Linear Algebra and Matrices

Martin Fluch

Spring 2007

May 14, 2007

Based closely on the book “Lineare Algebra I” by F. Lorenz, 1992.

To

Анна

Simplicity is beauty,
Mathematics is simplicity.

Contents

Introduction

Linear algebra is the branch of mathematics concerned with the study of vectors, vector spaces (also called linear spaces), linear transformations, and systems of linear equations. Vector spaces are a central theme in modern mathematics; thus, linear algebra is widely used in both abstract algebra and functional analysis. Linear algebra also has a concrete representation in analytic geometry and it is generalized in operator theory. It has extensive applications in the natural sciences and the social sciences, since nonlinear models can often be approximated by a linear model.

We will begin our studies by studying systems of linear equations. Without becomming to formal in the notation and language we will study the basic properties of the solutions for *homogeneous* and *nonhomogeneous systems of linear equations*. We will get known to the Gaussian algorithm for solving systems of linear equations. This algorithm will re-occur repeatedly in this lecture note. Towards the end of this first chapter two problems will in a natural way catch our attention. In order to solve them we need to begin to formalize the observations we made sofar.

The formalisation will begin by extracting the esential properties of the numbers we have used in the first chapter. This will lead to the concept of a *field* (Definition 2.1). Next we will formalize the properties which the solutions of a homogeneous system of equations posesses: the sum of two solutions of a homogeneous system of equations is again a solution this system of equations, and the same is true if we multiply a solution of such a system of equations by a number (that is an element of a field). This will lead to the concept of a *vector space* (Definition 2.4). Roughly spoken a vector space over a field is a set V where we can form sums of arbitrary elements and where we can multiply any element by *scalars* of the field and such that this addition and scalar multiplication satisfies certain rules which seem natural to us.

After we have made these essential definitions we will begin to study vector spaces more in detail. We will encounter basic but very important concepts of linear algebra, amongst others:

- linear combinations of vectors,
- basis of a vector spaces,
- linear dependence of vectors,
- the rank of a system of vectors (and related concepts),
- the dimension of a vector space.

Maybe one of the most essential results will be the theorem about the existence of a basis. It takes just 6 words to formulate this theorem which turns out to reach till the foundations of the mathematics: "*Every vector space has a basis.*" We will proof it only in the special case of finite dimensional vector spaces as the more general result will need some heavy machinery of axiomatic set theory. Though for completeness we have included this proof in the appendix of this lecture notes (together with an more detailed explanation about its importance; see Appendix C).

Towards the the end of this second chapter we will finally be able to answer the problems answer the problems which did arise in the end of the first chapter.

In the third chapter we will then start to study the relation ship between. We will introduce the concept of a *linear map* between vector spaces (Definition 3.1). The whole chapter is devoted to the study of these kind of maps. One of the main theme of this chapter will be the matrix description of linear maps between finite dimensional vector spaces. We will explore the relation ship between matrices and linear maps and what we can all conclude from that. Two theorems will provide us with the necessary information:

- "*The vector space* $\mathrm{Hom}_F(V,W)$ *of all linear maps between a* n *dimensional vector space* V *and a* m*-dimensional vector space* W *is isomorphic as vector spaces to the the vector space of all* $m \times n$*-matrices* $F^{m,n}$." (Theorem 3.27)
- "*The endomorphism ring* $\mathrm{End}_F(V)$ *of an* n*-dimensional vector space is isomorphic as* F*-algebras to the* F*-algebra of all* $n \times n$*-matrices* $M_n(F)$." (Theorem 3.35)

The proper understanding of these two theorems might take time but they are essential in order to really understand linear algebra. Other important topics of the third chapter will be:

- isomorphism and isomorphism of vector spaces,
- rank of a linear map,
- dimension formla for linear maps,
- the general and the special linear group,
- equivalence and similarity of matrices,
- normal form of matrices upto equivalence.

During the third chapter we will encounter that every invertible $n \times n$-matrix A can be written as a product

$$A = SD$$

where S is a matrix of the special linear group and D is a very simple diagonal matrix (Theorem 3.55). As a natural problem will arise the question about the uniqueness of this decomposition. It will turn out that the theory developed upto the third chapter is not enough to give a proper answer to this problem. We will need a new concept and this will lead to the the definition of a determinant function.

The fourth chapter will then be devoted to the first studies of determinant functions. First we introduce the new concept and show what hypotetical properties a determinant function would have. It will turn out that the detreminant function – in case it exists – must be unique. This will be the reason why we will later be able to give an afirmative answer about the uniqueness of the above decomposition. But we will first have to show that a determinant function exists and this is done in the second section of the chapter about determinants. When we are finally convinced about the existence of determinants we will study in the remaining part of the chapter some basic properties of the determinant function. The chapter will finish with the presentation of the Leibniz formula which shows the beauty and symmetries of the determinat function (Theorem 4.28).

CHAPTER 1

Systems of Linear Equations

Following closely [**Lor92**] we shall give an introducton to systems of linear equations where we will encounter the first time linear structures. In the next chapter we will then study linear structures in a more general setting.

1. Two Linear Equations with Two Variables

We consider the following system of two linar equations:

$$\begin{aligned} ax + by &= e \\ cx + dy &= f. \end{aligned} \tag{1}$$

Here a, b, c, d, e and f are numbers and x and y are variables. Our concern is which values of x and y satisfy the two equations above simultanously.

In order to avoid triviality we may assume that some of the numbers of the on the left hand side of (1) are not equal to zero. Let us assume that $a \neq 0$. Then we can substract c/a times the first equation of (1) from the second equation and we get

$$\begin{aligned} ax + by &= e \\ d'y &= f', \end{aligned} \tag{2}$$

with $d' = d - bc/a$ and $f' = f - ec/a$. Note that we can recover from this system of linear equations again the first one by adding c/a the first equation of (2) to the second equation of (2). We say that (1) and (2) are equivalent systems of linear equations.

If some values for x and y satisfy simultaneously the linear equations of (1) then they also satisfy the linear equations of (2) simultaneously. And the converse is also true. In general, equivalent systems of linear equations have exact the same solutions.

Note that the system of linear equations (2) is more easy to solve than the the first system. This is because in the second equation only appeares one variable instead of two, namely y. Since $a \neq 0$ we get that the second equation of (2) is equivalent with

$$(ad - bc)y = af - ec. \tag{3}$$

Thus the solveability of (1) depends very much on the fact whether the number

$$\delta(a, b, c, d) := ad - bc \tag{4}$$

is equal to 0 or not.

Case 1: Assume that $\delta(a, b, c, d) \neq 0$. Then (3) is equivalent with

$$y = \frac{af - ec}{ad - bc} = \frac{af - ec}{\delta(a, b, c, d)}$$

which we can also write as

$$y = \frac{\delta(a, e, c, f)}{\delta(a, b, c, d)}.$$

Thus the original system of linear equations (1) is equivalent with

$$\begin{aligned} ax + by &= e \\ y &= \delta(a,e,c,f)/\delta(a,b,c,d). \end{aligned} \tag{5}$$

It follows after a short calculation that since we have made the assumption that $a \neq 0$ that

$$x = \frac{ed - bf}{ad - bc} = \frac{\delta(e,b,f,d)}{\delta(a,b,c,d)}$$

is the unique solltion for x to the linear system (5).

Case 2: Assume that $\delta(a,b,c,d) = 0$. Straight from (3) follows that the system of linear equations is not always solveable. This is beacause the constants e and f of the righthand side of (3) can be choosen such that $af - ec \neq 0$ and in this case there exist no y satisfying the equation (3). This happens for example if $f = 1$ and $e = 0$, because then $af - ec = a \neq 0$ since we assumed that $a \neq 0$.

But in the case that $af - ec = 0$ we can choose the value for y freely and y together with

$$x = \frac{e - by}{a}$$

is a solution for the system of linear equations (1). Thus the system of linear equations (1) has a solution but this solution is not unique.

Collecting the above observations it is not difficult to prove the following result:

Proposition 1.1. *Consider the system of linear equations* (1) *in the two unknown variables x and y. Then we have the following three cases:*

Case 1: $a = b = c = d = 0$.

Then (1) *is solvable if and only if $e = f = 0$ and in this case any pair of numbers x and y is a sollution.*

Case 2: *Not all of the coefficients a, b, c, d are equal* 0*, but $\delta(a,b,c,d) = 0$.*

Then (1) *is not solvable if $\delta(a,e,c,f) = af - ec \neq 0$ or $\delta(e,b,f,d) = ed - bf \neq 0$. But if $\delta(a,e,c,f) = \delta(e,b,f,d) = 0$, then* (1) *is solveable and we have (if $a \neq 0$ what we can* always *achive by exchanging the equations or renaming the unknown variables) that all the solutions of* (1) *are given by*

$$x = \frac{1}{a}(e - bt), \quad x = t$$

where t denotes an arbitrary number.

Case 3: $\delta(a,b,c,d) \neq 0$.

Then (1) *posseses a unique sollution and this sollution is given by*

$$x = \frac{\delta(e,b,f,d)}{\delta(a,b,c,d)} \qquad \textit{and} \qquad y = \frac{\delta(a,e,c,f)}{\delta(a,b,c,d)}. \qquad \square$$

The above considerations give already an exhaustive description of what can happen when solving a system of linear equations, even if it consists of more equations and more variables: either the system will have exactly one sollution, or many solutions or no solutions. It will turn out that whether there exist a unique solution or not will always solely depend on the left hand side of the equations, that is on the coefficients of the variables (in the above case of the system of linear equations (1) these are the numbers a, b, c and d). Note that we will encounter the number $\delta(a,b,c,d)$ later in this lecture under the name "*determinant*".

2. Basic Notations for Systems of Linear Equations

We consider now an arbitrary system of linear equations

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &=& b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &=& b_2 \\ \vdots \qquad\quad \vdots \quad \vdots \qquad\quad \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &=& b_m \end{array} \tag{6}$$

of m equations in the n "unknown" variables $x_1, \ldots, x_n$.

The numbers a_{ij} $(i = 1, \ldots, m;\ j = 1, \ldots, n)$ are called the *coefficients* of the system of equations.[1] It is usefull to arrange the coefficents of the system of equations (6) in the following form:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{7}$$

Such a scheme of numbers is called a *matrix*. If one wants to emphasis the dimension of the above matrix, then one can also say that the scheme of numbers (7) is a $m \times n$-matrix. We consider the *columns* of this matrix:

$$v_1 := \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad v_2 := \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \quad \ldots \quad v_n := \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

One can consider these columns as $m \times 1$-matrices. Each of those matrices is a ordered system of m numbers and is also called an *m-tuple*. We can form also an m-tuple from the numbers $b_1, \ldots, b_m$ on the right side of the system of linear equations (6), namely

$$b := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Note the use of the symbol ":=" in the previous two equations. It does not denote an equality but rather denotes the definition of what is on the left side of the symbol. For example $x := 2$ means that x is *set by definition* equal to 2.

We can define in a natural way an addition of two m-tuples of numbers by

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{pmatrix} := \begin{pmatrix} c_1 + d_1 \\ c_2 + d_2 \\ \vdots \\ c_m + d_m \end{pmatrix}$$

and likewise the multiplication of an m-tuple by a number a by

$$a \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} := \begin{pmatrix} ac_1 \\ ac_2 \\ \vdots \\ ac_m \end{pmatrix}$$

[1] When we talk in the following of number one may think of real numbers, that is elements of the *field* $\mathbb{R}$. But all our considerations will be true in the case where $\mathbb{R}$ is replaced by an arbitrary field F. (What a field F is precisely we will introduce in the next chapter.)

Using this notation we can write the system of linear equations (6) in the following compact form:

$$x_1v_1 + x_2v_2 + \cdots + x_nv_n = b \tag{8}$$

We call the n-tuple

$$x := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

consisting of the numbers $x_1, \ldots, x_n$ which satisfy (6) or equivalently (8) a solution of the system of linear equations.

We will give the class of system of linear equations where the numbers $b_1, \ldots, b_m$ are all equal to 0 a special name by calling them *homogeneous* systems of linear equations.[2] If a system of linear equations (6) is given, then we call the system of linear equations which is derived from (6) by replacing the $b_1, \ldots, b_m$ by 0 the homogeneous system of linear equations *associated* with (6). We denote

$$0 := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

the m-tuple which consists of the numbers 0 only. (Note that we use the same symbol "0" for both, the number zero and the m-tuple consisting of only zeros.) Then the homogeneous system of linear equations which is assoziated with the system of linear equations (6) or equivalently (8) can be written as

$$x_1v_1 + x_2v_2 + \cdots + x_nv_n = 0. \tag{9}$$

Proposition 1.2. *One obtains all solutions of the nonhomogeneous system of linear equations* (6) *or equivalently* (8) *if one adds to a specific solution x of the system all the solutions of the homogeneous system* (9) *associated with it.*

PROOF. In order to avoid triviality we may assume that the system (6) has a solution x. Then if x' is a solution of the homogeneous system (9) we get

$$\begin{aligned}(x_1 + x_1')v_1 + (x_2 + x_1')v_2 + \cdots + (x_n + x_n')v_n = \\ (x_1v_1 + x_2v_2 + \cdots + x_nv_n) + (x_1'v_1 + x_2'v_2 + \cdots + x_n'v_n) = b + 0 = b.\end{aligned}$$

Thus $x + x'$ is also a solution of the system (6).

On the other hand, if both x and x' are solutions of the system (6), then $x' - x$ is a solution of the homogeneous system (9) since

$$\begin{aligned}(x_1' - x_1)v_1 + (x_2' - x_1)v_2 + \cdots + (x_n' - x_n)v_n = \\ (x_1'v_1 + x_2'v_2 + \cdots + x_n'v_n) - (x_1v_1 + x_2v_2 + \cdots + x_nv_n) = b - b = 0.\end{aligned}$$

Thus the solution $x' = x + (x' - x)$ is the sum of the specific solution x of the system (6) and the solution $x' - x$ of the homogeneous system (9). □

Let us denote by M the set of all solutions of (6). In the case that (6) is solveable we have that $M \neq \emptyset$ (where $\emptyset$ denotes the empty set). If M_0 denotes the set of all solutions of the homogeneous system of linear equations (8) assoziated with (6), then we can write the content of Proposition 1.2 in the compact form

$$M = x + M_0 \tag{10}$$

[2] A system of linear equations which is not necessarily homogeneous is called *nonhomogeneous*.

where x is some specific solution of (6) and $x + M_0 := \{x + x' : x' \in M\}$.

Now using (9) one can make the following observations about the solutions M_0 of an given homogeneous system of linear equations:

$$\text{If } x \in M_0 \text{ and } x' \in M_0 \text{ then also } x + x' \in M_0, \tag{11}$$

and:

$$\text{If } x \in M_0 \text{ and } c \text{ is any number then also } cx \in M_0. \tag{12}$$

Thus the set of solutions M_0 of some homogeneous system of linear equations cannot be just any set but must obey the requirements (11) and (12). If we denote by F^n the set of all n-tuples then we call a subset $U \subset F^n$ a *subspace* of F^n if we have

$$x, x' \in U \Rightarrow x + x' \in U \tag{13}$$

and

$$x \in U \Rightarrow cx \in U \text{ (for every number } c). \tag{14}$$

Thus, if we use the above notation we write the observations (11) and (12) in the following compact form.

Proposition 1.3. *The set of M_0 all solutions of a homogeneous system of linear equations in n unknown variables is a subspace of F^n.* □

Note that the requirement $M_0 \neq \emptyset$ is automatically satsified since the n-tuple 0 is always a solution of a homogeneuos system of linear equations in n-variables. Therefore we call the n-tuple 0 the *trivial solution* of a homogeneous system of liear equations. Now one gets as a consequence of Proposition 1.2 or equation (10) the following result.

Proposition 1.4. *A solvable nonhomogeneous system of linear equations has only a unique solution if and only if the associated homogeneous system of linear equations has only the trivial solution.* □

3. Elementary Transformations of Systems of Linear Equations and Elementary Row Transformations of Matrices

In this section we shall introduce a simple way to solve systems of linear equations: the Gaussian elimination algorithm.[3]

We begin with an arbitrary nonhomogeneous system of linear equations (6) in n variables. Our aim is – similar as in Section 1 – to convert this system of linear equations step by step into a system of linear equations which where the question about it solveability is easier to answer. Of course one step should transform a given system of linear equations S into a system of linear equations S' which is *equivalent* with the system S. That is, we demand that

$$x \text{ is a solution of } S \iff x \text{ is a solution of } S'$$

for every possible n-tuple x. Beside this natural requirement one is of course interested in keeping the transformations in each step as simple as possible. We will allow the following three *elementary transformations*:

(I) Adding a multiple of one equation to *another* equation.
(II) Exchanging two equations with each other.
(III) Multiplying one equation with a non-zero number.

[3]Named after the German mathematician and scientist Carl Friedrich Gauss, 1777–1855.

One verifies immediately that each of those transformations transform a system of linear equations into an equivalent one. Now the system of linear equations (6) is completely determined by the *extended coefficient matrix*

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix} \tag{15}$$

which is a $m \times (n+1)$-matrix. (If we will leave away the right most coloumn of the above matrix, then we will call it the *simple coefficient matrix* of the system of linear equations (6).) The transformations of type I, II and III will result in a natural way in row transformations of the extendet coefficient matrix (15). To simplify notation we will consider in the following elementary row transformations for arbitrary matrices. Therefore let

$$C := \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1q} \\ c_{21} & c_{22} & \cdots & c_{2q} \\ \vdots & & & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pq} \end{pmatrix} \tag{16}$$

be an arbitrary $p \times q$-matrix. We call the entries c_{ij} the *coefficients of the matrix* C. We shall denote by $u_1, u_2, \ldots, u_p$ the rows of the matrix C. That is for $i = 1, 2, \ldots p$ we let u_i be the (horizontaly written) q-tuple

$$u_i := (c_{i1}, c_{i2}, \ldots, c_{iq}).$$

An *elementary row transformation of a matrix* C shall now – in analogy to the above defined elementary transformations of a system of linear equations – mean one of the following three transformations.

(I) Adding a multiple of one row of C to *another* row of C.
(II) Exchanging two rows of C with each other.
(III) Multiplying one row of C with a non-zero number.

In order to simplify notation we will subsequently also allow *switching columns* as an elementary operation on matrices. Note that switching columns of the extended coefficient matrix (15) which do not involve the b_i represents renaming the unknown variables $x_1, \ldots, x_n$ in the system of linear equations (6).

