix associated to $f$ w.r.t. the bases $(u_1,\ldots,u_n)$ and $(v_1,\ldots,v_m)$, and let $M'(f)$ be the matrix associated to $f$ w.r.t. the bases $(u'_1,\ldots,u'_n)$ and $(v'_1,\ldots,v'_m)$. We have*

$$M'(f) = Q^{-1}M(f)P.$$

*Proof*. Since $f\colon E\to F$ can be written as $f = \mathrm{id}_F\circ f\circ \mathrm{id}_E$, since $P$ is the matrix of $\mathrm{id}_E$ w.r.t. the bases $(u'_1,\ldots,u'_n)$ and $(u_1,\ldots,u_n)$, and $Q^{-1}$ is the matrix of $\mathrm{id}_F$ w.r.t. the bases $(v_1,\ldots,v_m)$ and $(v'_1,\ldots,v'_m)$, by Proposition 2.12, we have $M'(f) = Q^{-1}M(f)P$.

As a corollary, we get the following result.

**Corollary 2.15** *Let $E$ be a vector space, and let $(u_1,\ldots,u_n)$ and $(u'_1,\ldots,u'_n)$ be two bases of $E$. Let $P$ be the change of basis matrix from $(u_1,\ldots,u_n)$ to $(u'_1,\ldots,u'_n)$. For any linear map $f\colon E\to E$, let $M(f)$ be the matrix associated to $f$ w.r.t. the basis $(u_1,\ldots,u_n)$, and let $M'(f)$ be the matrix associated to $f$ w.r.t. the basis $(u'_1,\ldots,u'_n)$. We have*

$$M'(f) = P^{-1}M(f)P.$$

## 2.8 Direct Sums

Before considering linear forms and hyperplanes, we define the notion of direct sum and prove some simple proositions. There is a subtle point, which is that if we attempt to define the direct sum $E\oplus F$ of two vector spaces using the cartesian product $E\times F$, we don't quite get the right notion because elements of $E\times F$ are ordered pairs, but we want $E\oplus F = F\oplus E$. Thus, we want to think of the elements of $E\oplus F$ as unordrered pairs of elements. It is possible to do so by considering the direct sum of a *family* $(E_i)_{i\in\{1,2\}}$, and more generally of a family $(E_i)_{i\in I}$. For simplicity, we begin by considering the case where $I = \{1,2\}$.

**Definition 2.16** Given a family $(E_i)_{i\in\{1,2\}}$ of two vector spaces, we define the *(external) direct sum* $E_1 \oplus E_2$ of the family $(E_i)_{i\in\{1,2\}}$ as the set

$$E_1 \oplus E_2 = \{\{\langle 1, u\rangle, \langle 2, v\rangle\} \mid u \in E_1,\ v \in E_2\},$$

with addition

$$\{\langle 1, u_1\rangle, \langle 2, v_1\rangle\} + \{\langle 1, u_2\rangle, \langle 2, v_2\rangle\} = \{\langle 1, u_1 + u_2\rangle, \langle 2, v_1 + v_2\rangle\},$$

and scalar multiplication

$$\lambda\{\langle 1, u\rangle, \langle 2, v\rangle\} = \{\langle 1, \lambda u\rangle, \langle 2, \lambda v\rangle\}.$$

We define the *injections* $in_1 \colon E_1 \to E_1 \oplus E_2$ and $in_2 \colon E_2 \to E_1 \oplus E_2$ as the linear maps defined such that,

$$in_1(u) = \{\langle 1, u\rangle, \langle 2, 0\rangle\},$$

and

$$in_2(v) = \{\langle 1, 0\rangle, \langle 2, v\rangle\}.$$

Note that

$$E_2 \oplus E_1 = \{\{\langle 2, v\rangle, \langle 1, u\rangle\} \mid v \in E_2,\ u \in E_1\} = E_1 \oplus E_2.$$

Thus, every member $\{\langle 1, u\rangle, \langle 2, v\rangle\}$ of $E_1 \oplus E_2$ can be viewed as an *unordered pair* consisting of the two vectors $u$ and $v$, tagged with the index 1 and 2, respectively.

**Remark:** In fact, $E_1 \oplus E_2$ is just the product $\prod_{i\in\{1,2\}} E_i$ of the family $(E_i)_{i\in\{1,2\}}$.

This is not to be confused with the cartesian product $E_1 \times E_2$. The vector space $E_1 \times E_2$ is the set of all ordered pairs $\langle u, v\rangle$, where $u \in E_1$, and $v \in E_2$, with addition and multiplication by a scalar defined such that

$$\langle u_1, v_1\rangle + \langle u_2, v_2\rangle = \langle u_1 + u_2, v_1 + v_2\rangle,$$
$$\lambda\langle u, v\rangle = \langle \lambda u, \lambda v\rangle.$$

There is a bijection between $\prod_{i\in\{1,2\}} E_i$ and $E_1 \times E_2$, but as we just saw, elements of $\prod_{i\in\{1,2\}} E_i$ are certain sets.

Of course, we can define $E_1 \times \cdots \times E_n$ for any number of vector spaces, and when $E_1 = \ldots = E_n$, we denote this product by $E^n$.

The following property holds.

**Proposition 2.16** *Given two vector spaces $E_1$ and $E_2$, $E_1 \oplus E_2$ is a vector space. For every pair of linear maps $f\colon E_1 \to G$ and $g\colon E_2 \to G$, there is a unique linear map $f + g\colon E_1 \oplus E_2 \to G$, such that $(f + g) \circ in_1 = f$, and $(f + g) \circ in_2 = g$, as in the following diagram:*

$$\begin{tikzcd}
E_1 \arrow[d, "in_1"'] \arrow[dr, "f"] & \\
E_1 \oplus E_2 \arrow[r, "f+g"] & G \\
E_2 \arrow[u, "in_2"] \arrow[ur, "g"'] &
\end{tikzcd}$$

*Proof*. Define

$$(f + g)(\{\langle 1, u\rangle, \langle 2, v\rangle\}) = f(u) + g(v),$$

for every $u \in E_1$ and $v \in E_2$. It is immediately verified that $f + g$ is the unique linear map with the required properties. $\square$

We already noted that $E_1 \oplus E_2$ is in bijection with $E_1 \times E_2$. If we define the *projections* $\pi_1\colon E_1 \oplus E_2 \to E_1$ and $\pi_2\colon E_1 \oplus E_2 \to E_2$, such that

$$\pi_1(\{\langle 1, u\rangle, \langle 2, v\rangle\}) = u,$$

and

$$\pi_2(\{\langle 1, u\rangle, \langle 2, v\rangle\}) = v,$$

we have the following proposition.