Let us now begin with a $p \times q$-matrix C as in (16). If all coefficients of C are zero then there is nothing to do. Thus we may assume that some coefficient is different from zero. In case this coefficient is in the i-th row, $i \neq 1$, we use a elementary row transformation of type II to switch the i-th with the first row. Thus in the first row of the so optained matrix – which we call again C – there is now at least one coefficient c_{1k} different from zero. If needed we can switch the k-th with the first column and thus we may assume from the beginning that

$$c_{11} \neq 0.$$

Now we add for every $i \geq 2$ the $(-c_{i1}/c_{11})$-times of the first row u_1 to the i-th row u_i. For the matrix C' with the rows u'_i holds then

$$\begin{aligned} u'_1 &:= u_1 \\ u'_i &:= u_i - (c_{i1}/c_{11})u_1. \end{aligned}$$

Then the matrix C' has the form

$$C' = \left(\begin{array}{cccc} c_{11} & c_{12} & \cdots & c_{1q} \\ \hline 0 & c'_{22} & \cdots & c'_{2q} \\ \vdots & & & \vdots \\ 0 & c'_{p2} & \cdots & c'_{pq} \end{array}\right)$$

Below the coefficient c_{11} of the matrix C' there are only zeros. If all the coefficients below the first row are zero, then there is nothing to do. Thus we may assume – for similar reasons as above – that

$$c'_{22} \neq 0.$$

Now we apply the same procedure as we had applied to the matrix C in the beginning to the columns $u'_2, \ldots, u'_p$, that is the matrix we optain from C' by leaving away the first row $u_1 = u'_1$. For every $i \geq 3$ we add the $(-c'_{i2}/c'_{22})$-times of u'_2 to u'_i. The matrix C'' which we optain, has now the rows:

$$\begin{aligned} u''_1 &:= u_1 \\ u''_2 &:= u'_2 \\ u''_i &:= u'_i - (c'_{i2}/c'_{22})u'_2. \end{aligned}$$

Thus the matrix C'' is of the form

$$C'' = \left(\begin{array}{ccccc} c_{11} & c_{12} & c_{13} & \cdots & c_{1q} \\ 0 & c'_{22} & c'_{23} & \cdots & c'_{2q} \\ \hline 0 & 0 & c''_{33} & \cdots & c''_{3q} \\ \vdots & & & & \vdots \\ 0 & 0 & c''_{p3} & \cdots & c''_{pq} \end{array}\right)$$

If we continue this procedure we will finally optain a $p \times q$-matrix D of the following form:

$$D = \left(\begin{array}{cccccc|c} d_{11} & & & & & & \\ 0 & d_{22} & & & * & & \\ 0 & 0 & d_{33} & & & & \\ 0 & 0 & 0 & d_{44} & & & * \\ \vdots & \vdots & \ddots & \ddots & \ddots & & \\ 0 & 0 & \ldots & 0 & 0 & d_{rr} & \\ \hline & & & 0 & & & 0 \end{array}\right) \tag{17}$$

where r is a certain natural number such that

$$0 \leq r \leq p \qquad \text{and} \qquad 0 \leq r \leq q \tag{18}$$

and

$$d_{ii} \neq 0 \qquad (1 \leq i \leq r). \tag{19}$$

Note that the two zeros below the horizontal line of the matrix (17) denote *zero-matrices*, that is matrices where every coefficient is equal to zero. Likewise the $*$ symbolizes arbitrary entries in the matrix D which we do not further care about. In the case that $r = 0$ it means that D is the $p \times q$-zero-matrix. On the other hand, it is clear how the cases $r = p$ or $r = q$ have to be understood.

Proposition 1.5. *Let C be an arbitrary $p \times q$-matrix. Then we can – using elementary row transformations of type I and type II and suitable column exchanges – transform C into a matrix D of the from given in* (17) *where the conditions* (18) *and* (19) *are satisfied.* □

Note that since $d_{ii} \neq 0$ $(1 \leq i \leq r)$ it is actually possible bring the matrix D into the following form.

$$\left(\begin{array}{cccccc|c} d_{11} & 0 & 0 & \dots & \dots & 0 & \\ 0 & d_{22} & 0 & & & 0 & \\ 0 & 0 & d_{33} & \ddots & & \vdots & \\ 0 & 0 & 0 & d_{44} & \ddots & \vdots & * \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 & \\ 0 & 0 & \dots & 0 & 0 & d_{rr} & \\ \hline & & & 0 & & & 0 \\ & & & & & & \end{array}\right) \tag{20}$$

The $r \times r$-matrix in left upper part of the matrix (20) has beside the coefficients on the diagonal only zeros. We call such a matrix a *diagonal matrix*.

And if we use row transformations of type III and multiply for every $1 \leq i \leq r$ the i-th row of the above matrix with the factor $1/d_{ii}$ we see that finally we can transform the matrix (17) into a matrix of the form

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & \dots & \dots & 0 & \\ 0 & 1 & 0 & & & 0 & \\ 0 & 0 & 1 & \ddots & & \vdots & \\ 0 & 0 & 0 & 1 & \ddots & \vdots & * \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 & \\ 0 & 0 & \dots & 0 & 0 & 1 & \\ \hline & & & 0 & & & 0 \\ & & & & & & \end{array}\right) \tag{21}$$

where the left upper part of the matrix (21) is a diagonal $r \times r$-matrix with only ones on the diagonal. We call this matrix the $r \times r$-*identity matrix* I_r. (In case there is no danger of confusion we may also denote the identity matrix just by the symbol I.)

Example. Consider for example the following 4×5-matrix, which we shall transform into the form (21):

$$C := \begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 3 & 9 & 10 & 1 & 2 \\ 0 & 2 & 7 & 3 & -1 \\ 2 & 8 & 12 & 2 & 1 \end{pmatrix}$$

Here the upper left coefficien $c_{11} = 1$ of the matrix C is already different from 0. Thus we can begin to add suitable multiples of the first row to the rows below. In our case this means we substract 3-times the first row from the second row and we substract 2-times the fist frow from the forth row.

$$\begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 3 & 9 & 10 & 1 & 2 \\ 0 & 2 & 7 & 3 & -1 \\ 2 & 8 & 12 & 2 & 1 \end{pmatrix} \quad \rightarrow \quad \begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 0 & 0 & -5 & -5 & 2 \\ 0 & 2 & 7 & 3 & -1 \\ 0 & 2 & 2 & -2 & 1 \end{pmatrix}$$

Then exchanging the second with the forth row yields:

$$\rightarrow \quad \begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 0 & 2 & 2 & -2 & 1 \\ 0 & 2 & 7 & 3 & -1 \\ 0 & 0 & -5 & -5 & 2 \end{pmatrix}$$

Now we can substract the second row once from the third and then in the next step we add the third row to the last and we optain:

$$\rightarrow \quad \begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 0 & 2 & 2 & -2 & 1 \\ 0 & 0 & 5 & 5 & -2 \\ 0 & 0 & -5 & -5 & 2 \end{pmatrix} \quad \rightarrow \quad \left(\begin{array}{ccc|cc} 1 & 3 & 5 & 2 & 0 \\ 0 & 2 & 2 & -2 & 1 \\ 0 & 0 & 5 & 5 & -2 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Now subtracting the 3/2-time of the second row to the first row and subtracting the 2/5-th of the third row to the second row yields

$$\rightarrow \quad \begin{pmatrix} 1 & 0 & 2 & 5 & -3/2 \\ 0 & 2 & 0 & -4 & 9/5 \\ 0 & 0 & 5 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Next we substract the 2/5-th of the third row to the first and then in the last step we multiply the second row by 1/2 and the third row by 1/5. We get

$$\rightarrow \quad \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 3 & -7/10 \\ 0 & 2 & 0 & -4 & 9/5 \\ 0 & 0 & 5 & 5 & -2 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right) \quad \rightarrow \quad \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 3 & -7/10 \\ 0 & 1 & 0 & -2 & 9/10 \\ 0 & 0 & 1 & 1 & -2/5 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Thus we have optained finally the form (21), and this by the way without the need of column transformations. This concludes the example.

Note that if one does not exchange columns than it is still possible to transform any $p \times q$-matrix into a matrix of the following form.

$$\left(\begin{array}{ccccccccccc|c} 0 & \dots & 0 & d_{1,j_1} & & & & & & & & \\ 0 & \dots & \dots & 0 & \dots & 0 & d_{2,j_2} & & & * & & * \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & \dots & d_{3,j_3} & & & \\ \vdots & & & & & & & & & & & \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \quad d_{r,j_r} & \\ \hline & & & & & & 0 & & & & & 0 \end{array}\right) \tag{22}$$

where $1 \leq j_1 < j_2 < \ldots < j_r \leq q$ and the $d_{i,j_i} \neq 0$ for $1 \leq i \leq r$. In particular it is possible to have all $d_{i,j_i} = 1$ by using elementary row transformations of type III.

Now if one applies Proposition 1.5 and the forth following considertations to the simple coefficient matrix C of a general system of linear equations (6), then we see that we can transform this matrix into a matrix of the form (17) or (20) or (21) or (22) (with $0 \leq r \leq m, n$) and this only by row transformations of type I, II or III and column exchanges (in the first three cases). If one applies the same row transformations to the extended coefficient matrix (15) and translates the result to the language of systems of linear equations, then one gets the following result.

Proposition 1.6. *Using elementary transformations of type I, II or III and after renaming the variables one can always transform an arbitrary system of linear equations* (6) *into an equivalent one of the form*

$$\begin{array}{rrrrr} x_1 & & + d_{1,r+1}x_{r+1} + \ldots + d_{1,n}x_n & = & b'_1 \\ & x_2 & + d_{2,r+1}x_{r+1} + \ldots + d_{2,n}x_n & = & b'_2 \\ & \ddots & \vdots & \vdots & \vdots \\ & & x_r + d_{1,r+1}x_{r+1} + \ldots + d_{1,n}x_n & = & b'_r \\ & & & 0 & = b'_{r+1} \\ & & & \vdots & \vdots \\ & & & 0 & = b'_m \end{array} \tag{23}$$

with $0 \leq r \leq m, n$. *In case the initial system of linear equations* (6) *is homogeneous then also the transformed system of linear equations* (23) *is homogeneous, that is the* $b'_i = 0$ *for all* $1 \leq i \leq m$ *and in this case one can leave away the* $m - r$ *last equations of system* (23). □

As an immediate consequence of this proposition we get the following two results.

Proposition 1.7. *A* homogenous *system of linear equations with* $n > m$ *unknown variables (that is more unknown variables then equations) has always non-trivial solutions.*

Proof. We may assume without any loss of generality that the system of linear equations is already given in the form (23). Then $r \leq m < n$. Let $x_{r+1}, \ldots, x_n$ be arbitrary numbers with at least one number different from zero. Then set

$$x_i := -d_{i,r+1}x_{r+1} - \ldots - d_{i,n}x_n$$

for $1 \leq i \leq r$. Then the n-tuple x is by construction a non-trivial solution of the system (23). □

Proposition 1.8. *Assume that we have a nonhomogeneous system of linear equations* (6) *with as many unknown variables as equations (that is* $m = n$*). If the homogeneous part of the system of linear equations has only the trivial solution then the system* (6) *has a unique solution.*

Proof. Since the homogeneous part of the system of linear equations is assumed to have only the trivial solution it means that we can transform the system of linear equations (6) with elementary row transformations of type I, II and III into a system of linear equations of the following form (after possible renaming of the unknown variables):

$$\begin{array}{rrr} x_1 & & = b'_1 \\ & x_2 & = b'_2 \\ & \ddots & \vdots \\ & & x_n = b'_n \end{array}$$

Now this system of linear equations has clearly only a unique solution and since it is equivalent with the given system of linear equations (6) it follows that also the system of linear equations (6) has only a unique solution. □

4. Methodes for Solving Homogeneous and Nonhomogeneous Systems of Linear Equations

In the last section we have introduced in connection with systems of linear equations the methode of elementary transformations of matrices. We used this methodes to develope the so called Gauss algorithm to solve systems of linear equations. In addition to this we could gain some theoretical insight into systems of linear equations, see Proposition 1.7 and 1.8. We shall summarize the calculation methodes for systems of linear equations. Proposition 1.2 and Proposition 1.3 suggest that is usefull to first study homogeneous systems of linear equations.

4.1. Methodes for Solving Homogeneous Systems of Linear Equations. Consider the homogeneous system of linear equations

$$\begin{array}{rcccl} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n = 0 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n = 0 \\ \vdots & & \vdots & \vdots & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n = 0 \end{array} \tag{24}$$

Let us denote by A the (simple) coefficient matrix of (24)

$$A := (a_{ij}). \tag{25}$$

Using elementary row transformations of type I, II and III together with exchanging columns we can transform the matrix A into the $m \times n$-matrix D of the form (21). That is we have then

$$D = \begin{pmatrix} I_r & B \\ 0 & 0 \end{pmatrix} \tag{26}$$

where I_r is the $r \times r$-identity matrix, B is a certain $r \times (n-r)$-matrix and the zeros stand for zero matrices of the appropriate format. Therefore is the system (24) is equivalent (after possible renaming of the unknown variables) to the system

$$\begin{array}{rcccl} x_1 & & + b_{1,1}x_{r+1} + \ldots + b_{1,n-r}x_n = 0 \\ & x_2 & + b_{2,1}x_{r+1} + \ldots + b_{2,n-r}x_n = 0 \\ & \ddots & \vdots & \vdots & \vdots \\ & & x_r + b_{1,1}x_{r+1} + \ldots + b_{1,n-r}x_n = 0 \end{array} \tag{27}$$

If $n = r$, then it is evident that this system has only the trivial solution 0. Thus we may assume that $n > r$. From (27) we see that if we choose arbitrary numbers $x_{r+1}, \ldots, x_n$ and if we then set the numbers $x_1, \ldots, x_r$ according to (27) to

$$x_i := -b_{i,1}x_{r+1} - b_{i,2}x_{r+2} - \ldots - b_{i,n-r}x_n \tag{28}$$

for $i = 1, 2, \ldots, r$, then

$$x := \begin{pmatrix} x_1 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} \tag{29}$$

is a solution of (27). Thus we see from this the following result.

Proposition 1.9. *In the case $n > r$ let $l_1, l_2, \ldots, l_{n-r}$ be the colums of the matrix*

$$\begin{pmatrix} -B \\ I_{n-r} \end{pmatrix} \tag{30}$$

where B is the matrix from (26), that is

$$l_1 := \begin{pmatrix} -b_{1,1} \\ -b_{2,1} \\ \vdots \\ -b_{r,1} \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \quad l_2 := \begin{pmatrix} -b_{1,2} \\ -b_{2,2} \\ \vdots \\ -b_{r,2} \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad l_{n-r} := \begin{pmatrix} -b_{1,n-r} \\ -b_{2,n-r} \\ \vdots \\ -b_{r,n-r} \\ 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{31}$$

Then the following is true: every solution of (27) can be written in an unique way as

$$x = t_1 l_1 + t_2 l_2 + \ldots + t_{n-r} l_{n-r} \tag{32}$$

whith certain numbers $t_1, t_2, \ldots, t_{n-r}$. On the other hand the expression (32) is for any numbers $t_1, t_2, \ldots, t_{n-r}$ a solution of (27).

PROOF. Assume that x as in (29) is a solution of (27). Set

$$t_1 := x_{r+1}, \quad t_2 := x_{r+2}, \quad \ldots, \quad t_{n-r} := x_n. \tag{33}$$

Since x is a solution of (27) we have the relations (28). But these state together with (33) nothing else but the relation between n-tuples as in (32).

Assume now on the other hand, that the n-tuple x in (29) can be written in the form (32). Then necessarily the relation (33) and (28) are satisfied. From the later it follows that x is a sollution of (27), the other relations states the uniqueness of the expression (32). □

Now Proposition 1.9 gives a very satisfying description of the subspace M_0 of all solutions of a homogeneous system of linear equations. If $M_0 \neq \{0\}$ then there exists elements $l_1, \ldots, l_{n-r} \in M_0$ such that every element of M_0 can be written in a unique way as a *linear combination* of the form (32). Such a system of elements $l_1, \ldots, l_{n-r}$ of M_0 is occasionally named a *system of fundamental solutions*. A set $\{l_1, \ldots, l_{n-r}\}$ which is made up of the elements of a system of fundamental solutions is also called a *basis* of M_0.

Example. Consider the following homogeneous system of linear equations

$$\begin{aligned} x_1 + 2x_2 + x_3 + x_4 + x_5 &= 0 \\ -x_1 - 2x_2 - 2x_3 + 2x_4 + x_5 &= 0 \\ 2x_1 - 4x_2 + 3x_3 - x_4 &= 0 \\ x_1 + 2x_2 + 2x_3 - 2x_4 - x_5 &= 0 \end{aligned} \tag{34}$$

with 4 equations in 5 unknown variables. Using elementary transformations its coefficient matrix transforms into the form (26) as follows.

$$\begin{pmatrix} 1 & 2 & 1 & 1 & 1 \\ -1 & -2 & -2 & 2 & 1 \\ 2 & -4 & 3 & -1 & 0 \\ 1 & 2 & 2 & -2 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & 1 & 1 \\ 0 & 0 & -1 & 3 & 2 \\ 0 & 0 & 1 & -3 & -2 \\ 0 & 0 & 1 & -3 & -2 \end{pmatrix} \rightarrow$$

$$\begin{pmatrix} 1 & 1 & 2 & 1 & 1 \\ 0 & -1 & 0 & 3 & 2 \\ 0 & 1 & 0 & -3 & -2 \\ 0 & 1 & 0 & -3 & -2 \end{pmatrix} \rightarrow \left(\begin{array}{cc|ccc} 1 & 1 & 2 & 1 & 1 \\ 0 & -1 & 0 & 3 & 2 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \rightarrow$$

$$\left(\begin{array}{cc|ccc} 1 & 0 & 2 & 4 & 3 \\ 0 & -1 & 0 & 3 & 2 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \rightarrow \left(\begin{array}{cc|ccc} 1 & 0 & 2 & 4 & 3 \\ 0 & 1 & 0 & -3 & -2 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Note that the third matrix is derived from the second one by exchanging the second and third column. If one takes this column exchange into account when applying Proposition 1.9 then one optains the following system

$$l_1 := \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad l_2 := \begin{pmatrix} -4 \\ 0 \\ 3 \\ 1 \\ 0 \end{pmatrix}, \quad l_3 := \begin{pmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}$$

of fundamental solutions for the homogeneous system of linear equations (34). This concludes the example.

4.2. Methodes for Solving Nonhomogeneous Systems of Linear Equations. We begin with a nonhomogeneous system of linear equations as in (6). Using elementary row transformation of type I, II and III and possible column exchanges we can bring the simple coefficient matrix of (6) into the form (26). Then one performs the same row transformations on the extended coefficient matrix of (6). As the last column one obtains

$$\begin{pmatrix} b'_1 \\ \vdots \\ b'_m \end{pmatrix}.$$

Thus the system of linear equations (6) – after possible renaming of the unknown variables – is equivalent to the following system of linear equations:

$$\begin{array}{rcl} x_1 & + b_{1,1}x_{r+1} + \ldots + b_{1,n-r}x_n & = b'_1 \\ x_2 & + b_{2,1}x_{r+1} + \ldots + b_{2,n-r}x_n & = b'_2 \\ \ddots & \vdots & \vdots \quad \vdots \\ & x_r + b_{1,1}x_{r+1} + \ldots + b_{1,n-r}x_n & = b'_r \\ & 0 & = b'_{r+1} \\ & \vdots & \vdots \\ & 0 & = b'_m \end{array} \tag{35}$$

Now this system of linear equations is exactly then solveable (and therefore also the system (6)) if

$$b'_{r+1} = b'_{r+2} = \ldots = b'_m = 0$$

holds.[4] In this case one can set $x_{r_+1} = \ldots = x_n = 0$ and one sees that the specaial n-tuple

$$x' := \begin{pmatrix} b'_1 \\ \vdots \\ b'_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is a sollution to (35). We have then for the set of all sollutions M of (6) that

$$M = x' + M_0$$

where M_0 is the solution space of the homogeneous system of linear equations associated with (6). Thus we have due to Proposition 1.9 the following result.

Proposition 1.10. *Assume that the nonhomogeneous system of linear equations* (6) *is solveable and that x' is some solution of* (6). *If $\{l_1, \ldots, l_{n-r}\}$ is a basis of the subspace M_0 of all solutions of the hogogeneous system of linear equations associated with* (6), *then any solution x of* (6) *can be written in a unique way in the form*

$$x = x' + t_1 l_1 + \ldots + t_{n-r} l_{n-r}$$

whith numbers $t_1, \ldots, t_{n-r}$. And vice versa, any such expression is a solution of (6). □

Example. We consider the following system of linear equations:

$$\begin{array}{rcl} x_1 + 3x_2 + 5x_3 + 2x_4 & = & 1 \\ 3x_1 + 9x_2 + 10x_3 + x_4 + 2x_5 & = & 0 \\ 2x_2 + 7x_3 + 3x_4 - x_5 & = & 3 \\ 2x_1 + 8x_2 + 12x_3 + 2x_4 + x_5 & = & 1 \end{array} \tag{36}$$

In the example on page 10 we have already considered the simple coefficient matrix of this system of linear equations. We perform now the same row transformations on the extended coefficient matrix and obtain:

$$\begin{pmatrix} 1 & 3 & 5 & 2 & 0 & 1 \\ 3 & 9 & 10 & 1 & 2 & 0 \\ 0 & 2 & 7 & 3 & -1 & 3 \\ 2 & 8 & 12 & 2 & 1 & 1 \end{pmatrix} \rightarrow \left(\begin{array}{ccc|cc|c} 1 & 3 & 5 & 2 & 0 & 1 \\ 0 & 2 & 2 & -2 & 1 & -1 \\ 0 & 0 & 5 & 5 & -2 & 4 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

Thus we see already at this stage that the system of linear equations (36) is *not* solveable: the last row of this extended coefficient matrix transforms into the equation $0 = 1$. This concludes our example.

Note that the above example shows that a nonhomogeneous system of linear equations with more equations than unknown variables is not always solveable (even though one might not see this on the first sight).

5. Two Problems

In the discussion so far we optained knowledge about the theory of solving systems of linear equations which one can describe as satisfying. After we gained insight in the general form of the solutions (Proposition 1.2 and Proposition 1.3) we have seen that using the Gauss algorithim we have a simple way to determine whether a given system of linear equations (6) is solveable or not. Further we have

[4] Note that in the case $r = m$ the system (6) is always solveable!

seen in the previous section that the Gauss algorithm is a useable tool to obtain all sollutions of a solveable system of linear equations.

But on the other hand our studies give inevitable raise to some theoretical questions which we want to point out at this place: As we have seen we can transform any matrix C using elementary row transformations and possible column exchanges into a matrix of the form (21). Now we might ask about the nature of the number r which appeares in this setting, namely does this number r only depend on the initial matrix C. That is, do we get every time the same number r even if we use different sequences of elementary row transformations and column exchanges to achive the general form (21). Let us write down this problem explicitely.

Problem 1. Is it possible to assign every matrix C in a natural way a number $r \geq 0$ such that this number does not change under an elementary row transformation and under a column exchange and such that a matrix of the form (21) is assigned exactly the number r?

The other question concerns the set of sollutions of a homogeneous system of linear equations. We have seen that the set M_0 of all solutions to a homogeneous system of linear equations (9) is always a subspace of the set of all possible n-tuples F^n. This suggest the following opposing question:

Problem 2. Exist for every subspace U of F^n a homogeneous system of linear equations (9) such that the set of solutions of (9) is exactly U?

It will now come with no surprise that if we want to find an answer to the above problems then we need to extend our theoretical horizont. In the next chapter we will introduce in a systematically way the basic concepts of Linear Algebra, which we have already prepared in this chapter. Towards the end of next the next chapter we will be able to answer these problems.[5]

[5]Problem 1 will be answered on page 45 and Problem 2 will be answered on page 49.

CHAPTER 2

Vector Spaces

1. Fields

In the previous chapter we have constantly spoken of "numbers" without saying anything in particular what kind of numbers we actually deal with. We did not use any particular properties of the numbers but rather only assumed some "well known properties" of the addition and multiplication of numbers as we know them for example from the real numbers $\mathbb{R}$ (which are known from calculus). Indeed, most of Linear Algebra – as for example the the treatment of systems of linear equations – does not depend on any particular properties of the numbers used.

In this section we shall introduce the concept of fields. In a nutshell a *field* F is a set together with two operations – one is called the addition and denoted by "$+$" and the other is called the multiplication and denoted by "$\cdot$" – satisfying a minimum of axioms which will be important for us.[1]

Definition 2.1. A *field* $F = (F, +, \cdot)$ is a triple consisting of a set F and two maps

$$+ \colon F \times F \to F, (x, y) \mapsto x + y$$

and

$$\cdot \colon F \times F \to F, (x, y) \mapsto xy$$

(the first one called *addition*, the latter is called *multiplication*) satisfying the following axioms:

(A1) The addition is *associative*, that is $x + (y + z) = (x + y) + z$ for every $x, y, z \in F$.

(A2) The addition is *commutative*, that is $x + y = y + x$ for every $x, y \in F$.

(A3) There exist exactly one element – which we denote by "0" – for which

$$x + 0 = x \qquad \text{and} \qquad 0 + x = x$$

holds for every $x \in F$. This element is also called the *zero* of F.

(A4) For every $x \in F$ there exists exactly one element in F – which we will denote by $-x$ – such that

$$x + (-x) = 0 \qquad \text{and} \qquad (-x) + x = 0.$$

(M1) The multiplication is associative, that is $x(yz) = (xy)z$ for every $x, y, z \in F$.

(M2) The multiplication is commutative, that is $xy = yx$ for every $x, y \in F$.

(M3) There exist exactly one element different from 0 – which we shall denote by "1" – for which

$$x1 = x \qquad \text{and} \qquad 1x = x$$

holds for every $x \in F$. This element is called the *one* of F.

[1] See Appendix A for the notation used for sets and maps.

(M4) For every $x \neq 0$ of F there exists exactly one element in F – which we shall denote by "$1/x$" (or alternatively denoted by "x^{-1}") – such that

$$x(1/x) = 1 \qquad \text{and} \qquad (1/x)x = 1.$$

(D) The addition and multiplication are bound by the *distributive law*, that is

$$(x+y)z = xz + yz \qquad \text{and} \qquad x(y+z) = xy + xz$$

for every $x, y, z \in F$.

Note that we could do with a smaller list of requirements. For example since the addition is required to be commutative (A2) it follows from $x + 0 = x$ that necessarily $0 + x = x$, too. Also the uniqueness requirements in (A3), (A4), (M3) and (M4) could have been left away.

Note further that in the language of Algebra the axioms (A1) to (A4) mean, that F is an *abelian group* under the addition, and the axioms (M1) to (M4) mean that $F^\bullet := F \setminus \{0\}$ is an abelian group under the multiplication. This terminology is not of further importance here, but it will reappear Section 11 of Chapter 3. There we will give a more detailed definition.

From the field axioms one can easily other known calculation rules, for example

$$x0 = 0 \tag{37}$$

$$(-x)y = -(xy) \tag{38}$$

$$xy = 0 \quad \Rightarrow \quad x = 0 \text{ or } y = 0 \tag{39}$$

$$x/y + u/v = (xv + yu)/yv \quad \text{for } y \neq 0 \text{ and } v \neq 0 \tag{40}$$

where $x/y := x(1/y) = xy^{-1}$. The proof of these calculation rules are left as an exercise.

Examples. Well known examples of fields are:

(1) The field of *rational numbers*

$$\mathbb{Q} := \left\{ \frac{r}{s} : r \in \mathbb{Z} \text{ and } s \in \mathbb{N}^+ \right\}.$$

(2) The field of *real numbers* $\mathbb{R}$.

(3) The field of *complex numbers*

$$\mathbb{C} := \{a + ib : a, b \in \mathbb{R}\}$$

where i denotes the *imaginary unit*.

If F is a field and $F' \subset F$ is a subset such that F' is a field under the *same* addition and multiplication as F, then F' is said to be a *subfield* of F. For example

$$\mathbb{Q}(\sqrt{3}) := \{x + y\sqrt{3} : x, y \in \mathbb{Q}\}$$

is a subfield of $\mathbb{R}$.

Let F be an arbitrary field. If $x \in F$ and $k > 0$ a positive integer, then we shall define

$$kx := \underbrace{x + \ldots + x}_{k \text{ times}}.$$

If $k = 0$ we define $kx := 0$ and in the case that $k < 0$ is a negative integer we define $kx := -(-k)x$.

Now observe that the field axioms do *not* exclude the possibility that for the field F holds

$$k \cdot 1 \neq 0 \tag{41}$$

for every integer $k \neq 0$.[2] If F is a field where $k \cdot 1 = 0$ is true for some integer $k > 0$ then we say that F has *positive characteristic*. More precisely we define the chararteristic of a field as follows.

Definition 2.2 (Characteristic of a Field)**.** Let F be a field. If there exists a smallest positive integer $k > 0$ such that

$$k \cdot 1 = 0$$

then we say that F has the *characteristic* k, in symbols $\mathrm{char}(F) := k$. Otherwise we say that F has the characteristic 0, that is $\mathrm{char}(F) := 0$.

Most of the time the characteristic of a field does not affect the results in what follows. It is still important to remember that as long as a field does not have characteristic 0 we cannot be sure that $k \cdot x \neq 0$ is always true for non-zero $x \in F$ and non-zero $k \in \mathbb{Z}$. In Appendix B examples for fileds with positive characteristic are presented.

We shall conclude this section about fields with a related definition. In Linear Algebra (and other fields of mathematics) one encounters mathematical objects which satisfy nearly all requirements of a field. For example the set of integers $\mathbb{Z}$ with the well known addition and multiplication satisfies all field axioms except (M4). In Chapter 3 we will encounter mathematical objects which satisfy even less of the filed axioms, namely all field axioms except (M2) and (M4).[3] This observation motivates the following definition.

Definition 2.3 (Ring)**.** A *ring (with unit)* is a triple $R = (R, +, \cdot)$ satisfying all field axioms except (M2) and (M4). If R is a ring which satisfies $(M2)$ then R is said to be *commutativ*. A ring which satisfies (39) is called *regular*.

Example. The set of integers

$$\mathbb{Z} := \{0, \pm 1, \pm 2, \pm 3, \ldots\}$$

is a commutative and regular ring under the usual addition and multiplication. It is called the *ring of integers*.

2. Vector Spaces

We are now ready to define the central algebraic object of Linear Algebra: vector spaces over a field F. Roughly spoken a vector space over a field F is a set V where we can compute the sum of any two elements and where we can multiply any element by elements of the F such that certain "natural rules" are satisfied. The precise definition of a vector space is the following.

Definition 2.4 (Vector Space)**.** Let F be a field. A *vector space over the field* F (or in short F*-vector space*) is a triple $V = (V, +, \cdot)$ consisting of a set V and two maps

$$+ \colon V \times V \to V, (x, y) \mapsto x + y$$

[2]Note that this does not contradict with (39) since there we assumed that $x, y \in F$ whereas in (41) we have $k \in \mathbb{Z}$ and $1 \in F$. We are used to that we can consider the integers $\mathbb{Z}$ as a subset of the field of rational, real or complex numbers. But in general we cannot assume that we can consider the integers as a subset of an arbitrary field.

[3]Amongst other these are the endomorphism ring $\mathrm{End}_F(V)$ and the full matrix ring $M_n(F)$ of degree n over F.

and

$$\cdot\colon F \times V \to V, (a, x) \mapsto ax$$

(the first one is called *addition*, the latter is called *scalar multiplication*) satisfying the following *vector space axioms:*

(A1) The addition is associative, that is $x + (y + z) = (x + y) + z$ for all $x, y, z \in V$.

(A2) The addition is commutative, that is $x + y = y + x$ for every $x, y \in V$.

(A3) There exists exactly one element in V – denoted by "0" – such that

$$x + 0 = x$$

for every $x \in V$.

(A4) For every $x \in V$ there exists exactly one element – denoted by "$-x$" – such that

$$x + (-x) = 0.$$

(SM1) $(ab)x = a(bx)$ for every $a, b \in F$ and $x \in V$.

(SM2) $1x = x$ for every $x \in V$.

(SM3) $a(x + y) = ax + ay$ for every $a \in F$ and $x, y \in V$.

(SM4) $(a + b)x = ax + bx$ for every $a, b \in F$ and $x \in V$.

Note that the above axioms (A1) to (A4) are structural exactly the same as the axioms (A1) to (A4) in Definition 2.1 of a field in the previous section. Again this means in the language of Algebra that V together with the addition is an abelian group.

Elements of the vector space V are called *vectors*. Note that we use the same symbol "0" to denote the zero of the field F and to denote the identity element of the addition in V. The identity elemnt of the addition in V is also called the *zero vector*. A vector space is sometimes also called a *linear space*.

From the vector space axioms follow esaily further calculation rules, for example

$$0x = 0 \text{ and } a0 = 0 \tag{42}$$

$$(-a)x = -ax = a(-x) \text{ and in particular } (-1)x = -x \tag{43}$$

$$ax = 0 \quad \Rightarrow \quad a = 0 \text{ or } x = 0 \tag{44}$$

$$a(x - y) = ax - ay \tag{45}$$

where $a \in F$, $x, y \in V$ and $x - y := x + (-y)$. The verification of these rules is left as an exercise. Note that the calculation rule (42) ensures that the use of the symbol "0" for both the zero element of the field F and the zero vector of the vector space V will not cause confusion.

Examples. (1) In the previous chapter we have already introduced the set F^n all n-tuples

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad x_i \in F$$

of a given field F. Together with the component wise addition and the scalar multiplication as defined in Section 2 of the previous chapter this set becomes a F-vector space.

(2) Consider the set $C^0(I)$ of all continuous maps $f\colon I \to \mathbb{R}$ from the interval $I \subset \mathbb{R}$ with values in $\mathbb{R}$. This set becomes in a natural way a $\mathbb{R}$ vector space.

(3) Let V be a vector space over the the field F. Then V is also a vector space over any subfield $F' \subset F$.
(4) Let E be a field and F a subfield of E. Then E is in a natural way a vector space over F. For example the field of real numbers $\mathbb{R}$ is in a natural way a $\mathbb{Q}$-vector space.
(5) As a generalisation of the first example consider an arbitrary set I. Then F^I is defined to be the set of all maps from the set I into the field F, that is

$$F^I := \{f : f \text{ is a map from } I \text{ to } F\}.$$

The set F^I becomes in a natural way an F-vector space: If x and y are elements of F^I, that is $x\colon I \to F$ and $y\colon I \to F$ are maps, then the sum of x and y is the map $x + y\colon I \to K$ defined by

$$(x + y)(i) := x(i) + y(i), \qquad i \in I, \tag{46}$$

and the scalar product of x with an element $a \in F$ is the map $ax\colon I \to F$ defined by

$$(ax)(i) := ax(i). \tag{47}$$

In particular, if we set $I := \{1, 2, \ldots, n\}$ then F^I is essentialy the same vector space as F^n in the first example.[4]

Definition 2.5. (Subspace of a Vector Space) Let V be a vector space over the field F and let $V' \subset V$ be a subset. If V' is a vector space under the addition and scalar multiplication of V, then V' is called a *(linear) subspace* of V.

Proposition 2.6. *(Subspace Criterion) A* non-empty *subset $U \subset V$ of a F-vector space V is a linear subspace of V if and only if the following two conditions are satisfied:*

$$x, y \in U \quad \Rightarrow \quad x + y \in U \tag{48}$$

$$a \in F, x \in U \quad \Rightarrow \quad ax \in U \tag{49}$$

PROOF. If U is a subspace of V, then of course (48) and (49) are satisfied. Thus assume that U is a non-empty subset of V satisfying the conditons (48) and (49). We need only to verify the vector space axioms (A3) and (A4). Let $x \in U$. Then $-x = (-1)x \in U$ by (43) and (49). Thus $(A4)$ is satisfied. Now since U is assumed to be not empty it follows that there exists an $x \in U$. Thus also $-x \in U$ due to (A4) and therefore $0 = x + (-x) \in U$ due to (48). □

Note the *essential* (!) difference between the definition of a linear subspace (Definition 2.5) and the Subspace Criterion (Proposition 2.6). The first one *defines* what a linear subspace is, the latter gives a criterion to *decide more easily* whether a subset $U \subset V$ of a vector space V is a linear subspace (according to the Definition 2.5) or not.

Examples. (1) Every vector space V has the *trivial subspace* $\{0\} \subset V$. A vector space which consists only of the zero vector 0 is also called the *zero space*. By abuse of notation we denote the zero space with the symbol "0", that is $0 := \{0\}$. Again there is no danger of confusion of the zero space with the zero element of the field F or with the zero vector of the vector space V if the reader is awake.

[4] We will later define precisely what we mean exactly by the term "essentialy the same" when we introduce in Section 2 in the next chapter the concept of isomorphisms and isomorphic vector spaces.

(2) The set of solutions M_0 of a homogeneous system of linear equations (9) is a subspace of F^n (see Proposition 1.3 in the previous chapter).
(3) The $\mathbb{R}$-vector space $C^0(I)$ is a subspace of $\mathbb{R}^I$.
(4) Let I be an arbitrary set and consider the F-vector space F^I from the previous example set. Denote by $F^{(I)}$ the subset of F^I consisting of all functions $f\colon I \to F$ such that

$$f(x) = 0 \qquad \text{for all } x \in I \text{ but finite many exceptions.}$$

Then $F^{(I)}$ is a linear subspace of F^I and $F^{(I)} \neq F^I$ if and only if I is not a finite set.

Note that the vector spaces of the form $F^{(I)}$ are of principal importance. Every F-vector space is of the type $F^{(I)}$ for a suitable set I.[5] This is the fundamental theorem of the theory of vector spaces and we will later return to this subject.