**Proposition 2.17** *Given two vector spaces $E_1$ and $E_2$, for every pair of linear maps $f\colon D \to E_1$ and $g\colon D \to E_2$, there is a unique linear map $f \times g\colon D \to E_1 \oplus E_2$, such that $\pi_1 \circ (f \times g) = f$, and $\pi_2 \circ (f \times g) = g$, as in the following diagram:*

$$\begin{tikzcd}
 & E_1 \\
D \arrow[ur, "f"] \arrow[r, "f\times g"] \arrow[dr, "g"'] & E_1 \oplus E_2 \arrow[u, "\pi_1"'] \arrow[d, "\pi_2"] \\
 & E_2
\end{tikzcd}$$

*Proof*. Define

$$(f \times g)(w) = \{\langle 1, f(w)\rangle, \langle 2, g(w)\rangle\},$$

for every $w \in D$. It is immediately verified that $f \times g$ is the unique linear map with the required properties. $\square$

**Remark:** It is a peculiarity of linear algebra that direct sums and products of finite families are isomorphic. However, this is no longer true for products and sums of infinite families.

When $U, V$ are subspaces of a vector space $E$, letting $i_1: U \to E$ and $i_2: V \to E$ be the inclusion maps, if $U \oplus V$ is isomomorphic to $E$ under the map $i_1 + i_2$ given by Proposition 2.16, we say that $E$ *is a direct (internal) sum* of $U$ and $V$, and we write $E = U \oplus V$ (with a slight abuse of notation, since $E$ and $U \oplus V$ are only isomorphic). It is also convenient to define the sum $U + V$ of $U$ and $V$.

**Definition 2.17** Given a vector space $E$, let $U, V$ be any subspaces of $E$. We define the *sum* $U + V$ of $U$ and $V$ as the set

$$U + V = \{w \in E \mid w = u + v, \text{ for some } u \in U \text{ and some } v \in V\}.$$

We say that $E$ *is the (internal) direct sum* of $U$ and $V$, denoted by $E = U \oplus V$,[3] if for every $x \in E$, there exist unique vectors $u \in U$ and $v \in V$, such that $x = u + v$.

It is immediately verified that $U + V$ is the least subspace of $E$ containing $U$ and $V$. Note that by definition, $U + V = V + U$, and $U \oplus V = V \oplus U$. The following two simple propositions holds.

**Proposition 2.18** *Let $E$ be a vector space. The following properties are equivalent:*

1. $E = U \oplus V$.
2. $E = U + V$ *and* $U \cap V = 0$.

*Proof*. First, assume that $E$ is the direct sum of $U$ and $V$. If $x \in U \cap V$ and $x \neq 0$, since $x$ can be written both as $x+0$ and $0+x$, we have a contradiction. Thus $U \cap V = 0$. Conversely, assume that $x = u + v$ and $x = u' + v'$, where $u, u' \in U$ and $v, v' \in V$. Then,

$$v' - v = u - u',$$

where $v' - v \in V$ and $u - u' \in U$. Since $U \cap V = 0$, we must have $u' = u$ and $v' = v$, and thus $E = U \oplus V$. $\square$

**Proposition 2.19** *Let $E$ be a vector space, and assume that $E = U \oplus V$. Then,*

$$\dim(E) = \dim(U) + \dim(V).$$

*Proof*. Let $(u_i)_{i \in I}$ be a basis of $U$ and let $(v_j)_{j \in J}$ be a basis of $V$, where $I$ and $J$ are disjoint. Clearly, $(u_i)_{i \in I} \cup (v_j)_{j \in J}$ is a basis of $E$. $\square$

We now give the definition of a direct sum for any arbitrary nonempty index set $I$. First, let us recall the notion of the product of a family $(E_i)_{i \in I}$. Given a family of sets $(E_i)_{i \in I}$, its product $\prod_{i \in I} E_i$, is the set of all functions $f: I \to \bigcup_{i \in I} E_i$, such that, $f(i) \in E_i$, for every $i \in I$. It is one of the many versions of the axiom of choice, that, if $E_i \neq \emptyset$ for every $i \in I$, then $\prod_{i \in I} E_i \neq \emptyset$. A member $f \in \prod_{i \in I} E_i$, is often denoted as $(f_i)_{i \in I}$. For every $i \in I$, we have the *projection* $\pi_i: \prod_{i \in I} E_i \to E_i$, defined such that, $\pi_i((f_i)_{i \in I}) = f_i$. We now define direct sums.

[3] Again, with a slight abuse of notation!

**Definition 2.18** Let $I$ be any nonempty set, and let $(E_i)_{i\in I}$ be a family of vector spaces. The *(external) direct sum* $\bigoplus_{i\in I} E_i$ of the family $(E_i)_{i\in I}$ is defined as follows:

$\bigoplus_{i\in I} E_i$ consists of all $f \in \prod_{i\in I} E_i$, which have finite support, and addition and multiplication by a scalar are defined as follows:

$$\begin{aligned}(f_i)_{i\in I} + (g_i)_{i\in I} &= (f_i + g_i)_{i\in I},\\ \lambda(f_i)_{i\in I} &= (\lambda f_i)_{i\in I}.\end{aligned}$$

We also have *injection maps* $in_i\colon E_i \to \bigoplus_{i\in I} E_i$, defined such that, $in_i(x) = (f_i)_{i\in I}$, where $f_i = x$, and $f_j = 0$, for all $j \in (I - \{i\})$.

The following proposition is an obvious generalization of Proposition 2.16.

**Proposition 2.20** *Let $I$ be any nonempty set, let $(E_i)_{i\in I}$ be a family of vector spaces, and let $G$ be any vector space. The direct sum $\bigoplus_{i\in I} E_i$ is a vector space, and for every family $(h_i)_{i\in I}$ of linear maps $h_i\colon E_i \to G$, there is a unique linear map*

$$\Big(\sum_{i\in I} h_i\Big)\colon \bigoplus_{i\in I} E_i \to G,$$

*such that, $(\sum_{i\in I} h_i) \circ in_i = h_i$, for every $i \in I$.*

**Remark:** When $E_i = E$, for all $i \in I$, we denote $\bigoplus_{i\in I} E_i$ by $E^{(I)}$. In particular, when $E_i = K$, for all $i \in I$, we find the vector space $K^{(I)}$ of Definition 2.9.

We also have the following basic proposition about injective or surjective linear maps.

**Proposition 2.21** *Let $E$ and $F$ be vector spaces, and let $f\colon E \to F$ be a linear map. If $f\colon E \to F$ is injective, then, there is a surjective linear map $r\colon F \to E$ called a retraction, such that $r \circ f = \mathrm{id}_E$. If $f\colon E \to F$ is surjective, then, there is an injective linear map $s\colon F \to E$ called a section, such that $f \circ s = \mathrm{id}_F$.*

*Proof*. Let $(u_i)_{i\in I}$ be a basis of $E$. Since $f\colon E \to F$ is an injective linear map, by Proposition 2.8, $(f(u_i))_{i\in I}$ is linearly independent in $F$. By Theorem 2.2, there is a basis $(v_j)_{j\in J}$ of $F$, where $I \subseteq J$, and where $v_i = f(u_i)$, for all $i \in I$. By Proposition 2.8, a linear map $r\colon F \to E$ can be defined such that $r(v_i) = u_i$, for all $i \in I$, and $r(v_j) = w$ for all $j \in (J - I)$, where $w$ is any given vector in $E$, say $w = 0$. Since $r(f(u_i)) = u_i$ for all $i \in I$, by Proposition 2.8, we have $r \circ f = \mathrm{id}_E$.