If U and W are two subsets of a vector space V, then we denote by $U+W$ the *sum* of U and W the set

$$U + W := \{u + w : u \in U \text{ and } w \in W\}.$$

Proposition 2.7. *Let V be a vector space over F and let U and W be two linear subspaces of V. Then both the intersection $U \cap W$ and the sum $U + W$ are linear subspaces of V.*

Proof. This is verified directly using the Subspace Criterion from Proposition 2.6. □

3. Linear Combinations and Basis of a Vector Space

Definition 2.8 (Linear Combination)**.** Let V be a F-vector space and $u_1, \ldots, u_m$ some vectors of V. Then we say that $v \in V$ is a *linear combination of the vectors* $u_1, \ldots, u_m$ if there exists elements $a_1, \ldots, a_m \in F$ such that

$$v = a_1 u_1 + \ldots + a_m u_m. \tag{50}$$

If $M \neq \emptyset$ is an arbitrary subset of V, then we say that $v \in V$ is a *linear combination of vectors in M* if there exists some vectors $u_1, \ldots, u_m \in M$ such that v is a linear combination of $u_1, \ldots, u_m$.

Note that a linear combination is always a *finite* sum! In Linear Algebra we do not consider infinite sums. Note also that the zero vector 0 is always in a trival way a linear combination of any collection $u_1, \ldots, u_m$, namely

$$0 = 0u_1 + \ldots + 0u_m.$$

Definition 2.9. Let M be an arbitrary subset of the F-vector space V. Then the *linear hull* or *span* of M in V is the set

$$\operatorname{span} M := \{v \in V : v \text{ is a linear combination of vectors in } M\}. \tag{51}$$

If $M = \emptyset$ we define

$$\operatorname{span} \emptyset := \{0\} \tag{52}$$

to be the trivial zero vector space 0.

[5] If one considers $F^{(I)}$ for $I = \emptyset$ to be the null vector space 0.

In the case that $M \neq \emptyset$ then $v \in \operatorname{span} M$ means that there exists some $u_1, \ldots, u_m$ such that v can be expressed as a sum as in (50). If the u_i are not pairwise different vectors, then we can always transform the linear combination into one with pairwise different vectors $u_i \in M$.

Using the vector space $F^{(M)}$ we can characterize the elements $v \in \operatorname{span} M$ in the following way.

Proposition 2.10. *Let M be a non-empty subset of the F-vector space V. Then $v \in V$ is an element of* $\operatorname{span} M$ *if and only if there exists a $x \in F^{(M)}$ such that*

$$v = \sum_{u \in M}{}' x(u)u. \tag{53}$$

Here the symbol $\sum'$ is used to denote that the sum in (53) is actual a *finite* sum (even though the set M might be infinite).

Proof of Proposition 2.10. "$\Rightarrow$": Assume that v is a vector of $\operatorname{span} M$. Then there exists vectors $u_1, \ldots u_m \in M$ and elements $a_1, \ldots, a_m \in F$ such that $v = a_1u_1 + \ldots + a_mu_m$. Then define a map $x\colon M \to F$ by

$$x(u) := \begin{cases} a_i & \text{if } u = u_i \text{ for some } 1 \leq i \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

Then we have by construction $x \in F^{(M)}$ and

$$\sum_{u \in M}{}' x(u)u = x(u_1)u_1 + \ldots + x(u_n)u_n = a_1u_1 + \ldots + a_nu_n = v.$$

"$\Leftarrow$": On the other hand, if there exists a $x \in F^{(M)}$ such that (53), then v is clearly a linear combination of some vectors of M, that is $v \in \operatorname{span} M$. □

Note that the notation in (53) is a very practical notation. If now $x, x' \in F^{(M)}$ and $a \in F$, then we have the following equations

$$\sum_{u \in M}{}' x(u)u + \sum_{u \in M}{}' x'(u)u = \sum_{u \in M}{}' (x + x')(u)u \tag{54}$$

and

$$a \sum_{u \in M}{}' x(u)u = \sum_{u \in M}{}' (ax)(u)u \tag{55}$$

due to the definition of the addition and scalar multiplication in (46) and (47) for vectors of the vector space F^M and thus also for vectors of the vector space $F^{(M)}$. Now using Proposition 2.10 the above two equations translate into the two observations:

$$v, v' \in \operatorname{span} M \quad \Rightarrow \quad v + v' \in \operatorname{span} M$$

and

$$v \in \operatorname{span} M, a \in F \quad \Rightarrow \quad av \in \operatorname{span} M.$$

Since $\operatorname{span} M$ is always non-empty (it contains the zero vector) we conclude using the Subspace Criterion 2.6 the following result.

Proposition 2.11. *Let M be an arbitrary subset of the F-vector space V. Then* $\operatorname{span} M$ *is linear subspace of V.* □

If M is an arbitrary subset of V. We claim that $M \subset \operatorname{span} M$. In the case that $M = \emptyset$ this is clear. Thus let $M \neq \emptyset$. If $v \in M$ then $v = 1v \in \operatorname{span} M$ and thus $M \subset \operatorname{span} M$ also in this case. On the other hand, if W is an arbitrary linear subspace of V with $M \subset W$, then clearly $\operatorname{span} M \subset W$. Let U be another linear subspace which has the above property of $\operatorname{span} M$, that is $M \subset U$ and U is containied in any linear subspace W which contains M. Then $U \subset \operatorname{span} M$ and on the other hand also $\operatorname{span} M \subset U$. That is, $U = \operatorname{span} M$. Thus we have shown the following usefull characterisation of the linear hull $\operatorname{span} M$ of M.

Proposition 2.12. *Let M be a subset of the vector space V. Then there exists a unique linear subspace U of V such that $M \subset U$ and having the property that*

$$\text{if } W \text{ is a linear subspace of } V \text{ with } M \subset W \text{ also then } U \subset W. \tag{56}$$

And we have precisely $U = \operatorname{span} M$. □

In short the above result means that the linear hull of M is precisely the smalles linear subspace of V containing M.

We shall collect a few easy to verify properties of the linear hull:

$$M \subset \operatorname{span} M \tag{57}$$

$$M \subset M' \quad \Rightarrow \quad \operatorname{span} M \subset \operatorname{span} M' \tag{58}$$

$$M = \operatorname{span} M \quad \Longleftrightarrow \quad M \text{ is a linear subspace of } V \tag{59}$$

$$\operatorname{span}(\operatorname{span} M) = \operatorname{span} M \tag{60}$$

$$\operatorname{span}(M \cup M') = \operatorname{span} M + \operatorname{span} M' \tag{61}$$

The first property we have already verified above, the proof of the remaining properties is left as an exercise.

Examples. (1) Let $V := C^0(\mathbb{R})$ be the $\mathbb{R}$-vector space of all continuous functions $f\colon \mathbb{R} \to \mathbb{R}$. Then the functions exp, sin and cos are vectors of the vector space V. We shall show that exp is not contained in the linear hull of $M := \{\sin, \cos\}$. We show this by assuming towards a contradiction that $\exp \in \operatorname{span}\{\sin, \cos\}$. By definition this means that there exists numbers $a_1, a_2 \in \mathbb{R}$ such that $\exp = a_1 \sin + a_2 \cos$. In other words this means

$$\exp(x) = a_1 \sin(x) + a_2 \cos(x) \qquad \text{for all } x \in \mathbb{R}.$$

But this would imply that $0 < \exp(0) = a_1 \sin(0) + a_2 \cos(0) = 0 + a_2 = a_2$ and on the other hand $0 < \exp(\pi) = a_1 \sin(\pi) + a_2 \cos(\pi) = 0 - a_2 = -a_2$, that is $a_2 > 0$ and at the same time $a_2 < 0$ which is a contradiction. Thus the assumption that exp is a linear combination of the functions sin and cos is shown to be wrong and it follows that

$$\exp \notin \operatorname{span}\{\sin, \cos\}.$$

(2) Recall the setting and notation of the frist chapter. Consider a system of m linear equations with n unknown variables over a field F. Denote by $v_1, \ldots, v_n \in F^m$ the columns of the simple coefficient matrix associated with the system of linear equations. Then the system of linear equations is apparently solveable if and only if

$$b \in \operatorname{span}\{v_1, \ldots, v_n\}$$

where b denotes the right most column of the extended coefficient associated with the system of linear equations.

Moreover the associated homogeneous system of linear equations has a non-trivial solution if and only if the zero vector $0 \in F^m$ is a *non-trivial* linear combination of the vectors $v_1, \ldots, v_n$.

The meaning of 0 being a non-trivial linear combination of the vectors $v_1, \ldots, v_n$ in the above example is given by the following definition.

Definition 2.13. Let $v_1, \ldots, v_n$ be some (not necessarily pairwise distinct) vectors of the F-vector space V. Then 0 is a *non-trivial linear combination* of the vectors $v_1, \ldots, v_n$ if there exist elements $a_1, \ldots, a_n \in F$ such that

$$a_1 v_1 + \ldots + a_n v_n = 0$$

and $a_i \neq 0$ for at least one $1 \leq i \leq n$.

Note that the above definition is of essential importance! This is already suggested by the close relation with homogeneous systems of linear equations which did lead to the definition.

Example. Consider the $\mathbb{R}$-vector space $V = \mathbb{R}^4$. Then 0 is a non-trivial linear combination of the three vectors

$$v_1 := \begin{pmatrix} -1 \\ -1 \\ 3 \\ 4 \end{pmatrix}, \quad v_2 := \begin{pmatrix} 4 \\ 10 \\ 3 \\ 5 \end{pmatrix} \quad \text{and} \quad v_3 := \begin{pmatrix} 3 \\ 7 \\ 1 \\ 2 \end{pmatrix}.$$

For example $v_1 - 2v_2 + 3v_3 = 0$.

We shall introduce one more essential concept:

Definition 2.14 (Basis of a Vector Space)**.** A subset $M \subset V$ of a vector space V is said to be a *basis* of V, if every vector $v \in V$ can be written in a unique way as a linear combination of vectors of M.

We say that a basis M is *finite* if the set M contains only finitely many elements.

The above definition is equivalent with: M is a basis of V if and only if for every $v \in V$ there exists exactly one $x \in F^{(M)}$ such that

$$v = \sum_{u \in M}{}' x(u)u \tag{62}$$

Note that the above definition of a basis contains *two* properties:

(1) We have that $V = \operatorname{span} M$. That is every vector $v \in V$ *can* be written as a linear combination of vectors of M.
(2) For every vector there exists *only one* way to write v as in (62) as a linear combination of the vectors in M. That is if $x, y \in F^{(M)}$ are two elements such that
$$v = \sum_{u \in M}{}' x(u)u = \sum_{u \in M}{}' y(u)u$$
then necessarily $x = y$, that is $x(u) = y(u)$ for every $u \in M$.

Examples. (1) A finite subset $M \subset V$ which consist of n elements, say

$$M = \{u_1, \ldots, u_n\},$$

is precisely then a basis of the F-vector space V if for ever vector $v \in V$ there exists a unique solution to the system of linear equations

$$x_1 u_1 + \ldots + x_n u_n = v. \tag{63}$$

In the previous chapter we have seen that the uniqueness requirement is equivalent with the homogeneous part of (63),

$$x_1u_1 + \ldots + x_nu_n = 0$$

having only the trivial solution (Proposition 1.8). And this is again equivalent with that there exists *no* non-trivial linear combination of the zero vector 0.

Thus a finite subset M is a basis of V if and only if $V = \operatorname{span} M$ and there exists no non-trivial linear combination of the zero vector 0 by vectors of M.

(2) Consider the vector space $V = F^2$. Then every subset

$$\left\{ \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right\} \subset V \tag{64}$$

where $ad - bc \neq 0$ is a basis of V.

(3) Consider the vector space F^n. Then the set consisting of the *canonical unit vectors*

$$e_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \quad e_n := \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{65}$$

forms a basis of F^n, the *standard basis of* F^n. The proof is left as an exercise.

(4) More general, if I is an arbitrary non-empty set, then the elements $e_i \in F^{(I)}$ defined by

$$e_i(j) := \begin{cases} 0 & \text{for } j \neq i, \\ 1 & \text{for } j = i. \end{cases} \tag{66}$$

for every $i \in I$ is a basis of the vector space $F^{(I)}$. This basis is called the *standard basis of* $F^{(I)}$. It follows that

$$x = \sum_{i \in I}{}' x(i)e_i \tag{67}$$

for every $x \in F^{(I)}$.

(5) Let V be a vector space which has a basis $M \neq \emptyset$, then V cannot be the zero vector space $\{0\}$. Because otherwise $0 = 1u = 0u$ for some $u \in M$ and this would contradict to the uniqueness requirement in the definition of a basis. This motivates the *agreement* that the empty set $\emptyset$ is the (only) basis of the zero vector space.

4. Linear Dependence and Existence of a Basis

In this section we will study the question whether a given vector space will have a basis or not. The outcome will be that *every* vector space has a basis. Throughout this section V will be a vector space over a fixed field F.

Definition 2.15. We say that V is *generated* (or *spanned*) by the subset $M \subset V$ if

$$V = \operatorname{span} M.$$

In this case M is called a *(linear) generating system* of V. We say that V is *finitely generated* if there exists a generating system M of V which consists of only finite many elements.

Examples. (1) Every vector space V has a generating system since trivialy $V = \operatorname{span} V$.

(2) The vector space F^n is finitely generated. A finite generating system is for example given by the canonical standard basis

$$\{e_1, \ldots, e_n\}.$$

(3) The empty set is a generating system of the zero vector space.

(4) The vectors

$$\begin{pmatrix}1\\2\end{pmatrix}, \quad \begin{pmatrix}3\\4\end{pmatrix}, \quad \begin{pmatrix}5\\6\end{pmatrix}$$

gives a generating system of $\mathbb{R}^2$, see the example (2) on page 28.

(5) The five vectors

$$\begin{pmatrix}1\\3\\0\\2\end{pmatrix}, \quad \begin{pmatrix}3\\9\\2\\8\end{pmatrix}, \quad \begin{pmatrix}5\\10\\7\\12\end{pmatrix}, \quad \begin{pmatrix}2\\1\\3\\2\end{pmatrix}, \quad \begin{pmatrix}0\\2\\-1\\1\end{pmatrix}$$

is not a generating system of the vector space $\mathbb{R}^4$, compare with the example on page 10.

(6) If I is an infinite set, then the vector space $F^{(I)}$ is not finitely generated.

(7) The vector space $C^0(\mathbb{R})$ of all continuous functions from $\mathbb{R}$ to $\mathbb{R}$ is not finitely generated.

We are now interested in "small" generating systems in the sense that the generating system does not contain any vectors which are not necesarily needed to generate the vector space. This leads to the following definition.

Definition 2.16 (Linear Dependence)**.** A subset $M \subset V$ of the vector space V is said to be a *linear independent* subset of V if for every $u \in M$ the linear subspace generated by $M \setminus \{u\}$ is a proper subspace of $\operatorname{span} M$. That is, for every $u \in M$ holds

$$\operatorname{span}(M \setminus \{u\}) \neq \operatorname{span} M.$$

A subset $M \subset V$ is said to be a *linear dependent* subset of V if it is not a linear independent set of V.

Note that if M is a linear dependent subset of V then this means that there exists at least one vector $u \in M$ such that $\operatorname{span}(M \setminus \{u\}) = \operatorname{span} M$.

Examples. (1) The empty set $\emptyset$ is always a linear independent subset of V.

(2) If $M \subset V$ conatins the zero vector 0 then M is always a linear depenedent subset of V.

(3) The vectors

$$\begin{pmatrix}1\\2\end{pmatrix}, \quad \begin{pmatrix}3\\4\end{pmatrix}, \quad \begin{pmatrix}5\\6\end{pmatrix}$$

form a linear dependent subset of $\mathbb{R}^2$ since any two of those vectors generate already $\mathbb{R}^2$.

(4) Assume that $\operatorname{char} F \neq 2$ (that is $1 + 1 \neq 0$ in F). And let $u, v \in V$ be two non-zero vectors of V. Then the set $\{u, u + v, u - v\}$ is always linear dependent.

Lemma 2.17. *Let $M \subset V$ be a subset. Let $v \in \operatorname{span} M$ such that $v \notin M$. Then $M \cup \{v\}$ is a linear dependent subset of V.*

PROOF. Since $M \subset M \cup \{v\}$ we have that $\operatorname{span} M \subset \operatorname{span}(M \cup \{v\})$. On the other hand, since $v \in \operatorname{span} M$ we have that $M \cup \{v\} \subset \operatorname{span} M$. But then also $\operatorname{span}(M \cup \{v\}) \subset \operatorname{span}(\operatorname{span} M) = \operatorname{span} M$. Thus we have the inclusion in both directions and therefore $\operatorname{span} M = \operatorname{span}(M \cup \{v\})$. Now since we assumed that $v \notin M$ we have $(M \cup \{v\}) \setminus \{v\} = M$. Therefore $\operatorname{span}((M \cup \{v\}) \setminus \{v\}) = \operatorname{span}(M \cup \{v\})$ and this shows that $M \cup \{v\}$ is a linear dependent subset of V. □

Proposition 2.18 (Characterisation of the Linear Dependency)**.** *A subset $M \subset V$ of a vector space V is linear depenedent if and only if there exists finitely many pairwise distinct vectors $v_1, \ldots, v_m \in M$ such that 0 is a non-trivial linear combination of the vectors $v_1, \ldots, v_m$.*

PROOF. "⇒": Assume that M is a linear dependent subset of V. Then there exists a $v_1 \in M$ such that $\operatorname{span} M = \operatorname{span}(M \setminus \{v_1\})$. Thus there exist pairwise distinct vectors $v_2, \ldots, v_m \in M \setminus \{v_1\}$ such that $v_1 = a_2 v_2 + \ldots + a_m v_m$ for some numbers $a_2, \ldots, a_m \in F$. But this means that

$$0 = 1v_1 - a_2 v_2 - \ldots - a_m v_m$$

is a non-trivial linear combinations of the zero vector 0 of pairwise distinct vectors $v_1, \ldots, v_m$.

"⇐": Assume that there exists pairwise distinct vectors $v_1, \ldots, v_m \in M$ such that the zero vector is a non-trivial linear combination of the these vectors, say

$$0 = a_1 v_1 + \ldots + a_m v_m$$

with some number $a_k \neq 0$. Without any loss we may assume that $a_1 \neq 0$ and in particular we may assume with out any loss of generality that $a_1 = -1$. Then

$$v_1 = a_2 v_2 + \ldots + a_m v_m \in \operatorname{span}(M \setminus \{v_1\}).$$

Thus $M \subset \operatorname{span}(M \setminus \{v_1\})$ and therefore $\operatorname{span} M \subset \operatorname{span}(M \setminus \{v_1\})$. On the other hand $\operatorname{span}(M \setminus \{v_1\}) \subset \operatorname{span} M$ since $M \setminus \{v_1\} \subset M$. Therefore we have the inclusion in both directions and this shows that we have the equality $\operatorname{span} M = \operatorname{span}(M \setminus \{v_1\})$ for $v_1 \in M$. Therefore M is a linear dependent subset of V. □

Now using the fact that a M is by definition a linear independent subset of V if and only if M is not a linear dependent subset of V one can use the previous proposition to formulate a characterisation of a linear indepenedent subset of V.