Now, assume that $f\colon E \to F$ is surjective. Let $(v_j)_{j\in J}$ be a basis of $F$. Since $f\colon E \to F$ is surjective, for every $v_j \in F$, there is some $u_j \in E$ such that $f(u_j) = v_j$. Since $(v_j)_{j\in J}$ is a basis of $F$, by Proposition 2.8, there is a unique linear map $s\colon F \to E$ such that $s(v_j) = u_j$. Also, since $f(s(v_j)) = v_j$, by Proposition 2.8 (again), we must have $f \circ s = \mathrm{id}_F$. □

The converse of Proposition 2.21 is obvious. We now have the following fundamental Proposition.

**Proposition 2.22** *Let $E$, $F$ and $G$, be three vector spaces, $f: E \to F$ an injective linear map, $g: F \to G$ a surjective linear map, and assume that* $\mathrm{Im}\, f = \mathrm{Ker}\, g$*. Then, the following properties hold. (a) For any section $s: G \to F$ of $g$,* $\mathrm{Ker}\, g \oplus \mathrm{Im}\, s$ *is isomorphic to $F$, and the linear map $f + s: E \oplus G \to F$ is an isomorphism.*[4]

*(b) For any retraction $r: F \to E$ of $f$,* $\mathrm{Im}\, f \oplus \mathrm{Ker}\, r$ *is isomorphic to $F$.*[5]

*Proof*. (a) Since $s: G \to F$ is a section of $g$, we have $g \circ s = \mathrm{id}_G$, and for every $u \in F$,

$$g(u - s(g(u))) = g(u) - g(s(g(u))) = g(u) - g(u) = 0.$$

Thus, $u - s(g(u)) \in \mathrm{Ker}\, g$, and we have $F = \mathrm{Ker}\, g + \mathrm{Im}\, s$. On the other hand, if $u \in \mathrm{Ker}\, g \cap \mathrm{Im}\, s$, then $u = s(v)$ for some $v \in G$ because $u \in \mathrm{Im}\, s$, $g(u) = 0$ because $u \in \mathrm{Ker}\, g$, and so,

$$g(u) = g(s(v)) = v = 0,$$

because $g \circ s = \mathrm{id}_G$, which shows that $u = s(v) = 0$. Thus, $F = \mathrm{Ker}\, g \oplus \mathrm{Im}\, s$, and since by assumption, $\mathrm{Im}\, f = \mathrm{Ker}\, g$, we have $F = \mathrm{Im}\, f \oplus \mathrm{Im}\, s$. But then, since $f$ and $s$ are injective, $f + s: E \oplus G \to F$ is an isomorphism. The proof of (b) is very similar. $\square$

Given a sequence of linear maps $E \xrightarrow{f} F \xrightarrow{g} G$, when $\mathrm{Im}\, f = \mathrm{Ker}\, g$, we say that the sequence $E \xrightarrow{f} F \xrightarrow{g} G$ is *exact at $F$*. If in addition to being exact at $F$, $f$ is injective and $g$ is surjective, we say that we have a *short exact sequence*, and this is denoted as

$$0 \longrightarrow E \xrightarrow{f} F \xrightarrow{g} G \longrightarrow 0.$$

The property of a short exact sequence given by Proposition 2.22 is often described by saying that $0 \longrightarrow E \xrightarrow{f} F \xrightarrow{g} G \longrightarrow 0$ is a (short) *split exact sequence*.

As a corollary of Proposition 2.22, we have the following result.

**Corollary 2.23** *Let $E$ and $F$ be vector spaces, and let $f: E \to F$ be a linear map. Then, $E$ is isomorphic to* $\mathrm{Ker}\, f \oplus \mathrm{Im}\, f$*, and thus,*

$$\dim(E) = \dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f).$$

*Proof*. Consider

$$\mathrm{Ker}\, f \xrightarrow{i} E \xrightarrow{f'} \mathrm{Im}\, f,$$

where $\mathrm{Ker}\, f \xrightarrow{i} E$ is the inclusion map, and $E \xrightarrow{f'} \mathrm{Im}\, f$ is the surjection associated with $E \xrightarrow{f} F$. Then, we apply Proposition 2.22 to any section $\mathrm{Im}\, f \xrightarrow{s} E$ of $f'$ to get an isomorphism between $E$ and $\mathrm{Ker}\, f \oplus \mathrm{Im}\, f$, and Proposition 2.19, to get $\dim(E) = \dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f)$. $\square$

The following Proposition will also be useful.

[4]The existence of a section $s: G \to F$ of $g$ follows from Proposition 2.21.

[5]The existence of a retraction $r: F \to E$ of $f$ follows from Proposition 2.21.

**Proposition 2.24** *Let $E$ be a vector space. If $E = U \oplus V$ and $E = U \oplus W$, then there is an isomorphism $f: V \to W$ between $V$ and $W$.*

*Proof.* Let $R$ be the relation between $V$ and $W$, defined such that

$$\langle v, w \rangle \in R \quad \text{iff} \quad w - v \in U.$$

We claim that $R$ is a functional relation that defines a linear isomorphism $f: V \to W$ between $V$ and $W$, where $f(v) = w$ iff $\langle v, w \rangle \in R$ ($R$ is the graph of $f$). If $w - v \in U$ and $w' - v \in U$, then $w' - w \in U$, and since $U \oplus W$ is a direct sum, $U \cap W = 0$, and thus $w' - w = 0$, that is $w' = w$. Thus, $R$ is functional. Similarly, if $w - v \in U$ and $w - v' \in U$, then $v' - v \in U$, and since $U \oplus V$ is a direct sum, $U \cap V = 0$, and $v' = v$. Thus, $f$ is injective. Since $E = U \oplus V$, for every $w \in W$, there exists a unique pair $\langle u, v \rangle \in U \times V$, such that $w = u + v$. Then, $w - v \in U$, and $f$ is surjective. We also need to verify that $f$ is linear. If

$$w - v = u$$

and

$$w' - v' = u',$$

where $u, u' \in U$, then, we have

$$(w + w') - (v + v') = (u + u'),$$

where $u + u' \in U$. Similarly, if

$$w - v = u$$

where $u \in U$, then we have

$$\lambda w - \lambda v = \lambda u,$$

where $\lambda u \in U$. Thus, $f$ is linear. $\square$

Given a vector space $E$ and any subspace $U$ of $E$, Proposition 2.24 shows that the dimension of any subspace $V$ such that $E = U \oplus V$ depends only on $U$. We call $\dim(V)$ the *codimension* of $U$, and we denote it by $\mathrm{codim}(U)$. A subspace $U$ of codimension 1 is called a *hyperplane*.

The notion of rank of a linear map or of a matrix is an important one, both theoretically and practically, since it is the key to the solvabilty of linear equations.

**Definition 2.19** Given two vector spaces $E$ and $F$ and a linear map $f: E \to F$, the *rank* $\mathrm{rk}(f)$ of $f$ is the dimension $\dim(\mathrm{Im}\, f)$ of the image subspace $\mathrm{Im}\, f$ of $F$.

We have the following simple proposition.

**Proposition 2.25** *Given a linear map $f: E \to F$, the following properties hold:*

*(i)* $\mathrm{rk}(f) = \mathrm{codim}(\mathrm{Ker}\, f)$.