Proposition 2.19 (Characterisation of the Linear Independency)**.** *A subset $M \subset V$ of the vector space V is a linear indepened subset of V if and only if for every finitely many, pairwise distinct vectors $v_1, \ldots, v_m \in M$ it follows from*

$$a_1 v_1 + \ldots + a_m v_m = 0$$

that $a_1 = \ldots = a_m = 0$. □

Note that the above proposition is often used in literature to *define* linear independence. Further it follows from the same proposition that if $M' \subset M$ is a subset of a linear independent subset M of V, then M' is also a linearly indepenedent subset of V.

Definition 2.20. Let $M \subset V$ be a linear indepenedent subset of V. Then we say that M is a *maximal* linear indepenedent subset of $M' \subset V$ if for every $u \in M' \setminus M$ the set $M \cup \{u\}$ is linear dependent.

In particular is M a maximal linear independent set of V if for every $u \in V \setminus M$ the set is $M \cup \{u\}$ is linear dependent.

Lemma 2.21. *Let $M \subset V$ be a linear independent subset of V.*

(1) *If $v \in V$ is a vector such that $M \cup \{v\}$ is a linear depenendent set, then $v \in \operatorname{span} M$.*
(2) *If M is a maximal linear independent subset of $M' \subset V$, then $M' \subset \operatorname{span} M$. (In particular $\operatorname{span} M' = \operatorname{span} M$.)*

PROOF. (1) Assume that M is linear independent and $M \cup \{v\}$ is linear dependent. It follows from Proposition 2.18 that there exists finitely many pairwise distinct vectors $v_1, \ldots, v_m \in M$ and numbers $a, a_1, \ldots, a_m \in F$ (not all equal to zero) such that

$$av + a_1 v_1 + \ldots + a_m v_m = 0. \tag{68}$$

If $a = 0$ then it follow by Proposition 2.18 that M is linear dependend, but this is silly. Thus $a \neq 0$ and we can multibly (68) by a^{-1} and get

$$v = (-a_1 a^{-1})v_1 + \ldots + (-a_1 a^{-1})v_m.$$

Thus $v \in \operatorname{span} M$ as claimed.

(2) Let $u \in M'$ be an arbitrary element. If $u \in M$ then clearly $u \in \operatorname{span} M$. On the other hand, if $u \notin M$, then $M \cup \{u\}$ is linear dependend (due to the maximality of M) and thus by the previous result it follows that $u \in \operatorname{span} M$, too. Alltogether this means that $M' \subset \operatorname{span} M$. □

We are now ready to characterize the property of a set $M \subset V$ to be a basis of the vector space V in several different ways.

Proposition 2.22 (Characterisation of a Basis)**.** *Let $M \subset V$ be a subset of the F-vector space V. Then the following statements are equivalent:*

(1) *M is a basis of V.*
(2) *M is a minimal generating system of V.*
(3) *M generates V and M is linearly independent.*
(4) *M is a maximal linear independent subset of V.*
(5) *M is a maximal linear independent subset of* every *generating system E of V which contains M.*
(6) *There* exists *a generating system E of V which contains M as a maximal linear independent subset of E.*
(7) *There exists a bijective[6] map $f\colon F^{(M)} \to V$ such that:*
   (a) *$f(x+y) = f(x) + f(y)$ and $f(ax) = af(x)$ for all $x, y \in F^{(M)}$ and $a \in F$,*
   (b) *$f(e_u) = u$ for every $u \in M$ were $\{e_u : u \in M\}$ is the standard basis of $F^{(M)}$ as defined in the example on page 28.*

PROOF. We will follow the following deduction scheme in our proof:

$$\begin{array}{ccccc} & & (2) & & \\ & & \Updownarrow & & \\ (1) & \Leftrightarrow & (3) & \Rightarrow & (4) \\ \Updownarrow & & \Uparrow & & \Downarrow \\ (7) & & (6) & \Leftarrow & (5) \end{array}$$

"(2) ⇔ (3)": That the statements (2) and (3) are equivalent follows straight from the definition of the linear independence.

"(1) ⇒ (3)": Assume that M is a basis of V. From the first property of a basis it follows that $V = \operatorname{span} M$, that is V is generated by M. It remains to show

[6]For the notation of a bijective map see Appendix A.

that M is linearly independent. Let us assume towards a contradiction that M is *not* linearly independent. Then this means that there exists finite many vectors $v_1, \ldots, v_m \in M$ and numbers $a_1, \ldots, a_m \in F$ (not all equal to zero) such that

$$0 = a_1 v_1 + \ldots + a_m v_m.$$

On the other hand we can also express the zero vector in the following form:

$$0 = 0v_1 + \ldots + 0v_m.$$

But this contradicts to the property of a basis which says that by definition there is only one unique way to write the zero vector as a linear combination of vectors of M. Therefore our assumption that M is linearly dependent is shown to be false and it follows that M is linearly independent.

"(3) $\Rightarrow$ (4)": We assume that (3) holds, that is M is a linearly independent generating system of V. We have to show that for every $v \in V$ which is not a vector of M the set $M \cup \{v\}$ becomes linearly dependent. But since M is assumed to be a generating system of V it follows that $v \in \operatorname{span} M$ and thus $M \cup \{v\}$ is linearly dependent by Lemma 2.17.

"(4) $\Rightarrow$ (5)": If M is a maximal linear independent subset of V then M is also a maximal linear independent subset of any subset $E \subset V$ which contains M.

"(5) $\Rightarrow$ (6)": Evident since V is a generating system of V containing M.

"(6) $\Rightarrow$ (3)": We assume (6). Since M is already assumed to be linearly independent it remains to show that $V = \operatorname{span} M$. Since M is a maximal linear independent subset of E it follows from the second part of Lemma 2.21 that $E \subset \operatorname{span} M$. In particular this means that $\operatorname{span} E = \operatorname{span} M$ and since E is a generating system of V we get that $V = \operatorname{span} M$.

"(3) $\Rightarrow$ (1)": We assume that $V = \operatorname{span} M$ (that is M satisfies the first property of a basis) and that M is linearly independent. Thus we need to show that M satisfy also the second property of a basis. Assume therefore that $x, y \in F^{(M)}$ are two elements such that

$$\sum_{u \in M}{}' x(u)u = \sum_{u \in M}{}' y(u)u.$$

Then

$$\sum_{u \in M}{}' (x(u) - y(u))u = 0$$

and since M is assumed to be linearly independent this means that $x(u) - y(u) = 0$ for every $u \in M$ (Proposition 2.19). Thus $x(u) = y(u)$ for every $u \in M$ and this means that $x = y$ as required by the definition of a basis.

"(1) $\Rightarrow$ (7)": We define a function $f\colon F^{(M)} \to V$ by

$$f(x) := \sum_{u \in M}{}' x(u)u.$$

Then f is surjective due to the first property of a basis and injective due to the second property of a basis. This means that f is a bijective map. Further f satisfies the condition (a), see (54) and (55). Further we have by the definition of the map f and of the elements e_v, see (66), for every $v \in M$ that

$$f(e_v) = \sum_{u \in M}{}' e_v(u)u = v$$

and thus also the condition (b) is satisfied.

"(7) $\Rightarrow$ (1)": Assume the condition (7) and let $v \in V$ be an arbitrary vector. Since f is surjective there exists a $x \in F^{(M)}$ such that $f(x) = v$. If one applies f to the equation (67) one gets using the relations (54) and (55)

$$v = f(x) = \sum_{u \in M}{}' f(x(u)e_u) = \sum_{u \in M}{}' x(u) f(e_u) = \sum_{u \in M}{}' x(u)u$$

and thus $v \in \operatorname{span} M$. This is the first property of a basis. From the injectivity of f follows that also the second property of a basis is satisfied. □

Now one *very* important result for Linear Algebra is the following theorem about the existence of a basis for a vector space which we will state without a proof.

Theorem 2.23 (Existence of a Basis)**.** *Every vector space V has a basis.* □

The idea of the proof is roughly the following: Due to the characterisation of a basis (Proposition 2.22) it is enough to show that every vector space V has a *maximal* linear independent subset. The existence of such a set is proven using a result from set theory, called *Zorn's Lemma*. For more details see the Appendix C.

Example. The field of real numbers $\mathbb{R}$ can be considered as a vector space over the field of rational numbers $\mathbb{Q}$. Then the above theorem states that there *exists* a $\mathbb{Q}$-linear independent subset $B \subset \mathbb{R}$ such that for every $x \in \mathbb{R}$ there exists *finitely many* (!) elements $b_1, \ldots, b_m \in B$ and numbers $a_1, \ldots, a_m \in \mathbb{Q}$ such that

$$x = a_1 b_1 + \ldots + a_m b_m.$$

Note that the above theorem does not give an answer how this set looks like. The theorem only states that this subset of $\mathbb{R}$ exists.

For finitely generated vector space the existence of a basis is easier prove. We state therefore the bit weaker form of the above theorem.

Theorem 2.24 (Existence of a Basis for Finitely Generated Vector Spaces)**.** *Let V be a finitely generated vector space. Then V has a finite basis. More precisely, every finite generating system of V contains a basis of V.*

Proof. Let E be a finite generating system. Then since E is finite it must contain a minimal generating system M. But then M is by Proposition 2.22, (2) a basis of V. □

5. The Rank of a Finite System of Vectors

In the following we will consider nearly exclusively finitely generated vector spaces. As a self-evident foundation we will make extensive use of Theorem 2.24 which we obtained in a very direct way.

In this section we will introduce another improtant concept which will recure over and over again: the rank of a system of vectors. Variants of this concept exists and they all relate closely to each other.[7] The rank of a system of vectors will turn out to be a very fruitful concept.

Assume that we are given a *finite* system of vectors $u_1, \ldots, u_m$ of a vector space V. The *rank* of this collection shall be the maximal number of linear independent vectors in this collection. More precisely this is formulate in the following way.

[7] For example in the end of this chapter we will define what we mean by the rank of a matrix and in the next chapter we will assign to every linear map f between finite dimensional vector spaces a number which we will call the rank of f.

Definition 2.25 (Rank of a System of Vectors)**.** Let $u_1, \ldots, u_m$ be a finite system of vectors (not necessarily distinct) of the vector space V. We say that the *rank* of this system of vectors is r, in symbols

$$\operatorname{rank}(u_1, \ldots, u_m) := r,$$

if the following two conditions are satisfied:

(1) There *exists* a linear independent subset of $\{u_1, \ldots, u_m\}$ which consists of exactly r vectors.
(2) Every subset of $\{u_1, \ldots, u_m\}$ which consists of $r + 1$ vectors is linear dependent.

If we speak in the following of a *system* $u_1, \ldots, u_m$ of vectors of a vector space V then we mean always the *m-tuple* $(u_1, \ldots, u_m)$ of vectors in V. One must differentiate this from the *set* $\{u_1, \ldots, u_m\}$. In the later case the order of the vectors is not important, the order in which the the vectors $u_1, \ldots, u_m$ appeare in the m-tuple are essential!

Note that we have clearly

$$\operatorname{rank}(u_1, \ldots, u_m, u_{m+1}) \geq \operatorname{rank}(u_1, \ldots, u_m),$$

that is adding a vector to a system of vectors does not decrease the rank of the system, and that the rank of a system is for obvious reason bounded by m, that is

$$\operatorname{rank}(u_1, \ldots, u_m) \leq m.$$

Equality holds, that is

$$\operatorname{rank}(u_1, \ldots, u_m) = m,$$

if and only if the system of vectors $u_1, \ldots, u_m$ are linearly independent. Summarizing these observations yields the following result.

Proposition 2.26. *Let $u_1, \ldots, u_m$ a system of vectors of the F-vector space V. Then the following statements are equivalent:*

(1) *The system $u_1, \ldots, u_m$ is linearly independent.*
(2) *We have* $\operatorname{rank}(u_1, \ldots, u_m) = m$.
(3) *The set $\{u_1, \ldots, u_m\}$ is a linear independent subset of V consisting* exactly *of m vectors.*
(4) *If there exists elemenst $a_1, \ldots, a_m \in F$ such that*

$$a_1 u_1 + \ldots + a_m u_m = 0$$

then necessarily $a_1 = \ldots = a_m = 0$. □

Note that if we add the zero vector 0 to a system of vectors $u_1, \ldots, u_m$, then the rank of this system does not change. That is we have the equality

$$\operatorname{rank}(u_1, \ldots, u_m, 0) = \operatorname{rank}(u_1, \ldots, u_m).$$

This observation generalizes to the following result.

Proposition 2.27. *Let $u_1, \ldots, u_m$ be a system of vectors of the vector space V and $u \in \operatorname{span}\{u_1, \ldots, u_m\}$. Then*

$$\operatorname{rank}(u_1, \ldots, u_m, u) = \operatorname{rank}(u_1, \ldots, u_m).$$

PROOF. Let us simplify the notation by denoting with r the rank of the system of vectors $u_1, \ldots, u_m$, that is we set

$$r := \operatorname{rank}(u_1, \ldots, u_m). \tag{69}$$

We know already that $\operatorname{rank}(u_1, \ldots, u_m, u) \geq r$. Thus it remains to show that also $\operatorname{rank}(u_1, \ldots, u_m, u) \leq r$ is true. In order to show this we need to prove that there exists no linear independent subset $\{u_1, \ldots, u_m, u\}$ which consists of $r+1$ elements.

Thus let us assume towards a contradiction that there actually exists a linear independent set $T \subset \{u_1, \ldots, u_m, u\}$ which consists of $r+1$ elements. Due to (69) we have that necessarily $u \in T$ and that the set $M := T \setminus \{u\}$ is a maximal linear independent subset of $\{u_1, \ldots, u_m\}$. Thus it follows from the second part of Lemma 2.21 that

$$\operatorname{span} M = \operatorname{span}\{u_1, \ldots, u_m\}. \tag{70}$$

But due to assumption that $u \in \operatorname{span}\{u_1, \ldots, u_m\}$ we have that $u \in \operatorname{span} M$. Thus $M \cup \{u\} = T$ is linear dependent by Lemma 2.17. But this is a contradiction to our assumption that T is linearly independent.

Therefore there cannot exists a linearly independent subset of $\{u_1, \ldots, u_m, u\}$ with $r+1$ elements and this shows that $\operatorname{rank}(u_1, \ldots, u_m, u) \leq r$. □

Next we turn our attention to the question how we can determine the rank of a given system $u_1, \ldots, u_m$ of vectors of a F-vector space V. Since we can always replace the vector space V by the finitely generated vector space spanned by the vectors $u_1, \ldots, u_m$ we will assume in the following that

$$V \textit{ is a finitely generated } F\textit{-vector space.} \tag{71}$$

Thus Theorem 2.24 applies and we know that V has a finite basis. Assume that M is such a finite basis of V, say

$$M = \{b_1, \ldots, b_n\} \tag{72}$$

consists of exactly n distinct vectors of V. Then every vector $u \in V$ can be written as

$$u = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + \ldots + a_n b_m \tag{73}$$

with uniquely determined numbers $a_i \in F$. Observe that the n-tuple $(a_1, \ldots, a_n)$ does not – besides of the vector u – only depend on the *set* M but also on the *numbering* of the basis vectors $b_1, \ldots, b_n$. Therefore it is useful to make the following definition.

Definition 2.28 (Ordered Basis)**.** Let V be a finitely generated vector space. An *ordered basis* of V is a n-tuple $(b_1, \ldots, b_n)$ of pairwise distinct vectors of V such that the set $\{b_1, \ldots, b_n\}$ is a basis of V in the sense of Definition 2.14.

It is a straight consequence of Theroem 2.24 that every finitely generated vector space V posesses an ordered basis. In the following we will usually make use of an ordered basis of a vector space. In order to simplify the notation we will therefore agree to call an ordered basis simply just a basis. Therefore it will be the context which will tell whether the therm "basis" will refer to an set or an ordered system. When we speak in the following of the *canonical basis* of the vector space F^n then we will nearly always mean the ordered system $e_1, \ldots, e_n$ of the canonical unit vector (65) of the vector space F^n.

If V is a finitely generated F-vector space with $b_1, \ldots, b_n$ being an (ordered) basis of V then every vector $u \in V$ can be expressed in a unique way in the form (73). One calles the $a_i \in F$ the *coordinates* of the vector u with respect to

the basis $b_1, \ldots, b_n$. The n-tuple $(a_1, \ldots, a_n) \in F^n$ is called the *coordinate vector* of u with respect to the basis $b_1, \ldots, b_n$.

Next we will describe an

Algorithm to Compute the Rank of a Given System of Vectors. We first declare – similar to Chapter 1 – what we mean by elementary transformations of a system of vectors.

Definition 2.29. Under an *elementary transformation* of a system of vectors $(u_1, \ldots, u_m)$ of a F-vector space we shall understand one of the following operations which transforms the system again into a system of m elements:

(I) Replacing one u_i in $(u_1, \ldots, u_m)$ by $u_i + au_j$ with $a \in F$ and $i \neq j$.
(II) Exchanging two vectors in the system.
(III) Replacing one vector u_i in $(u_1, \ldots, u_m)$ by au_i where $a \in F \setminus \{0\}$.

Next we shall show, that we may use these transformations in order to determine the rank of a vector system.

Proposition 2.30. *The rank of a system of vectors $(u_1, \ldots, u_m)$ is not changed under the elementary transformations of Definition 2.29.*

PROOF. (I) We may assume with out any loss of generality that $i = 1$ and $j = 2$. Then $u_1' := u_1 + au_2$ is a linear combination of the vectors $u_1, \ldots, u_m$ and we have

$$\operatorname{rank}(u_1, \ldots, u_m) = \operatorname{rank}(u_1, \ldots, u_m, u_1') \geq \operatorname{rank}(u_1', u_2, \ldots, u_m)$$

by Proposition 2.27 and the observations on page 34. On the other hand $u_1 = u_1' - au_2$ and thus the system $u_1, \ldots, u_m$ can be obtained from the system $u_1', u_2, \ldots, u_m$ by an elementary transformation of type (I). Thus we have

$$\operatorname{rank}(u_1', u_2, \ldots, u_m) = \operatorname{rank}(u_1, \ldots, u_m, u_1') \geq \operatorname{rank}(u_1, \ldots, u_m)$$

for the same reason as above. Alltogether this means that

$$\operatorname{rank}(u_1, \ldots, u_m) = \operatorname{rank}(u_1', u_2, \ldots, u_m)$$

as needed to be shown.

(II) The claim is obvious for a transformation of type (II).