*(ii)* $\mathrm{rk}(f) + \dim(\mathrm{Ker}\, f) = \dim(E)$.

*(iii)* $\mathrm{rk}(f) \leq \min(\dim(E), \dim(F))$.

*Proof*. Since by Proposition 2.23, $\dim(E) = \dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f)$, and by definition, $\mathrm{rk}(f) = \dim(\mathrm{Im}\, f)$, we have $\mathrm{rk}(f) = \mathrm{codim}(\mathrm{Ker}\, f)$. Since $\mathrm{rk}(f) = \dim(\mathrm{Im}\, f)$, (ii) follows from $\dim(E) = \dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f)$. As for (iii), since $\mathrm{Im}\, f$ is a subspace of $F$, we have $\mathrm{rk}(f) \leq \dim(F)$, and since $\mathrm{rk}(f) + \dim(\mathrm{Ker}\, f) = \dim(E)$, we have $\mathrm{rk}(f) \leq \dim(E)$. □

The rank of a matrix is defined as follows.

**Definition 2.20** Given a $m \times n$-matrix $A = (a_{i\,j})$ over the field $K$, the *rank* $\mathrm{rk}(A)$ of the matrix $A$ is the maximum number of linearly independent columns of $A$ (viewed as vectors in $K^m$).

In view of Proposition 2.3, the rank of a matrix $A$ is the dimension of the subspace of $K^m$ generated by the columns of $A$. Let $E$ and $F$ be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis of $E$, and $(v_1, \ldots, v_m)$ a basis of $F$. Let $f\colon E \to F$ be a linear map, and let $M(f)$ be its matrix w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$. Since the rank $\mathrm{rk}(f)$ of $f$ is the dimension of $\mathrm{Im}\, f$, which is generated by $(f(u_1), \ldots, f(u_n))$, the rank of $f$ is the maximum number of linearly independent vectors in $(f(u_1), \ldots, f(u_n))$, which is equal to the number of linearly independent columns of $M(f)$, since $F$ and $K^m$ are isomorphic. Thus, we have $\mathrm{rk}(f) = \mathrm{rk}(M(f))$, for every matrix representing $f$.

We will see later, using duality, that the rank of a matrix $A$ is also equal to the maximal number of linearly independent rows of $A$.

If $U$ is a hyperplane, then $E = U \oplus V$ for some subspace $V$ of dimension 1. However, a subspace $V$ of dimension 1 is generated by any nonzero vector $v \in V$, and thus we denote $V$ by $Kv$, and we write $E = U \oplus Kv$. Clearly, $v \notin U$. Conversely, let $x \in E$ be a vector such that $x \notin U$ (and thus, $x \neq 0$). We claim that $E = U \oplus Kx$. Indeed, since $U$ is a hyperplane, we have $E = U \oplus Kv$ for some $v \notin U$ (with $v \neq 0$). Then, $x \in E$ can be written in a unique way as $x = u + \lambda v$, where $u \in U$, and since $x \notin U$, we must have $\lambda \neq 0$, and thus, $v = -\lambda^{-1}u + \lambda^{-1}x$. Since $E = U \oplus Kv$, this shows that $E = U + Kx$. Since $x \notin U$, we have $U \cap Kx = 0$, and thus $E = U \oplus Kx$. This argument shows that a hyperplane is a maximal proper subspace $H$ of $E$.

In the next section, we shall see that hyperplanes are precisely the Kernels of nonnull linear maps $f\colon E \to K$, called linear forms.

## 2.9 The Dual Space $E^*$ and Linear Forms

We already observed that the field $K$ itself is a vector space (over itself). The vector space $\mathrm{L}(E; K)$ of linear maps $f: E \to K$, the linear forms, plays a particular role. We take a quick look at the connection between $E$ and $\mathrm{L}(E; K)$, its *dual space*.

**Definition 2.21** Given a vector space $E$, the vector space $\mathrm{L}(E; K)$ of linear maps $f: E \to K$ is called the *dual space (or dual)* of $E$. The space $\mathrm{L}(E; K)$ is also denoted by $E^*$, and the linear maps in $E^*$ are called *the linear forms*, or *covectors*. The dual space $E^{**}$ of the space $E^*$ is called the *bidual* of $E$.

As a matter of notation, linear forms $f: E \to K$ will also be denoted by starred symbol, such as $u^*$, $x^*$, etc.

Given a linear form $u^* \in E^*$ and a vector $v \in E$, the result $u^*(v)$ of applying $u^*$ to $v$ is also denoted by $\langle u^*, v\rangle$. This defines a binary operation $\langle -, -\rangle: E^* \times E \to K$ satisfying the following properties:

$$\begin{aligned}
\langle u_1^* + u_2^*, v\rangle &= \langle u_1^*, v\rangle + \langle u_2^*, v\rangle \\
\langle u^*, v_1 + v_2\rangle &= \langle u^*, v_1\rangle + \langle u^*, v_2\rangle \\
\langle \lambda u^*, v\rangle &= \lambda\langle u^*, v\rangle \\
\langle u^*, \lambda v\rangle &= \lambda\langle u^*, v\rangle.
\end{aligned}$$

The above identities mean that $\langle -, -\rangle$ is a *bilinear map*. In view of the above identities, given any fixed vector $v \in E$, the map $\tilde{v}: E^* \to K$ defined such that

$$\tilde{v}(u^*) = \langle u^*, v\rangle$$

for every $u^* \in E^*$ is a linear map from $E^*$ to $K$, that is, $\tilde{v}$ is a linear form in $E^{**}$. Again from the above identities, the map $c_E: E \to E^{**}$, defined such that

$$c_E(v) = \tilde{v}$$

for every $v \in E$, is a linear map. We shall see that it is injective, and that it is an isomorphism when $E$ has finite dimension.

**Definition 2.22** Given a vector space $E$ and its dual $E^*$, we say that a vector $v \in E$ and a linear form $u^* \in E^*$ are *orthogonal* if $\langle u^*, v\rangle = 0$. Given a subspace $V$ of $E$ and a subspace $U$ of $E^*$, we say that $V$ *and* $U$ *are orthogonal* if $\langle u^*, v\rangle = 0$ for every $u^* \in U$ and every $v \in V$. Given a subset $V$ of $E$ (resp. a subset $U$ of $E^*$), the *orthogonal* $V^0$ *of* $V$ is the subspace $V^0$ of $E^*$ defined such that

$$V^0 = \{u^* \in E^* \mid \langle u^*, v\rangle = 0,\ \text{for every } v \in V\}$$

(resp. the *orthogonal* $U^0$ *of* $U$ is the subspace $U^0$ of $E$ defined such that

$$U^0 = \{v \in E \mid \langle u^*, v\rangle = 0,\ \text{for every } u^* \in U\}).$$

Observe that $E^0 = 0$, and $\{0\}^0 = E^*$.

It is also easy to see that if $M \subseteq N \subseteq E$, then $N^0 \subseteq M^0 \subseteq E^*$. From this, it follows that $V \subseteq V^{00}$ for every subspace $V$ of $E$, and that $U \subseteq U^{00}$ for every subspace $U$ of $E^*$. However, even though $V = V^{00}$ is always true, when $E$ is of infinite dimension, it is not always true that $U = U^{00}$.

Given a vector space $E$ and any basis $(u_i)_{i\in I}$ for $E$, we can associate to each $u_i$ a linear form $u_i^* \in E^*$, and the $u_i^*$ have some remarkable properties.

**Definition 2.23** Given a vector space $E$ and any basis $(u_i)_{i\in I}$ for $E$, by Proposition 2.8, for every $i \in I$, there is a unique linear form $u_i^*$ such that

$$u_i^*(u_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

for every $j \in I$. The linear form $u_i^*$ is called the *coordinate form* of index $i$ w.r.t. the basis $(u_i)_{i\in I}$.

Given an index set $I$, authors often define the so called "Kronecker symbol" $\delta_{i\,j}$, such that

$$\delta_{i\,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

for all $i, j \in I$. Then, $u_i^*(u_j) = \delta_{i\,j}$.

Given a vector space $E$ and a subspace $U$ of $E$, by Theorem 2.2, every basis $(u_i)_{i\in I}$ of $U$ can be extended to a basis $(u_j)_{j\in I\cup J}$ of $E$, where $I \cap J = \emptyset$. We have the following important theorem.

**Theorem 2.26** *Let $E$ be a vector space. The following properties hold:*

*(a) For every basis $(u_i)_{i\in I}$ of $E$, the family $(u_i^*)_{i\in I}$ of coordinate forms is linearly independent.*

*(b) For every subspace $V$ of $E$, we have $V^{00} = V$.*

*(c) For every subspace $V$ of finite codimension $m$ of $E$, for every subspace $W$ of $E$ such that $E = V \oplus W$ (where $W$ is of finite dimension $m$), for every basis $(u_i)_{i\in I}$ of $E$ such that $(u_1, \ldots, u_m)$ is a basis of $W$, the family $(u_1^*, \ldots, u_m^*)$ is a basis of the orthogonal $V^0$ of $V$ in $E^*$. Furthermore, we have $V^{00} = V$, and*

$$\dim(V) + \dim(V^0) = \dim(E).$$

*(d) For every subspace $U$ of finite dimension $m$ of $E^*$, the orthogonal $U^0$ of $U$ in $E$ is of finite codimension $m$, $U^{00} = U$, and*

$$\dim(U) + \dim(U^0) = \dim(E).$$

*Proof*. (a) Assume that
$$\sum_{i\in I}\lambda_i u_i^* = 0,$$
for a family $(\lambda_i)_{i\in I}$ (of scalars in $K$). Since $(\lambda_i)_{i\in I}$ has finite support, there is a finite subset $J$ of $I$ such that $\lambda_i = 0$ for all $i \in I - J$, and we have
$$\sum_{j\in J}\lambda_j u_j^* = 0.$$
Applying the linear form $\sum_{j\in J}\lambda_j u_j^*$ to each $u_j$ ($j \in J$), by Definition 2.23, since $u_i^*(u_j) = 1$ if $i = j$ and 0 otherwise, we get $\lambda_j = 0$ for all $j \in J$, that is $\lambda_i = 0$ for all $i \in I$ (by definition of $J$ as the support). Thus, $(u_i^*)_{i\in I}$ is linearly independent.

(b) Let $(u_i)_{i\in I\cup J}$ be a basis of $E$ such that $(u_i)_{i\in I}$ is a basis of $V$ (where $I \cap J = \emptyset$). Clearly, we have $V \subseteq V^{00}$. If $V \neq V^{00}$, then $u_{j_0} \in V^{00}$ for some $j_0 \in J$ (and thus, $j_0 \notin I$). Since $u_{j_0} \in V^{00}$, $u_{j_0}$ is orthogonal to every linear form in $V^0$. In particular, let $g^*$ be the linear form defined such that
$$g^*(u_i) = \begin{cases} 1 & \text{if } i = j_0 \\ 0 & \text{if } i \neq j_0, \end{cases}$$
where $i \in I \cup J$. Then, it is clear that $g^*(u_i) = 0$ for all $i \in I$, and thus that $g^* \in V^0$. However, $g^*(u_{j_0}) = 1$, contradicting the fact that $u_{j_0}$ is orthogonal to every linear form in $V^0$. Thus, $V = V^{00}$.

(c) Let $J = I - \{1, \ldots, m\}$. Every linear form $f^* \in V^0$ is orthogonal to every $u_j$, for $j \in J$, and thus, $f^*(u_j) = 0$, for all $j \in J$. For such a linear form $f^* \in V^0$, let
$$g^* = f^*(u_1)u_1^* + \cdots + f^*(u_m)u_m^*.$$
We have $g^*(u_i) = f^*(u_i)$, for every $i$, $1 \leq i \leq m$. Furthermore, by definition, $g^*$ vanishes on all $u_j$, where $j \in J$. Thus, $f^*$ and $g^*$ agree on the basis $(u_i)_{i\in I}$ of $E$, and so, $g^* = f^*$. This shows that $(u_1^*, \ldots, u_m^*)$ generates $V^0$, and since it is also a linearly independent family, $(u_1^*, \ldots, u_m^*)$ is a basis of $V^0$. It is then obvious that $\dim(V) + \dim(V^0) = \dim(E)$, and by part (b), we have $V^{00} = V$.

(d) Let $(u_1^*, \ldots, u_m^*)$ be a basis of $U$. Note that the map $h\colon E \to K^m$ defined such that
$$h(v) = (u_1^*(v), \ldots, u_m^*(v))$$
for every $v \in E$, is a linear map, and that its kernel $\operatorname{Ker} h$ is precisely $U^0$. Then, by Proposition 2.23,
$$\dim(E) = \dim(\operatorname{Ker} h) + \dim(\operatorname{Im} h) \leq \dim(U^0) + m,$$
since $\dim(\operatorname{Im} h) \leq m$. Thus, $U^0$ is a subspace of $E$ of finite codimension at most $m$, and by (c), we have $\dim(U^0) + \dim(U^{00}) = \dim(E)$. However, it is clear that $U \subseteq U^{00}$, which implies

$\dim(U) \leq \dim(U^{00})$, and so $\dim(U^0) + \dim(U) \leq \dim(E)$. Thus, $U^0$ is of finite codimension at least $m$, which implies $\mathrm{codim}(U^0) = m$. But then, $\mathrm{codim}(U^0) = m = \dim(U)$, and since $\dim(U^0) + \dim(U^{00}) = \dim(E)$, we have $\dim(U^{00}) = m$, and we must have $U = U^{00}$. $\square$

Part (a) of Theorem 2.26 shows that

$$\dim(E) \leq \dim(E^*).$$

When $E$ is of finite dimension $n$ and $(u_1, \ldots, u_n)$ is a basis of $E$, by part (c), the family $(u_1^*, \ldots, u_n^*)$ is a basis of the dual space $E^*$, called the *dual basis* of $(u_1, \ldots, u_n)$. By part (c) and (d) of theorem 2.26, the maps $V \mapsto V^0$ and $U \mapsto U^0$, where $V$ is a subspace of finite codimension of $E$ and $U$ is a subspace of finite dimension of $E^*$, are inverse bijections. These maps set up a *duality* between subspaces of finite codimension of $E$, and subspaces of finite dimension of $E^*$.