(III) The proof is done analogous as in the case of a transformation of type (I). □

Now let $u_1, \ldots, u_m$ be a given system of vectors of a finitely generated F-vector space V. Let $b_1, \ldots, b_n$ be an ordered basis of V. Then every vector of the system $u_1, \ldots, u_m$ can be written in a unique way as a linear combination of the basis vectors $b_1, \ldots, b_n$:

$$u_i = \sum_{j=1}^{n} a_{ij} b_j, \quad i = 1, 2, \ldots, m \tag{74}$$

with numbers $a_{ij} \in F$. Let us arrange these numbers in the following $m \times n$-matrix:

$$A := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \tag{75}$$

We shall call this matrix the *coordinate matrix* of the vector system $u_1, \ldots, u_m$ with respect to the ordered basis $b_1, \ldots, b_n$. Note that the rows of this matrix are

by definition exactly the coordinate vectors of the vectors $u_1, \ldots, u_m$ with respect to the basis $b_1, \ldots, b_n$.

An elementary transformation of the system $u_1, \ldots, u_m$ is equivalent with an elementary row transformation of the coordinate matrix (75). We have seen in Chapter 1 how we can transform this matrix into a matrix of the form (21). To achive this form we possible need to exchange columns, but such a column exchange corresponds to re-numbering the elements $b_1, \ldots, b_n$ of the basis and this is irrelevant for computing $\operatorname{rank}(u_1, \ldots, u_m)$. Thus we optain the following result.

Theorem 2.31 (Computation of the Rank of Finite Systems of Vectors)**.** *Let V be a finitely generated F-vector space and let $b_1, \ldots, b_n$ be a basis of V. Then one can transform step by step using elementary transformation any finite system of vectors $u_1, \ldots, u_m$ of V into a systme $u'_1, \ldots, u'_m$ such that the coordinate matrix of this transformed system with respect to a basis which differs from the basis $b_1, \ldots, b_n$ at most by a re-numeration is of the following form*

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & \cdots & \cdots & 0 & \\ 0 & 1 & 0 & & & \vdots & \\ 0 & 0 & 1 & & & \vdots & * \\ \vdots & & & \ddots & & \vdots & \\ \vdots & & & & \ddots & 0 & \\ 0 & 0 & 0 & \cdots & 0 & 1 & \\ \hline & & & 0 & & & 0 \end{array}\right) \tag{76}$$

where the upper left part is the $r \times r$-identity matrix and r is a certain natural number $0 \leq r \leq m, n$. Then

$$\operatorname{rank}(u_1, \ldots, u_m) = r \tag{77}$$

Proof. We only have to show the last claim. Since we know by Proposition 2.30 that elementary transformations do not change the rank of a system of vectors we need just to show that if $u_1, \ldots, u_m$ is a system of vectors such that its coordinate matrix with respect to some basis $b_1, \ldots, b_n$ is of the form (76) then (77) holds.

Since the last $m - r$ rows of the matrix (76) contains only zeros it follows that in this case the vectors $u_i = 0$ for $i \geq r + 1$. Thus we have that necessarily

$$\operatorname{rank}(u_1, \ldots, u_m) = \operatorname{rank}(u_1, \ldots, u_r, 0, \ldots, 0) = \operatorname{rank}(u_1, \ldots, u_r) \leq r.$$

It remains to show that also $\operatorname{rank}(u_1, \ldots, u_r) \geq r$ holds. Now from the definition of the coefficients of the coordinate matrix – see (74) and (75) – we get for $0 \leq i \leq r$ the equations

$$u_i = b_i + \sum_{j=r+1}^{n} a_{ij} b_j = b_i + u_i^*$$

where $u_i^* \in \operatorname{span}\{b_{r+1}, \ldots, b_n\}$. Now assume that $c_i \in F$ are numbers such that

$$c_1 u_1 + \ldots + c_r u_r = 0$$

is a linear combination of the zero vector. Then

$$\begin{aligned} c_1 u_1 + \ldots + c_r u_r &= c_1 b_1 + \ldots + c_r b_r + u^* \qquad (\text{with } u^* \in \operatorname{span}\{b_{r+1}, \ldots, b_n\}) \\ &= 0 \end{aligned}$$

if and only if $c_1 = \ldots = c_r = 0$ since the $b_1, \ldots, b_n$ are linearly independent and therefore there 0 is only the trivial linear combination of these vectors. Thus the vectors $u_1, \ldots, u_r$ are linearly independent and it follows $\operatorname{rank}(u_1, \ldots, u_m) = \operatorname{rank}(u_1, \ldots, u_r) \geq r$.

Alltogether we have shown that $\operatorname{rank}(u_1, \ldots, u_m) \leq r$ and $\operatorname{rank}(u_1, \ldots, u_m) \geq r$ and thus equality holds. □

Example. Consider following system of four vectors of the vector space $V = \mathbb{R}^5$:

$$u_1 := \begin{pmatrix} 1 \\ 3 \\ 5 \\ 2 \\ 0 \end{pmatrix}, \quad u_2 := \begin{pmatrix} 3 \\ 9 \\ 10 \\ 1 \\ 2 \end{pmatrix}, \quad u_3 := \begin{pmatrix} 0 \\ 2 \\ 7 \\ 3 \\ -1 \end{pmatrix}, \quad u_4 := \begin{pmatrix} 2 \\ 8 \\ 12 \\ 2 \\ 1 \end{pmatrix}$$

Its coordinate matrix with respect to the canonical basis $e_1, \ldots, e_5$ of $\mathbb{R}^5$ is the 4×5-matrix from the example on page 10 which can be transformed using elementary row transformations as follows:

$$\begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 3 & 9 & 10 & 1 & 2 \\ 0 & 2 & 7 & 3 & -1 \\ 2 & 8 & 12 & 2 & 1 \end{pmatrix} \rightarrow \left(\begin{array}{ccc|cc} 1 & 3 & 5 & 2 & 0 \\ 0 & 2 & 2 & -2 & 1 \\ 0 & 0 & 5 & 5 & -2 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccc|cc} 1 & 0 & 0 & 3 & -7/10 \\ 0 & 1 & 0 & -2 & 9/10 \\ 0 & 0 & 1 & 1 & -2/5 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Thus from Theorem 2.31 it follows that $\operatorname{rank}(u_1, u_2, u_3, u_4) = 4$. Note that actually we can already see this from from the second matrix in the above chain of matrices, that is, we do not necessarily need to completely transform the coordinate matrix into the form required by Theorem 2.31.

6. The Dimension of a Vector Space

We are now ready to introduce an other very important concept in Linear Algebra, namely the dimension of a vector Space.

Definition 2.32. We say that a F-vector space V has *dimension* n if V has a basis $b_1, \ldots, b_n$ which consists of exactly n (distinct) vectors.

The aim of this section is to show that if V has dimension m and dimension n for some integers m and n, then necessarily $m = n$ (Theorem 2.34).

Straight from Theorem 2.31 we get the following result.

Theorem 2.33. *Let $u_1, \ldots, u_m$ a system of vectors of a vector space V of dimension n. Then*

$$\operatorname{rank}(u_1, \ldots, u_m) \leq n.$$ □

Corollary. *Let M be a linearly independent set of a n-dimensional vector space V. Then M has at most n elements.*

Proof. Assume towards a contradiction that M contains at least $n+1$ pairwise distinct vectors $u_1, \ldots, u_{n+1}$. Then these vectors form a linearly independent system of vectors and we have $\operatorname{rank}(u_1, \ldots, u_{n+1}) = n + 1$. But this contradicts to Theorem 2.33 which states that $\operatorname{rank}(u_1, \ldots, u_{n+1}) \leq n$! Thus the assumption was wrong and necessarily M has at most n elements. □

Theorem 2.34 (Invariance of the Dimension)**.** *Let V be a finitely generated vector space. Then every basis of V has the same number of elements. In other words, every basis of an n-dimensional vector space consist precisely of n elements.*

PROOF. Let $E \subset V$ be a finite generating system of V. Apparently E has a maximal linear independent subset M. Then M is a basis of V by Proposition 2.22. Since E is finite M must be finite, too. Say M has precisely n elements. Then V has dimension n by Definition 2.32.

Now let M' be another basis of V. Then M' is a linear independent subset of V and thus M' contains at most n elements by the corollary to Theorem 2.33. Denote by m the number of elements in M'. We have then $m \leq n$ and V has dimension m by Definition 2.32. Thus M as a linear independent subset of V has at most m elements by the corollary to Theorem 2.33, that is $n \leq m$. Therefore we have alltogether $m = n$. □

The above theorem justifies that we can speak of *the* dimension of a vector space and that we denote the dimension of V in symbols by

$$\dim V := n$$

in case that V has dimension n. In this case we say also that V is a *finite dimensional* vector space. Note that $\dim V = n$ for some natural number n if and only if V is finitely generated. Thus a vector space is finite dimensional if and only if it is finitely generated. Note further that the trivial zero space $V = 0$ has dimension 0 since its only basis contains 0 elements (recall that we agreed that the we consider the empty set $\emptyset$ to be the basis of the zero space).

If V is not finitely generated, then we aggree to say that V has infinite dimension and we denote this fact in symbols by

$$\dim V := \infty.$$

The dimension of a F vector space V depends not only on the set V but also on the field F. For example consider the $\mathbb{R}$-vector space $\mathbb{R}$. Then $\dim \mathbb{R} = 1$. On the other hand if one considers $\mathbb{R}$ as a vector space over the rational numbers then $\mathbb{R}$ is infinite dimensional (compare the example on page 33). Thus the notation

$$\dim_F V$$

might be used to emphasis the field F of the F-vector space. Using this notation we have for example

$$\dim_{\mathbb{R}} \mathbb{R} = 1 \qquad \text{and} \qquad \dim_{\mathbb{Q}} \mathbb{R} = \infty.$$

Theorem 2.35. *Let $u_1, \ldots, u_m$ be a system of vectors of a n-dimensional vector space V. Then this system of vectors forms a basis of V if and only if*

$$\operatorname{rank}(u_1, \ldots, u_m) = n \qquad \textit{and} \qquad m = n$$

PROOF. "⇒": If $u_1, \ldots, u_m$ is a basis of V the vectors are linearly independent. Thus $\operatorname{rank}(u_1, \ldots, u_m) = m$ by Proposition 2.26. It follows from Theorem 2.34 that $m = n$.

"⇐": By assumption we have that $\operatorname{rank}(u_1, \ldots, u_m) = m$. Thus $M := \{u_1, \ldots, u_m\}$ is a linear independent subset of V consisting of exactly m elements by Proposition 2.26. Since $m = n$ this means by the corollary to Theorem 2.33 that M is a maximal linear independent subset of V and therefore a basis of V by Proposition 2.22. □

If M is a basis of a vector space, then M is a linear independent set and with it every subset $N \subset M$ is linearly independent, too. Allso the converse is true: if N is an linear independent subset of a vector space V, then one can always extend N to a maximal linear independent subset M of V, which then is a basis of V due to Proposition 2.22. We shall prove this result for finite dimensional vector spaces:

Theorem 2.36 (Basis Extension Theorem)**.** *Let N be a linear independent subset of a finite dimensional vector space V. Then N can be extendet to a basis of V, that is there exists a basis M of V such that $N \subset M$.*

PROOF. Let N be an arbitrary linear independent subset of V. Due to the corollary to Theorem 2.33 the set N has at most n elements. Let B be a basis of V. Then the set

$$E := B \cup N$$

is a finite generating system of V which contains the linear independent subset N. Since E is finite it must surely contain a maximal linear independent subset M with $N \subset M$. Then M is a basis of V by Proposition 2.22 which has by construction the properties required by the theorem. □

Theorem 2.37. *Let U a subspace of the vector space V. Then*

$$\dim U \leq \dim V. \tag{78}$$

In particular a subspace of a finite dimensional vector space is again finite dimensional.

If V is finite dimensional then $U = V$ if and only if $\dim U = \dim V$.

PROOF. If $\dim V = \infty$ then the inequality (78) is clearly satisfied. Thus we assume that V is finite dimensional and set $n := \dim V$.

We first need to show that then U is finite dimensional. Let M be a linear independent subset of U. Then M is also a linear independent subset of V and has therefore at most n elements due to the corollary to Theorem 2.33. Let m be the maximal number such that there exists a linear independent subset of U with m elements (this number exists since n is a finite number). Let M be such a linear independent subset of U with m elements. Then M is a maximal linear independent subset of U and therefore a basis of U by Proposition 2.22. Thus U is finite dimensional and $\dim U = m \leq n = \dim V$. This proves the first part.

Assume that V is finite dimensinal. If $U = V$, then clearly $\dim U = \dim V$. If on the other hand $\dim U = \dim V$, then U has a basis which consists of $n := \dim U$ elements. But then M is also a basis of V due to Theorem 2.35 and it follows that $U = V$. □

We shall conclude this section with an application of the basis extension theorem from above which yield a nearly self-evident result:

Proposition 2.38. *Let V be an arbitrary vector space (finite or infinite dimensional) and let $u_1, \ldots, u_m$ be an arbitrary finite system of vectors of V. Then*

$$\dim \operatorname{span}\{u_1, \ldots, u_m\} = \operatorname{rank}(u_1, \ldots, u_m).$$

That is, the rank of the system $u_1, \ldots, u_m$ is the equal to the dimension of the subspace spanned by the vectors of the system.

PROOF. Let us use the following abreviation to simplify the notation:

$$U := \operatorname{span}\{u_1, \ldots, u_m\}, \qquad r := \operatorname{rank}(u_1, \ldots, u_m), \qquad k := \dim U.$$

Theorem 2.24 states that $\{u_1, \ldots, u_m\}$ contains a basis M of U and since $\dim U = k$ it follows that M consists of exactly k distinct vectors. We may assume without any loss of generality that $M = \{u_1, \ldots, u_k\}$. Then

$$k = \operatorname{rank}(u_1, \ldots, u_k) \leq \operatorname{rank}(u_1, \ldots, u_k, \ldots, u_m) = r$$

On the other hand is $u_1, \ldots, u_m$ a system of vectors of the k-dimensional vector space U and thus $r \leq k$ by Theorem 2.33. Alltogether we have therefore $k = r$. □

Corollary. *Let $u \in V$. Then $u \in \operatorname{span}\{u_1, \ldots, u_m\}$ if and only if*

$$\operatorname{rank}(u_1, \ldots, u_m, u) = \operatorname{rank}(u_1, \ldots, u_m).$$

PROOF. "⇒": This follows from Proposition 2.27.

"⇐": Due to $\operatorname{rank}(u_1, \ldots, u_m, u) = \operatorname{rank}(u_1, \ldots, u_m)$ we get

$$\dim \operatorname{span}\{u_1, \ldots, u_m, u\} = \dim \operatorname{span}\{u_1, \ldots, u_m\}.$$

But now $\operatorname{span}\{u_1, \ldots, u_m\}$ is a subspace of $\operatorname{span}\{u_1, \ldots, u_m, u\}$ of the same finite dimension and thus $\operatorname{span}\{u_1, \ldots, u_m\} = \operatorname{span}\{u_1, \ldots, u_m, u\}$ by Theorem 2.37. □

7. Direct Sum and Linear Complements

In this section V shall always denote a *finite* dimensional vector space if not otherwise stated. This section will make extensively use of the basis extension theorem of the previous section and illustrate its powerfull application.

If U and W are two subspaces of V, then we have seen already that their sum $U + W$ and their intersection $U \cap W$ are subspace of V (see Proposition 2.7). We are in the following interested in the situation where $U + W$ is the largest possible subspace of V, namely the whole space V, and where $U \cap W$ is the smallest possible subspace of V, namely the zero vector space 0.

Definition 2.39 (Direct Sum)**.** Let V be an arbitrary vector space (finite or infinite dimensional). Let U and W be two subspaces of V. We say that V is the *(internal) direct sum* of the subspaces U and W if the two following conditions are satisfied:

$$V = U + W \qquad \text{and} \qquad U \cap W = 0.$$

If V is the direct sum of U and W, then we may express this fact in symbols by $V = U \oplus W$.[8]

If W is a subspace of V such that $U \cap W = 0$ then we say that W is a *transversal* space of U in V. If even $V = U \oplus W$, then we say that W is a *linear complement* of U in V.

Note that if W is a linear complement (transversal space) of U in V then also U is a linear complement (transversal space) of W in V.

As an application of the basis extension theorem we can see that if we are given a subspace U of the finite dimensional vector space V then there always a linear complement W of U in V. To see this we choose a basis M of U and extend it to a basis B of V. We set $M' := B \setminus M$. Then $W := \operatorname{span} M'$ is clearly a subspace of V such that

$$V := U \oplus W.$$

Thus we have just proven the following result.

[8]Note there exists also an construction of an external direct sum which is normaly denoted with the same symbol "⊕" which very closely related to the concept of the internal direct sum. But we do not consider this construction here.

Theorem 2.40 (Existence of Linear Complements)**.** *Let U be a given subspace of a finite dimensional vector space V. Then there exists a linear complement W of U, that is there exists a subspace W of V such that*

$$V = U \oplus W \qquad \square$$

Note that the linear complement of a given subspace U is *not* unique, that is there exist in general many linear complements of a given subspace U.

Proposition 2.41. *Let U and W be subspaces of an arbitrary vector space V (finite or infinite dimensional). Then $V = U \oplus W$ if and only if for every vector $v \in V$ there exists unique vectors $u \in U$ and $w \in W$ such that*

$$v = u + w. \tag{79}$$

PROOF. "⇒": We assume that $V = U \oplus W$. That we can find vectors $u \in U$ and $w \in W$ such that (79) holds is clear. Thus we need only to show that this decomposition is unique.

Therefore let $u' \in U$ and $w' \in W$ some vectors (possible distinct from u and w) such that $v = u' + w'$. Then from $u + w = v = u' + w'$ follows that

$$u - u' = w' - w. \tag{80}$$

Now $u - u' \in U$ and $w' - w \in W$ and thus it follows from (80) that both $u - u'$ and $w' - w$ are elemnts of $U \cap W = 0$. Thus $u - u' = 0$ and $w' - w = 0$ or in other words $u = u'$ and $w = w'$. Thus the decomposition (79) is seen to be unique.