One should be careful that this bijection does not extend to subspaces of $E^*$ of infinite dimension.

When $E$ is of infinite dimension, for every basis $(u_i)_{i \in I}$ of $E$, the family $(u_i^*)_{i \in I}$ of coordinate forms is never a basis of $E^*$. It is linearly independent, but it is "too small" to generate $E^*$. For example, if $E = \mathbb{R}^{(\mathbb{N})}$, where $\mathbb{N} = \{0, 1, 2, \ldots\}$, the map $f \colon E \to \mathbb{R}$ that sums the nonzero coordinates of a vector in $E$ is a linear form, but it is easy to see that it cannot be expressed as a linear combination of coordinate forms. As a consequence, when $E$ is of infinite dimension, $E$ and $E^*$ are not isomorphic.

When $E$ is of finite dimension $n$ and $(u_1, \ldots, u_n)$ is a basis of $E$, we noted that the family $(u_1^*, \ldots, u_n^*)$ is a basis of the dual space $E^*$ (called the dual basis of $(u_1, \ldots, u_n)$). Let us see how the coordinates of a linear form $\phi^*$ vary under a change of basis.

Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two bases of $E$, and let $P = (a_{i\,j})$ be the change of basis matrix from $(u_1, \ldots, u_n)$ to $(v_1, \ldots, v_n)$, so that

$$v_j = \sum_{i=1}^{n} a_{i\,j} u_i,$$

and let $P^{-1} = (b_{i\,j})$ be the inverse of $P$, so that

$$u_i = \sum_{j=1}^{n} b_{j\,i} v_j.$$

Since $u_i^*(u_j) = \delta_{i\,j}$ and $v_i^*(v_j) = \delta_{i\,j}$, we get

$$v_j^*(u_i) = v_j^*(\sum_{k=1}^{n} b_{k\,i} v_k) = b_{j\,i},$$

and thus

$$v_j^* = \sum_{i=1}^{n} b_{j\,i} u_i^*,$$

and similarly

$$u_i^* = \sum_{j=1}^{n} a_{i\,j} v_j^*.$$

Since

$$\varphi^* = \sum_{i=1}^{n} \varphi_i u_i^* = \sum_{i=1}^{n} \varphi_i' v_i^*,$$

we get

$$\varphi_j' = \sum_{i=1}^{n} a_{i\,j} \varphi_i.$$

Comparing with the change of basis

$$v_j = \sum_{i=1}^{n} a_{i\,j} u_i,$$

we note that this time, the coordinates $(\varphi_i)$ of the linear form $\varphi$ change in the *same direction* as the change of basis. For this reason, we say that the coordinates of linear forms are *covariant*. By abuse of language, it is often said that linear forms are *covariant*, which explains why the term *covector* is also used for a linear form.

**Remark:** In many texts using tensors, vectors are often indexed with lower indices. If so, it is more convenient to write the coordinates of a vector $x$ over the basis $(u_1, \ldots, u_n)$ as $(x^i)$, using an upper index, so that

$$x = \sum_{i=1}^{n} x^i u_i,$$

and in a change of basis, we have

$$v_j = \sum_{i=1}^{n} a_j^i u_i$$

and

$$u^i = \sum_{j=1}^{n} a_j^i v^j.$$

Dually, linear forms are indexed with upper indices. Then, it is more convenient to write the coordinates of a covector $\varphi^*$ over the dual basis $(u^{*1}, \ldots, u^{*n})$ as $(\varphi_i)$, using a lower index, so that

$$\varphi^* = \sum_{i=1}^{n} \varphi_i u^{*i}$$

and in a change of basis, we have

$$u^{*i} = \sum_{j=1}^{n} a^i_j v^{*j}$$

and

$$\varphi'_j = \sum_{i=1}^{n} a^i_j \varphi_i.$$

With these conventions, the index of summation appears once in upper position and once in lower position, and the summation sign can be safely omitted, a trick due to Einstein. For example, we can write

$$\varphi'_j = a^i_j \varphi_i$$

as an abbreviation for

$$\varphi'_j = \sum_{i=1}^{n} a^i_j \varphi_i.$$

We will now pin down the relationship between a vector space $E$ and its bidual $E^{**}$.

**Proposition 2.27** *Let $E$ be a vector space. The following properties hold:*

*(a) The linear map $c_E\colon E \to E^{**}$ defined such that*

$$c_E(v) = \tilde{v},$$

*that is, $c_E(v)(u^*) = \langle u^*, v\rangle$ for every $u^* \in E^*$, is injective.*

*(b) When $E$ is of finite dimension $n$, the linear map $c_E\colon E \to E^{**}$ is an isomorphism (called the canonical isomorphism).*

*Proof*. (a) Let $(u_i)_{i\in I}$ be a basis of $E$, and let $v = \sum_{i\in I} v_i u_i$. If $c_E(v) = 0$, then in particular, $c_E(v)(u^*_i) = 0$ for all $u^*_i$, and since

$$c_E(v)(u^*_i) = \langle u^*_i, v\rangle = v_i,$$

we have $v_i = 0$ for all $i \in I$, that is, $v = 0$, showing that $c_E\colon E \to E^{**}$ is injective.

If $E$ is of finite dimension $n$, by Theorem 2.26, for every basis $(u_1, \ldots, u_n)$, the family $(u^*_1, \ldots, u^*_n)$ is a basis of the dual space $E^*$, and thus the family $(u^{**}_1, \ldots, u^{**}_n)$ is a basis of the bidual $E^{**}$. This shows that $\dim(E) = \dim(E^{**}) = n$, and since by part (a), we know that $c_E\colon E \to E^{**}$ is injective, in fact, $c_E\colon E \to E^{**}$ is bijective (because an injective map carries a linearly independent family to a linearly independent family, and in a vector space of dimension $n$, a linearly independent family of $n$ vectors is a basis, see Proposition 2.3). $\square$

When a vector space $E$ has infinite dimension, $E$ and its bidual $E^{**}$ are never isomorphic.

When $E$ is of finite dimension and $(u_1, \ldots, u_n)$ is a basis of $E$, in view of the canonical isomorphism $c_E: E \to E^{**}$, the basis $(u_1^{**}, \ldots, u_n^{**})$ of the bidual is identified with $(u_1, \ldots, u_n)$.

Proposition 2.27 can be reformulated very fruitfully in terms of pairings.

**Definition 2.24** Given two vector spaces $E$ and $F$ over $K$, a *pairing between $E$ and $F$* is a bilinear map $\langle -, - \rangle: E \times F \to K$. Such a pairing is *non-singular* if for every $u \in E$, if $\langle u, v \rangle = 0$ for all $v \in F$, then $u = 0$, and for every $v \in F$, if $\langle u, v \rangle = 0$ for all $u \in E$, then $v = 0$.

For example, the map $\langle -, - \rangle: E^* \times E \to K$ defined earlier is a non-singular pairing (use the proof of (a) in lemma 2.27).

Given a pairing $\langle -, - \rangle: E \times F \to K$, we can define two maps $\varphi: E \to F^*$ and $\psi: F \to E^*$ as follows: For every $u \in E$, we define the linear form $\varphi(u)$ in $F^*$ such that

$$\varphi(u)(v) = \langle u, v \rangle$$

for every $v \in F$, and we define the linear form $\psi(u)$ in $E^*$ such that

$$\psi(v)(u) = \langle u, v \rangle$$

for every $u \in E$. We have the following useful proposition.

**Proposition 2.28** *Given two vector spaces $E$ and $F$ over $K$, for every non-singular pairing $\langle -, - \rangle: E \times F \to K$ between $E$ and $F$, the maps $\varphi: E \to F^*$ and $\psi: F \to E^*$ are linear and injective. Furthermore, if $E$ and $F$ have finite dimension, then this dimension is the same and $\varphi: E \to F^*$ and $\psi: F \to E^*$ are bijections.*

*Proof*. The maps $\varphi: E \to F^*$ and $\psi: F \to E^*$ are linear because a pairing is bilinear. If $\varphi(u) = 0$ (the null form), then

$$\varphi(u)(v) = \langle u, v \rangle = 0$$

for every $v \in F$, and since $\langle -, - \rangle$ is non-singular, $u = 0$. Thus, $\varphi: E \to F^*$ is injective. Similarly, $\psi: F \to E^*$ is injective. When $F$ has finite dimension $n$, we have seen that $F$ and $F^*$ have the same dimension. Since $\varphi: E \to F^*$ is injective, we have $m = \dim(E) \leq \dim(F) = n$. The same argument applies to $E$, and thus $n = \dim(F) \leq \dim(E) = m$. But then, $\dim(E) = \dim(F)$, and $\varphi: E \to F^*$ and $\psi: F \to E^*$ are bijections. $\square$

When $E$ has finite dimension, the non-singular pairing $\langle -, - \rangle: E^* \times E \to K$ yields another proof of the existence of a natural isomorphism between $E$ and $E^{**}$. Interesting non-singular pairings arise in exterior algebra. We now show the relationship between hyperplanes and linear forms.

## 2.10 Hyperplanes and Linear Forms

Actually, Proosition 2.29 below follows from parts (c) and (d) of Theorem 2.26, but we feel that it is also interesting to give a more direct proof.

**Proposition 2.29** *Let $E$ be a vector space. The following properties hold:*

*(a) Given any nonnull linear form $f^* \in E^*$, its kernel $H = \mathrm{Ker}\, f^*$ is a hyperplane.*

*(b) For any hyperplane $H$ in $E$, there is a (nonnull) linear form $f^* \in E^*$ such that $H = \mathrm{Ker}\, f^*$.*

*(c) Given any hyperplane $H$ in $E$ and any (nonnull) linear form $f^* \in E^*$ such that $H = \mathrm{Ker}\, f^*$, for every linear form $g^* \in E^*$, $H = \mathrm{Ker}\, g^*$ iff $g^* = \lambda f^*$ for some $\lambda \neq 0$ in $K$.*

*Proof*. (a) If $f^* \in E^*$ is nonnull, there is some vector $v_0 \in E$ such that $f^*(v_0) \neq 0$. Let $H = \mathrm{Ker}\, f^*$. For every $v \in E$, we have

$$f^*\left(v - \frac{f^*(v)}{f^*(v_0)} v_0\right) = f^*(v) - \frac{f^*(v)}{f^*(v_0)} f^*(v_0) = f^*(v) - f^*(v) = 0.$$

Thus,

$$v - \frac{f^*(v)}{f^*(v_0)} v_0 = h \in H,$$

and

$$v = h + \frac{f^*(v)}{f^*(v_0)} v_0,$$

that is, $E = H + Kv_0$. Also, since $f^*(v_0) \neq 0$, we have $v_0 \notin H$, that is, $H \cap Kv_0 = 0$. Thus, $E = H \oplus Kv_0$, and $H$ is a hyperplane.

(b) If $H$ is a hyperplane, $E = H \oplus Kv_0$ for some $v_0 \notin H$. Then, every $v \in E$ can be written in a unique way as $v = h + \lambda v_0$. Thus, there is a well-defined function $f^*\colon E \to K$, such that, $f^*(v) = \lambda$, for every $v = h + \lambda v_0$. We leave as a simple exercise the verification that $f^*$ is a linear form. Since $f^*(v_0) = 1$, the linear form $f^*$ is nonnull. Also, by definition, it is clear that $\lambda = 0$ iff $v \in H$, that is, $\mathrm{Ker}\, f^* = H$.

(c) Let $H$ be a hyperplane in $E$, and let $f^* \in E^*$ be any (nonnull) linear form such that $H = \mathrm{Ker}\, f^*$. Clearly, if $g^* = \lambda f^*$ for some $\lambda \neq 0$, then $H = \mathrm{Ker}\, g^*$. Conversely, assume that $H = \mathrm{Ker}\, g^*$ for some nonnull linear form $g^*$. From (a), we have $E = H \oplus Kv_0$, for some $v_0$ such that $f^*(v_0) \neq 0$ and $g^*(v_0) \neq 0$. Then, observe that

$$g^* - \frac{g^*(v_0)}{f^*(v_0)} f^*$$

is a linear form that vanishes on $H$, since both $f^*$ and $g^*$ vanish on $H$, but also vanishes on $Kv_0$. Thus, $g^* = \lambda f^*$, with

$$\lambda = \frac{g^*(v_0)}{f^*(v_0)}.$$

□

If $E$ is a vector space of finite dimension $n$ and $(u_1, \ldots, u_n)$ is a basis of $E$, for any linear form $f^* \in E^*$, for every $x = x_1u_1 + \cdots + x_nu_n \in E$, we have

$$f^*(x) = \lambda_1 x_1 + \cdots + \lambda_n x_n,$$

where $\lambda_i = f^*(u_i) \in K$, for every $i$, $1 \leq i \leq n$. Thus, with respect to the basis $(u_1, \ldots, u_n)$, $f^*(x)$ is a linear combination of the coordinates of $x$, as expected.

We leave as an exercise the fact that every subspace $V \neq E$ of a vector space $E$, is the intersection of all hyperplanes that contain $V$. We now consider the notion of transpose of a linear map and of a matrix.

## 2.11 Transpose of a Linear Map and of a Matrix

Given a linear map $f: E \to F$, it is possible to define a map $f^\top: F^* \to E^*$ which has some interesting properties.

**Definition 2.25** Given a linear map $f: E \to F$, the *transpose* $f^\top: F^* \to E^*$ *of* $f$ is the linear map defined such that

$$f^\top(v^*) = v^* \circ f,$$

for every $v^* \in F^*$. Equivalently, the linear map $f^\top: F^* \to E^*$ is defined such that

$$\langle v^*, f(u)\rangle = \langle f^\top(v^*), u\rangle,$$

for all $u \in E$ and all $v^* \in F^*$.

It is easy to verify that the following properties hold:

$$\begin{aligned}
(f^\top)^\top &= f, \\
(f + g)^\top &= f^\top + g^\top \\
(g \circ f)^\top &= f^\top \circ g^\top \\
\mathrm{id}_E^\top &= \mathrm{id}_{E^*}.
\end{aligned}$$

Note the reversal of composition on the right-hand side of $(g \circ f)^\top = f^\top \circ g^\top$.