"⇐": By assumption $V = U + W$. Thus we need only to show that $U \cap W = 0$. Assume that $v \in U \cap W$. Then both $v = v + 0$ and $v = 0 + v$ are decompositions of the form (79). But due to the uniqueness requirement this means that $v = 0$. Therefore $U \cap W = 0$ is the zero space. Alltogether this shows that $V = U \oplus W$. $\square$

Corollary. *Assume that $V = U \oplus W$ and let $u \in U$ and $w \in W$. Then $u + w = 0$ if and only if $u = 0$ and $w = 0$.* $\square$

Theorem 2.42 (Dimension Formula for Subspaces)**.** *Let U and W be two finite dimensional subspaces of an arbitrary (finite or infinite dimensional) F-vector space V. Then we have the following dimension formular:*

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W) \tag{81}$$

PROOF. Again we will use in this proof the basis extension theorem. We set $m := \dim(U \cap W)$ and choose a basis $a_1, \ldots, a_m$ of $U \cap W$. On one hand we extend this basis to a basis

$$a_1, \ldots, a_m, b_1, \ldots, b_r \tag{82}$$

of the subspace U and on the other hand we extend the same basis to a basis

$$a_1, \ldots, a_m, c_1, \ldots, c_s \tag{83}$$

of the subspace W. Note that in this notation we have $\dim U = m + r$ and $\dim W = m + s$. We claim now that

$$a_1, \ldots, a_m, b_1, \ldots, b_r, c_1, \ldots, c_s \tag{84}$$

is a basis of $U + W$. As soon as we have shown it follows that the dimension formula (82), since then

$$\begin{aligned} \dim(U + W) = m + r + s &= (m + r) + (m + s) - m \\ &= \dim(U) + \dim(W) - \dim(U \cap W). \end{aligned}$$

It is clear that (84) is a generating system of $U + W$. It remains to show that the vectors $a_1, \ldots, a_m, b_1, \ldots, b_r, c_1, \ldots, c_s$ are linearly independent because then they form a basis of $U + W$ due to Proposition 2.22. Therefore we assume that the zero vector 0 is a linear combination of those vectors, that is we assume that

$$\sum_{i=1}^{m} \lambda_i a_i + \sum_{i=1}^{r} \mu_i b_i + \sum_{i=1}^{s} \nu_i c_i = 0 \tag{85}$$

for some coefficients $\lambda_i, \mu_i, \nu_i \in F$. We need to show that all those coefficients are necessarily equal to zero. To simplify the notation let us denote the three partial sums of (85) by a, b and c, that is in this notation (85) reads

$$a + b + c = 0. \tag{85'}$$

Then $a \in U \cap W$, $b \in U$ and $c \in W$ and it follows from (85′) that $b = -a - c \in W$ and therefore $b \in U \cap W$. But this means that the coefficients $\mu_1, \ldots, \mu_i$ are all equal to zero due to the choice of the vectors $b_1, \ldots, b_r$ and thus $b = 0$. Similarly one deduces that also the coefficients $\nu_1, \ldots, \nu_s$ are all equal to zero and thus $c = 0$. But then from (85′) follows that $a = 0$ and thus the remaining coefficients $\lambda_1, \ldots, \lambda_m$ are necessarily all equal to zero, too. Thus we have seen that the vectors (84) are linearly independent and since they generate $U + W$ this means that those vectors form a basis of $U + W$ as we wanted to show in order to complete the proof. □

Example. Let U and W be two 2-dimensional subspaces of the 3-dimensional $\mathbb{R}$-vector space $\mathbb{R}^3$ and assume that $U \neq W$. Then it is easy to see that $U + W = \mathbb{R}^3$ and it follows from the above dimension formula that

$$\dim(U \cap W) = \dim U + \dim W - \dim \mathbb{R}^3 = 2 + 2 - 3 = 1.$$

Thus $U \cap W$ is a 1-dimensional subspace of $\mathbb{R}^3$ and in particular $U \cap W$ is not the zero space 0. That is there exists a non-zero vector $v \in U \cap W$ and $U \cap W = \operatorname{span}\{v\} = \{av : a \in \mathbb{R}\}$. This set is a straight line in $\mathbb{R}^3$ passing through the origin 0. The example shows that two planes in $\mathbb{R}^3$ which contain the origin 0 of $\mathbb{R}^3$ – here U and W – intersect always in a straight line passes through the origin of $\mathbb{R}^3$ – here $U \cap W$.

As a consequence of Theorem 2.42 we shall note the following criterion for a subspace W being a transversal space or even a linear complement of a given space U in V.

Proposition 2.43. *Let V be an arbitrary vector space (finite or infinite dimensional) and let U and W be two finite dimensional subspaces of V. Then we have that*

(1) *W is a transversal space of U in V if and only if*

$$\dim(U + W) = \dim U + \dim W,$$

and

(2) *W is a linear complement of U in V if and only if*

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

Proof. This result is consequence of the following two obeservations in the given setting that U and V are finite dimensional subspaces of V: The condition $U \cap W = 0$ is equivalent with $\dim(U \cap W) = 0$ and the condition $V = U + W$ is equivalent with $\dim V = \dim(U + W)$. □

Note that the second requirement of the above proposition is equivalent with the vector space V being the direct sum of the finite dimensional subspaces U and W, that is $V = U \oplus W$. In particular is V in this case finite dimensional, too.

8. Row and Column Rank of a Matrix

In this section we shall study a bit in more detail the connection between the computation of the rank of a system of vectors and elementary transformations of a matrix.

Definition 2.44 (Row and Column Rank of a Matrix)**.** Let

$$A := \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \tag{86}$$

be an arbitrary $m \times n$-matrix with coefficients in a field F. We denote by $u_1, \dots, u_m$ the rows of the matrix A, considered as vectors of the vector space F^m. Similarly we denote by $v_1, \dots, v_n$ the columns of the matrix, considered as vectors of the vector space F^n. Then the *row rank* of the matrix A is the rank of the system of vectors $u_1, \dots, u_m$ and is denoted in symbols by

$$\operatorname{rank}_{\mathrm{r}} A := \operatorname{rank}(u_1, \dots, u_m).$$

Similarly we define the *column rank* of the matrix A to be the rank of the system of vectors $v_1, \dots, v_n$ and we denote the column rank of A in symbols by

$$\operatorname{rank}_{\mathrm{c}} A := \operatorname{rank}(v_1, \dots, v_n).$$

Note that the row rank of a matrix remains unchanged under elementary row transformations due to Proposition 2.30. Similarly the column rank remains invariant under elementray row transformations.[9]

The natural question is how the row and column rank of a matrix relate to each other. The answer will be simple: both numbers are always equal. This will mean that it will make sense to assign a rank to a matrix using either ways of computation. The aim of this section will be to show this equality of row and column rank. In the next section we will then use the result of this section to answer the two open problems from Section 5 of Chapter 1.

Let $u_1, \dots, u_m$ be a system of vectors of a finite dimensional F-vector space V, and let $b_1, \dots, b_n$ be a basis of V. Let us for a while use the following notation: for every vector $u \in V$ we shall denote by $\tilde{u}$ the coordinate vector of u with respect to the basis $b_1, \dots, b_n$. (This notation does express the dependency of the coodinate vectors on the basis $b_1, \dots, b_n$ but we do not mind this here cause we keep the basis fixed for the current consideration.) Now the zero vectore 0 is a non-trivial linear combination of the vectors $u_1, \dots, u_m$ if and only if 0 is a non-trivial linear combination of the corresponding coordinate vectors $\tilde{u}_1, \dots, \tilde{u}_m$. This is because for arbitrary elements $\lambda_1, \dots, \lambda_m \in F$ we have

$$\Big(\sum_{i=1}^{n} \lambda_i u_i\Big)^{\sim} = \sum_{i=1}^{n} \lambda_i \tilde{u}_i$$

and $\tilde{v} = 0$ is equivalent with $v = 0$. As a consequence of this we get more generaly

$$\operatorname{rank}(u_1, \dots, u_m) = \operatorname{rank}(\tilde{u}_1, \dots, \tilde{u}_m). \tag{87}$$

Let A be the coordinate matrix of the system of vectors $u_1, \dots, u_m$ with respect to the basis $b_1, \dots, b_n$. Then by its definition the rows of the matrix A are precisely the vectors $\tilde{u}_1, \dots, \tilde{u}_m \in F^n$. Therefore we have shown the follwoing result.

[9] Elementary column transformations of type I, II and III of a matrix are defined in the very analogous way as we have defined elementary row transformations for matrices on page 8 in the previous chapter.

Proposition 2.45. *Let $u_1, \ldots, u_m$ be a system of vectors of the finite dimensional F-vector space V. Let A the coordinat matrix of this system with respect to an arbitrary basis $b_1, \ldots, b_n$ of V. Then*

$$\operatorname{rank_r}(A) = \operatorname{rank}(u_1, \ldots, u_m). \qquad \square$$

In particular this means that the column rank of the coordinate matrix A is only dependent on the system of vectors $u_1, \ldots, u_m$ and independent of the choice of the basis $b_1, \ldots, b_n$.

In Section 5 we have shown that the rank of a system of vectors $u_1, \ldots, u_k$ is unchanged under elementary transformations and that the elementary transformations are in a one-to-one correspondence to row transformations of the coordinate matrix A of the system $u_1, \ldots, u_m$ with respect to the basis $b_1, \ldots, b_n$. But the rank of the system $u_1, \ldots, u_m$ is by definition exactly the column rank of the matrix A and thus we have shown in Section 5 that the column rank of a matrix A is left invariant under elementary row transformations (Proposition 2.30).

But what happens in the case of an elementary column transformation of a matrix. It appears that there is a beautiful analogy to the case of a row transformation. Without going into detail, the elementary column transformations of the coordinate matrix A of the system $u_1, \ldots, u_m$ with respect to the basis $b_1, \ldots, b_n$ are in a one-to-one correspondence with elementary transformations of the basis system $b_1, \ldots, b_n$. Since a elementary transformation of a vector system does not change the rank of the system (Proposition 2.30) it means that a basis $b_1, \ldots, b_n$ of V is transformed to the system $b'_1, \ldots, b'_n$ which is again a basis of V. Now the above mentioned one-to-one correspondence (which needs to be proven but we shall ommit this not too difficult proof here) means that if A' is derived from the coordinate matrix A by an elementary column transformation, then there exists a basis $b'_1, \ldots, b'_n$ such that the coordinate matrix of the system $u_1, \ldots, u_m$ with respect to the new basis $b'_1, \ldots, b'_n$ is precisely equal to A'. Thus it follows by Proposition 2.45 that

$$\operatorname{rank_r}(A') = \operatorname{rank_r}(A)$$

since both numbers are equal to $\operatorname{rank}(u_1, \ldots, u_m)$. In other words this means that the row rank is invariant under elementary column transformations.

Combining this observation with Proposition 2.30 yields the following result about the row rank of a matrix.

Proposition 2.46. *The row rank of a matrix remains invariant under elementray row and column transformations. That is, if A' is a matrix which is derived from the matrix A by elemetary row and column transformations then*

$$\operatorname{rank_r}(A) = \operatorname{rank_r}(A') \qquad \square$$

Note that the above result gives now an **Answer to Problem 1** as stated in the end of Chapter 1: the row rank of the matrix C is exactly the number which satisfies the requirements stated by the problem.

Now there exists no essential difference in the definition of row and column rank of a matrix. Let us denote by ${}^tA$ the *transposed* matrix of A which is the $n \times m$-matrix derived from (86) by mirrowing it along the top-left to botom-right diagonal, that is we set

$${}^tA := \begin{pmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \vdots & & & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{pmatrix} \tag{88}$$

Then the above statement means more precisely that naturaly

$$\operatorname{rank}_{\mathrm{r}} {}^{t}A = \operatorname{rank}_{\mathrm{c}} A$$

and

$$\operatorname{rank}_{\mathrm{c}} {}^{t}A = \operatorname{rank}_{\mathrm{r}} A$$

and that an elementary column (row) transformations of ${}^{t}A$ corresponds to a elementary row (column) transformation of A. Thus we get from Proposition 2.46 the following result about the invariance of the column rank.

Proposition 2.47. *The column rank of a matrix remains invariant under elementray row and column transformations. That is, if A' is a matrix which is derived from the matrix A by elemetary row and column transformations then*

$$\operatorname{rank}_{\mathrm{c}}(A) = \operatorname{rank}_{\mathrm{c}}(A') \qquad \square$$

From Chapter 1 we know that we can transform any $m \times n$-matrix A into a matrix of the form

$$\left(\begin{array}{ccccc|c} 1 & 0 & \dots & \dots & 0 & \\ 0 & 1 & & & \vdots & \\ \vdots & & \ddots & & \vdots & * \\ \vdots & & & \ddots & 0 & \\ 0 & \dots & \dots & 0 & 1 & \\ \hline & & 0 & & & 0 \end{array}\right) = \begin{pmatrix} I_r & * \\ 0 & 0 \end{pmatrix} \tag{89}$$

by using only row transformations and possible column transformations of type II (that is exchanging two columns). Now it is evident how to continue from this form using elementary column transformations to transform this matrix into a matrix of the form

$$A' := \left(\begin{array}{ccccc|c} 1 & 0 & \dots & \dots & 0 & \\ 0 & 1 & & & \vdots & \\ \vdots & & \ddots & & \vdots & 0 \\ \vdots & & & \ddots & 0 & \\ 0 & \dots & \dots & 0 & 1 & \\ \hline & & 0 & & & 0 \end{array}\right) = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \tag{90}$$

where the upper left block is the $r \times r$-identity matrix I_r for some number $0 \leq r \leq m, n$. Let us – before we continue – write this result down in a theorem which in a way extends Proposition 1.5 from the the previous chaptere were we only allowed row transformations.

Theorem 2.48. *Let F be a field. Then any $m \times n$-matrix can be transformed into a matrix of the form* (90) *by using elementary row* and *column transformations.* $\square$

Now it is apparent that for the matrix A' in (90) holds

$$\operatorname{rank}_{\mathrm{r}} A' = \operatorname{rank}_{\mathrm{c}} A' = r.$$

Since the elementary row and column transformations do not change the row and column rank of a matrix this means that

$$\operatorname{rank}_{\mathrm{r}} A = \operatorname{rank}_{\mathrm{r}} A' = \operatorname{rank}_{\mathrm{c}} A' = \operatorname{rank}_{\mathrm{c}} A,$$

that is row and column rank of a matrix agree. Thus we have shown the following result.

Theorem 2.49. *Let A be an arbitrary matrix with coefficients in a field F. Then the row rank of A is equal to the column rank of A, that is*

$$\operatorname{rank}_{\mathrm{r}} A = \operatorname{rank}_{\mathrm{c}} A. \qquad \square$$

Definition 2.50 (Rank of a Matrix)**.** Let A be a matrix with coefficients in a field F. Then the *rank* of A is defined to be the row rank (or equivalently the column rank) of A, in symbols

$$\operatorname{rank} A := \operatorname{rank}_{\mathrm{r}} A.$$

Let us collect the basic properties of the rank of a matrix:

(1) The rank of a matrix A is equal to the maximal number of linear independent rows of A and this number is at the same time also equal to the maximal number of linear independent columns of A.
(2) If A is the coefficient matrix of the system $u_1, \ldots, u_m$ of vectors of a vector space V with respect to an arbitrary basis of V, then
$$\operatorname{rank} A = \operatorname{rank}(u_1, \ldots, u_m).$$
(3) The rank of a matrix is invariant under elementary row and column transformations.
(4) Every matrix A over an arbitray field F can be transformed – using suitable elementary row and column transformations – into a matrix of the form
$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$
where r is the rank of the matrix A.

Algorithm for the Computation of Basis for a Subspace. As an application (and repetition) of our knowledge we have obtained so far we shall present an algorithm to compute the basis of a given subspace of a finite dimensional vector space V.

Let V be a vector space over a field F of dimension n and let $e_1, \ldots, e_n$ be a basis of V. Assume that a subspace U of V is given by the span of a system $u_1, \ldots, u_m$ of vectors in V, that is

$$U := \operatorname{span}\{u_1, \ldots, u_m\}.$$

Then the subspace U is determined by the coordinate matrix $A = (a_{ij})$ of $u_1, \ldots, u_m$ with respect to the basis $e_1, \ldots, e_n$:

$$u_i = \sum_{j=1}^{n} a_{ij} e_j \qquad (i = 1, \ldots, m).$$

Then it follows from Proposition 2.38 and Definition 2.50 that

$$\dim U = \operatorname{rank}(u_1, \ldots, u_m) = \operatorname{rank} A.$$

We compute the rank of the matrix A with the help of the Gauss Algorithm (see Chapter 1) using suitable elementary row transformations and column exchanges. By this we obtain finaly a matrix of the form

$$C := \begin{pmatrix} I_r & B \\ 0 & 0 \end{pmatrix} \tag{91}$$

where the matrix B is a certain $r \times (n-r)$-matrix over F. It follows that $r = \operatorname{rank} A = \dim U$.

But in addition to the dimension of U we have also also found a basis for U: let us denote by $b'_1, \ldots, b'_r$ the vectors of V of which the coordinate vectors are precisely the first r rows of the matrix C in (91). Then the system has rank r. If we revert the column exchanges which have been done to optain the matrix C in (91) then we get a system $b_1, \ldots, b_r$ of vectors *from the subspace* U which has rank $r = \dim U$. Thus the system $b_1, \ldots, b_r$ is a basis of U. Note that $b_1, \ldots, b_r$ is indeed a system of vectors of U since the elementary transformations done to the system $u_1, \ldots, u_m$ do not lead out of the subspace U.

9. Application to Systems of Linear Equations

Already in Chapter 1 we have seen how to decide whether a system of linear equations is solveable or not (see Section 4.2 in the previous chapter). The answer given there can be considered to be satisfying our needs. Nonetheless the progress we have made in this chapter about vector spaces gives us the possibility to get a deeper theoretical insight into systems of linear equations.

Assume that we are give an arbitrary system of linear equations of m equations in n unknown variables over a field F as in (6):

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\quad \vdots \quad \vdots \qquad\quad \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{92}$$

Denote by A the simple coefficient matrix of this system of linear equations and by C its extended coefficent matrix. Let $v_1, \ldots, v_n$ be the columns of the matrix A and denote by b the right most column of the extendet coefficient matrix. Then the m-tuples $v_1, \ldots, v_m$ and b are vectors of F^m. Now the system of linear equation (92) is solveable in F if and only if there exists elements $x_1, \ldots, x_n \in F$ such that

$$x_1v_1 + \ldots + x_nv_n = b.$$

This is again equivalent that $b \in \operatorname{span}\{v_1, \ldots, v_n\}$. The corrolary to Proposition 2.38 then states that this is the case if and only if

$$\operatorname{rank}(v_1, \ldots, v_n, b) = \operatorname{rank}(v_1, \ldots, v_n).$$

Now the first one is by definition the rank of the extended coefficient matrix C and the latter is the rank of the simple coefficent matrix A. Thus we have shown the following solvability criterion for systems of linear equations.

Proposition 2.51. *Let C denote the extended and let A denote the simple coefficient matrix of the system of linear equations* (92)*. Then* (92) *is solveable if and only if*

$$\operatorname{rank} C = \operatorname{rank} A. \qquad \square$$

We conclude this section by giving an **Answer to Problem 2** as it has been stated in the end of Chapter 1:

Proposition 2.52. *Let F be a field and U a subspace of the vector space F^n. Then there exists a homogeneous system of linear equatinons such that its solution space is exactly equal to U.*

PROOF. Due to Theorem 2.37 the subspace U has a finite generating system $u_1, \ldots, u_m$. Using elementary row transformations and column exchanges we can transform the coordinate matrix of $u_1, \ldots, u_k$ with respect to some basis $b_1, \ldots, b_n$ of V to a matrix of the form

$$C := \begin{pmatrix} I_s & B \\ 0 & 0 \end{pmatrix}$$

with some $0 \leq s \leq m, n$ where B is some $s \times (n-s)$-matrix. We consider the system of vectors obtained from the columns of the matrix

$$\begin{pmatrix} I_s \\ {}^tB \end{pmatrix}. \tag{93}$$

(Recall that ${}^tB$ denotes the transposed matrix of B as we have defined on page 45.) Reverting the column exchanges which have possible done to obtain the matrix C we see that we may assume that the subspace U has been given by a matrix of the form (93).