The following proposition shows the relationship between orthogonality and transposition.

**Proposition 2.30** *Given a linear map $f: E \to F$, for any subspace $U$ of $E$, we have*

$$f(U)^0 = (f^\top)^{-1}(U^0).$$

*As a consequence,* $\mathrm{Ker}\, f^\top = (\mathrm{Im}\, f)^0$ *and* $\mathrm{Ker}\, f = (\mathrm{Im}\, f^\top)^0$.

*Proof*. We have

$$\langle v^*, f(u)\rangle = \langle f^\top(v^*), u\rangle,$$

for all $u \in E$ and all $v^* \in F^*$, and thus, we have $\langle v^*, f(u)\rangle = 0$ for every $u \in U$, i.e., $v^* \in f(U)^0$, iff $\langle f^\top(v^*), u\rangle = 0$ for every $u \in U$, i.e., $v^* \in (f^\top)^{-1}(U^0)$, proving that

$$f(U)^0 = (f^\top)^{-1}(U^0).$$

Since we already observed that $E^0 = 0$, letting $U = E$ in the above identity, we obtain that

$$\operatorname{Ker} f^\top = (\operatorname{Im} f)^0.$$

The identity $\operatorname{Ker} f = (\operatorname{Im} f^\top)^0$ holds because $(f^\top)^\top = f$. $\square$

The following theorem shows the relationship between the rank of $f$ and the rank of $f^\top$.

**Theorem 2.31** *Given a linear map $f: E \to F$, the following properties hold.*

*(a) The dual $(\operatorname{Im} f)^*$ of $\operatorname{Im} f$ is isomorphic to $\operatorname{Im} f^\top = f^\top(F^*)$.*

*(b) $\operatorname{rk}(f) \leq \operatorname{rk}(f^\top)$. If $\operatorname{rk}(f)$ is finite, we have $\operatorname{rk}(f) = \operatorname{rk}(f^\top)$.*

*Proof*. (a) Consider the linear maps

$$E \xrightarrow{p} \operatorname{Im} f \xrightarrow{j} F,$$

where $E \xrightarrow{p} \operatorname{Im} f$ is the surjective map induced by $E \xrightarrow{f} F$, and $\operatorname{Im} f \xrightarrow{j} F$ is the injective inclusion map of $\operatorname{Im} f$ into $F$. By definition, $f = j \circ p$. To simplify the notation, let $I = \operatorname{Im} f$. Since $E \xrightarrow{p} I$ is surjective, let $I \xrightarrow{s} E$ be a section of $p$, and since $I \xrightarrow{j} F$ is injective, let $F \xrightarrow{r} I$ be a retraction of $j$. Then, we have $p \circ s = \mathrm{id}_I$ and $r \circ j = \mathrm{id}_I$, and since $(p \circ s)^\top = s^\top \circ p^\top$, $(r \circ j)^\top = j^\top \circ r^\top$, and $\mathrm{id}_I^\top = \mathrm{id}_{I^*}$, we have

$$s^\top \circ p^\top = \mathrm{id}_{I^*},$$

and

$$j^\top \circ r^\top = \mathrm{id}_{I^*}.$$

Thus, $F^* \xrightarrow{j^\top} I^*$ is surjective, and $I^* \xrightarrow{p^\top} E^*$ is injective. Since $f = j \circ p$, we also have

$$f^\top = (j \circ p)^\top = p^\top \circ j^\top,$$

and since $F^* \xrightarrow{j^\top} I^*$ is surjective, and $I^* \xrightarrow{p^\top} E^*$ is injective, we have an isomorphism between $(\operatorname{Im} f)^*$ and $f^\top(F^*)$.

(b) We already noted that part (a) of Theorem 2.26 shows that $\dim(E) \leq \dim(E^*)$, for every vector space $E$. Thus, $\dim(\operatorname{Im} f) \leq \dim((\operatorname{Im} f)^*)$, which, by (a), shows that $\operatorname{rk}(f) \leq \operatorname{rk}(f^\top)$. When $\dim(\operatorname{Im} f)$ is finite, we already observed that as a corollary of Theorem 2.26, $\dim(\operatorname{Im} f) = \dim((\operatorname{Im} f)^*)$, and thus, we have $\operatorname{rk}(f) = \operatorname{rk}(f^\top)$. $\square$

The following proposition shows the relationship between the matrix representing a linear map $f: E \to F$ and the matrix representing its transpose $f^\top: F^* \to E^*$.

**Proposition 2.32** *Let $E$ and $F$ be two vector spaces, and let $(u_1, \ldots, u_n)$ be a basis for $E$, and $(v_1, \ldots, v_m)$ be a basis for $F$. Given any linear map $f: E \to F$, if $M(f)$ is the $m \times n$-matrix representing $f$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$, the $n \times m$-matrix $M(f^\top)$ representing $f^\top: F^* \to E^*$ w.r.t. the dual bases $(v_1^*, \ldots, v_m^*)$ and $(u_1^*, \ldots, u_n^*)$ is the transpose $M(f)^\top$ of $M(f)$.*

*Proof*. Recall that the entry $a_{i\,j}$ in row $i$ and column $j$ of $M(f)$ is the $i$-th coordinate of $f(u_j)$ over the basis $(v_1, \ldots, v_m)$. By definition of $v_i^*$, we have $\langle v_i^*, f(u_j)\rangle = a_{i\,j}$. The entry $a_{j\,i}^\top$ in row $j$ and column $i$ of $M(f^\top)$ is the $j$-th coordinate of $f^\top(v_i^*)$ over the basis $(u_1^*, \ldots, u_n^*)$, which is just $f^\top(v_i^*)(u_j) = \langle f^\top(v_i^*), u_j\rangle$. Since

$$\langle v_i^*, f(u_j)\rangle = \langle f^\top(v_i^*), u_j\rangle,$$

we have $a_{i\,j} = a_{j\,i}^\top$, proving that $M(f^\top) = M(f)^\top$. $\square$

We now can give a very short proof of the fact that the rank of a matrix is equal to the rank of its transpose.

**Proposition 2.33** *Given a $m \times n$ matrix $A$ over a field $K$, we have $\mathrm{rk}(A) = \mathrm{rk}(A^\top)$.*

*Proof*. The matrix $A$ corresponds to a linear map $f: K^n \to K^m$, and by Theorem 2.31, $\mathrm{rk}(f) = \mathrm{rk}(f^\top)$. By Proposition 2.32, the linear map $f^\top$ corresponds to $A^\top$. Since $\mathrm{rk}(A) = \mathrm{rk}(f)$, and $\mathrm{rk}(A^\top) = \mathrm{rk}(f^\top)$, we conclude that $\mathrm{rk}(A) = \mathrm{rk}(A^\top)$. $\square$

Thus, given an $m \times n$-matrix $A$, the maximum number of linearly independent columns is equal to the maximum number of linearly independent rows. There are other ways of proving this fact that do not involve the dual space, but instead some elementary transformations on rows and columns.