We consider now the homogeneous system of linear equations which has the coefficent matrix

$$(-{}^tB, I_{n-s}).$$

It follows from Proposition 1.9 from Chapter 1 that the columns of the matrix (93) is a basis – after possible reordering of the unknown variables – of the solution base. Therefor this solution space is equal to the given subspace U. □

CHAPTER 3

Linear Maps

1. Definition and Simple Properties

So far we have only studied vector spaces on their own. In this chapter we will study maps between vector spaces. Our interest will not lie in arbitrary maps but rather in maps which preserve the linear structure of vector spaces.

Note that in Appendix A there is given a short summary about basic mathematical terms which are frequently used when speaking about maps.

Definition 3.1 (Linear Map)**.** Let V and W be two vector spaces over the same field F. Then a map $f\colon V \to W$ is said to be a *linear map* if

$$f(x+y) = f(x) + f(y) \tag{94}$$

$$f(ax) = af(x) \tag{95}$$

for all vectors $x, y \in V$ and all $a \in F$.

Examples. (1) Let $f\colon V \to W$ be given by $f(x) := 0$ for every vector $x \in V$. Then the map f is linear since

$$f(x+y) = 0 = 0 + 0 = f(x) + f(y)$$

and

$$f(ax) = 0 = a0 = af(x)$$

for every $x, y \in V$ and $a \in F$. This linear map, which maps every vector of V to the zero vector 0 of W is called the *trivial linear map*.

(2) Let V be a F-vector space and let $c \in F$ be a fixed element. Then the map $f\colon V \to V$ given by $f(x) := cx$ for every vector $x \in V$ is linear since

$$f(x+y) = c(x+y) = cx + cy = f(x) + f(y)$$

and

$$f(ax) = c(ax) = a(cx) = af(x)$$

for every $x, y \in V$ and $a \in F$.

(3) Let F be a field and consider the F-vector space F^2. Let $a, b, c, d \in F$ some fixed elements. Then the map $f\colon F^2 \to F^2$ given by

$$f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} := \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}$$

is linear as can be easily verified.

(4) An example from calculus: consider the $\mathbb{R}$-vector space $C^\infty(\mathbb{R})$ of all infinite many times differentiable real valued function on $\mathbb{R}$. Then the differential operator

$$D\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}), f \mapsto D(f) := f'$$

which maps every function f to its derivative f' is linear because the differentiation rules are linear.

(5) Another example from calculus: consider the $\mathbb{R}$-vector space $C^0([0,1])$ of all continuous real valued functions defined on the unit interval $[0,1]$. Then the integral operator

$$I\colon C^0([0,1]) \to C^0([0,1]), f \mapsto I(f) := F$$

which maps every function f to the antiderivative F defined by $F(x) := \int_0^x f(t)dt$ is linear because the integration is linear.

(6) Let $f\colon V \to W$ be a map between two F-vector spaces. Then f is a linear map if and only if

$$f(ax + by) = af(x) + bf(y)$$

for every $x, y \in V$ and $a, b \in F$.

In mathematics maps between objects which are compatible with the mathematical structure of those objects are often called *homomorphism*. Now the properties (94) and (95) state that a linear map is compatible with the linear structure of a vector space and thus linear maps are also called *homomorphisms* of vector spaces.

Lemma 3.2. *Let V and W be vector spaces over the same field F and let $f\colon V \to W$ be a linear map. Then $f(0) = 0$ and $f(-x) = -f(x)$ for every $x \in V$.*

PROOF. This is verified by two simple calculations: $f(0) = f(0 \cdot 0) = 0 \cdot f(0) = 0$ and $f(-x) = f(-1 \cdot x) = -1 \cdot f(x) = -f(x)$. □

Proposition 3.3. *Let V, V' and V'' be vector spaces over the same field F. If $f\colon V \to V'$ and $g\colon V' \to V''$ are two linear maps, then their composite*

$$g \circ f\colon V \to V'', x \mapsto (g \circ f)(x) := g(f(x))$$

is also a linear map.

PROOF. Let $x, y \in V$. Then $(g \circ f)(x + y) = g(f(x + y)) = g(f(x) + f(y)) = g(f(x)) + g(f(y)) = (g \circ f)(x) + (g \circ f)(y)$. Similarly if $x \in V$ and $a \in F$, then $(g \circ f)(ax) = g(f(ax)) = g(af(x)) = ag(f(x)) = a(g \circ f)(x)$. Therefore both requirements of a linear map are satisfied for the composite map $g \circ f$ as had to be shown. □

Before we study linear maps in more detail we will classify them according to their mapping properties.

Definition 3.4. Let V and W be two vector spaces over the same field F. We call a injective linear map $f\colon V \to W$ a *monomorphism* and a surjective linear map is called a *epimorphism*. Finally an *isomorphism* is a bijective linear map.

Example. Let $u_1, \ldots, u_n$ be a system of vectors of an arbitrary vector space V over a field F. Then the map

$$f\colon F^n \to V, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_1u_1 + \ldots + x_nu_n$$

is a linear map from F^n to V. This map maps the canonical basis vectors $e_1, \ldots, e_n$ of F^n to the vectors $f(e_k) = u_k$ $(1 \leq k \leq n)$. The map f is apparently a monomorphism if and only if the vectors $u_1, \ldots, u_n$ are linearly independent. The map f is apparently an epimorphism if and only the $u_1, \ldots, u_n$ are a generating system of V. In other words f is an isomorphism if and only if $u_1, \ldots, u_n$ is a basis of V.

Note that in the above example we have essentially defined the map $f\colon F^n \to V$ by specifying the images of the standard basis vectors of F^n. More generaly the following result holds.

Proposition 3.5. *Let V and W be two vector spaces over the same field F. Let B be a basis of V and $f_0\colon B \to W$ an arbitrary map from the basis B to the vector space W. Then f_0 extends in a unique way to a linear map $f\colon V \to W$. That is, there exists a unique linear map $f\colon V \to W$ such that $f(b) = f_0(b)$ for every $b \in B$.*

Proof. Let us first prove the uniqueness. Assume that $f'\colon V \to W$ is an other linear map satisfying $f'(b) = f_0(b)$ for every $b \in B$. Let $v \in V$ be an arbitrary vector. We want to show that then $f(v) = f'(v)$. Since B is a basis ov V there exists a unique $x \in F^{(B)}$ such that

$$v = \sum_{b\in B}{}' x(b)b. \tag{96}$$

Then by the linearity of f and f' follows

$$f(v) = \sum_{b\in B}{}' f(x(b)b) = \sum_{b\in B}{}' x(b)f(b) = \sum_{b\in B}{}' x(b)f'(b) = \sum_{b\in B}{}' f'(x(b)b) = f'(v).$$

Therefore $f(v) = f'(v)$ for every $v \in V$ and this means $f = f'$. Thus the linear map f is unique if it exists.

It remains to show that there indeed exists a linear map $f\colon V \to W$ extending $f_0\colon B \to W$ in the requiret way. We do the proof in a constructive way by *defining* f in the following way: if $v \in V$ is an arbitrary vector, then we know that there exists a unique $x \in F^{(B)}$ such that (96) holds. Set

$$f(v) := \sum_{b\in B}{}' x(b)f_0(b).$$

Due to the uniquensess of x it follows that this defines a well defined map $f\colon V \to W$. We have to show that this map is linear and that it extends f_0. This is left as an exercise. □

In other words the above proposition states that a linear map $f\colon V \to W$ is completely characterized by the images of an arbitrary basis B of V.

Example. Assume that

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

is an arbitrary $m \times n$ matrix with coefficients in a field F. Denote by $v_1, \ldots, v_n$ the columns of the matrix A, which are then vectors of F^m. Then there exists precisely one linear map

$$f\colon F^n \to F^m$$

which maps the vectors $e_1, \ldots, e_n$ of the standard basis of F^n to the vectors $v_1, \ldots, v_n$. If $x \in F^n$ is an arbitrary vector

$$x = \sum_{i=1}^{n} x_i e_i = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

then the value $f(x)$ is given by

$$f(x) = \sum_{i=1}^{n} x_i v_i = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}$$

2. Isomorphisms and Isomorphism of Vector Spaces

Proposition 3.6. *Let V, V' and V'' be three vector spaces over the same field F. And let $f\colon V \to V'$ and $g\colon V' \to V''$ be two isomorphisms. Then the composite map $g \circ f\colon V \to V''$ and the inverse map $f^{-1}\colon V' \to V$ are isomorphisms.*

Proof. Left as an exercise. □

Definition 3.7 (Isomorphismus of Vector Spaces)**.** Let V and W be two vector spaces over the same field F. Then we say that V is *isomorphic* to W (as F-vector spaces) if there exists an isomorphism $f\colon V \to W$. If V and W are isomorphic vector spaces then we denote this fact in symbols by $V \cong W$.

Note the above defined relation has the following properties: If V, V' and V'' are vector spaces over the same field F, then

(1) $V \cong V$,
(2) $V \cong V' \Rightarrow V' \cong V$,
(3) $V \cong V'$ and $V' \cong V'' \Rightarrow V \cong V''$.

In mathematical terms the first property means that the relation "$\cong$" is *reflexive*, the second property means that the relation "$\cong$" is *symmetric* and the third property means that "$\cong$" is *transitive*. A relation which is reflexive, symmetric and transitive is also called a *equivalence relation*.

The importance of the concept of isomorphic vector spaces is apparent: if $f\colon V \to W$ is an isomorphism of vector spaces then every property of objects of V which is based on the vector space structure is transfered automatically to the coresponding images of those images in W. For example if $b_1, \ldots, b_n$ is a basis of V then $f(b_1), \ldots, f(b_n)$ is a basis of W. Such a claim does actually not need a seperate proof since it is a consequence of the nature of an isomorphism. In future we will not justify such kind of claims by more then saying "due to isomorphy reasons". For example due to isomorphy reason we have that if $u_1, \ldots, u_n$ are vectors of V, then

$$\operatorname{rank}(f(u_1), \ldots, f(u_n)) = \operatorname{rank}(u_1, \ldots, u_n).$$

Examples. (1) Let $v_1, \ldots, v_n$ be a linearly independent system of vectors in V and assume that $f\colon V \to W$ is an isomorphism of F-vector spaces. Then $f(v_1), \ldots, f(v_n)$ is a linearly independent system of vectors of W.

Proof: Since $v_1, \ldots, v_n$ is a linearly independent system of vectors it follos that they are all pairwise distinct vectors in V and since f is a monomorphism it follows that also $f(v_1), \ldots, f(v_n)$ are pairwise distinct vectors. Let $a_1, \ldots, a_n \in F$ such that

$$a_1 f(v_1) + \ldots + a_n f(v_n) = 0. \tag{97}$$

We want to show that this linear combination is infact the trivial linear combination of the zero vector. Due to the lienarity of f we have that

$$\begin{aligned} 0 &= a_1 f(v_1) + \ldots + a_n f(v_n) \\ &= f(a_1 v_1) + \ldots + f(a_n v_n) \\ &= f(a_1 v_1 + \ldots + a_n v_n). \end{aligned}$$

Now f is a monomorphism and since $f(0) = 0$ and $f(a_1v_1+\ldots+a_nv_n) = 0$ we get that $a_1v_1 + \ldots + a_nv_n = 0$. But since $v_1, \ldots, v_n$ was assumed to be a linearly independent system of vectors it follows that $a_1 = \ldots = a_n = 0$. That is, the above linear combination (97) is actually the trivial linear combination of the zero vector. Thus $f(v_1), \ldots, f(v_n)$ is a linearly independent system of vectors in W. □

Note that we used in this proof only the fact that an isomorphism is a monomorphism! Thus the result is also true if f is only a monomorphism.

(2) Let $v_1, \ldots, v_n$ be a generating system of V and assume that $f\colon V \to W$ is an isomorphism of F-vector spaces. Then $f(v_1), \ldots, f(v_n)$ is a generating system of W.

Proof: We have to show that every $y \in W$ can be written as a linear combination of the vectors $f(v_1), \ldots, f(v_n)$. Therefore let $y \in W$ be an arbitrary vector. Since f is an epimorphism it follows that there exist $x \in V$ such that $y = f(x)$. Since $v_1, \ldots, v_n$ is a generating system of V it follows that we can write

$$x = a_1v_1 + \ldots + a_nv_n$$

with some elements $a_1, \ldots, a_n \in F$. But then by the linearity of f we have that

$$\begin{aligned} y &= f(x) \\ &= f(a_1v_1 + \ldots + a_nv_n) \\ &= f(a_1v_1) + \ldots + f(a_nv_n) \\ &= a_1f(v_1) + \ldots + a_nf(v_n) \end{aligned}$$

and thus y is a linear combination of the vectors $f(v_1), \ldots, f(v_n)$. Since this is true for any $y \in W$ it follows that $f(v_1), \ldots, f(v_n)$ is a generating system of W. □

Note also here, that we needed not the full "power" of an isomorphism. We just used the fact that an isomorphism is an epimprphism. Thus the result is also true if f is just an epimorphism.

(3) Let $v_1, \ldots, v_n$ be a basis of V and assume that $f\colon V \to W$ is an isomorphism of F-vector spaces. Then $f(v_1), \ldots, f(v_n)$ is a basis of W.

Proof: This follows now from the previous two results since a basis is a linear independent generating system. □

Note that here we need the full "power" of an isomorphism: a linear map f between two isomorphic vector spaces – in this case V and W – maps a basis of V to a basis of W if and only if f is an isomorphism.

Proposition 3.8. *Let V be a n-dimensional vector space over the field F. Let $B = (b_1, \ldots, b_n)$ be an ordered basis of V. Then the map*

$$i_B\colon F^n \to V, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_1b_1 + \ldots + x_nb_n$$

is an isomorphism of vector spaces.

PROOF. This follows straight from the considerations of the example on page 52 in the previous section. □

We call the isomorphism i_B of the previous proposition the *basis isomorphism* of V with respect to the basis B. Its inverse

$$c_B\colon V \to F^n, x_1b_1 + \ldots + x_nb_n \mapsto \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

is called the *coordinate isomorphism* of V with respect to the basis B.

We can now prove a very strong result about finite dimensional vector spaces over the same field F, namely that they are characterized by their dimension.

Proposition 3.9. *Let V and W be two finite dimensional vector spaces over the same field F. Then*

$$V \cong W \iff \dim V = \dim W.$$

PROOF. "$\Rightarrow$": If V and W are two isomorphic vector spaces then they have apparently the same dimension, that is $\dim V = \dim W$.[1]

"$\Leftarrow$" Assume that $\dim V = \dim W$, say both vector spaces are n-dimensional. Then V has a finite basis B with n elements and W has a finite basis C with n elements. Then we have the coordinate isomorphism $c_B\colon V \to F^n$ of V with respect to the basis B and the basis isomorphism $i_C\colon F^n \to W$ of W with respect to the basis C. Its composite is apparently an isomorphism

$$i_C \circ c_B\colon V \to W.$$

Therefore $V \cong W$. □

Note that in general it is *not* (!) true that if $\dim V = \infty$ and $\dim W = \infty$, that then $V \cong W$. But one can show that $V \cong W$ if and only if there exists a bijective map between a basis of V and a basis of W.

3. Dimension Formula for Linear Maps

Definition 3.10. Let V and W be vector spaces over the same field F and let $f\colon V \to W$ be a linear map. Then the *kernel* of f is the set

$$\ker f := \{x \in V : f(x) = 0\}$$

and the *image* of f is the set

$$\operatorname{im} f := \{y \in W : \text{there exists a } x \in V \text{ such that } f(x) = y\}.$$

The image and the kernel of a linear map $f\colon V \to W$ are not just subsets of V and W, but share even more structure. We have namely the following result.

Proposition 3.11. *Let V and W be vector spaces over the same field F and let $f\colon V \to W$ be a linear map. Then $\ker f$ is a subspace of V and $\operatorname{im} f$ is a subspace of W.*

PROOF. We show first that $\ker f$ is a subspace of V. First note that $f(0) = 0$ and therefore $0 \in \ker f$. In partitcular $\ker f$ is not empty. If $x_1, x_2 \in \ker f$, then $f(x_1 + x_2) = f(x_1) + f(x_2) = 0 + 0 = 0$ and thus $x_1 + x_2 \in \ker f$. If $x \in \ker f$ and $a \in F$, then $f(ax) = af(x) = a0 = 0$ and thus also $ax \in \ker f$. Thus all the requirements of the supspace criterion of Proposition 2.6 are staisfied and it follows that $\ker f$ is a subspace of V.

Next we show that $\operatorname{im} f$ is a subspace of W. First of all it is clear that $\operatorname{im} f$ is not empty since V is not empty. Next, if $y_1, y_2 \in \operatorname{im} f$, then there exists $x_1, x_2 \in V$ such

[1] Note that this holds also in the case that V and W are not finite dimensional vector spaces.

that $f(x_1) = y_1$ and $f(x_2) = y_2$. Then $y_1 + y_2 = f(x_1) + f(x_2) = f(x_1 + x_2) \in \operatorname{im} f$ since $x_1 + x_2 \in V$. If $y \in \operatorname{im} f$ and $a \in F$ then there exists a $x \in V$ such that $f(x) = y$ and we have that $ay = af(x) = f(ax) \in \operatorname{im} f$ since $ax \in V$. Thus all the requirements of the supspace criterion of Proposition 2.6 are staisfied and it follows that $\operatorname{im} f$ is a subspace of W. □

Proposition 3.12. *Let V and W be vector spaces over the same field F and let $f\colon V \to W$ be a linear map. Then the following two statements hold:*

(1) *f is a monomorphism if and only if* $\ker f = 0$.
(2) *f is an epimorphism if and only if* $\operatorname{im} f = W$.

PROOF. Note that the second claim is just the definition of an epimorphism. Thus we need only to prove the first claim.

"⇒": Let $x \in \ker f$. Then $f(x) = 0 = f(0)$ but this means that $x = 0$ since f is assumed to be monomorphism. Therefore $\ker f = \{0\}$ is the trival zero vector space.

"⇐": Assume that $\ker f = \{0\}$ and let $x, y \in V$ such that $f(x) = f(y)$. Then $f(x - y) = f(x) - f(y) = 0$ and thus $x - y \in \ker f$. But since $\ker f = \{0\}$ this means that $x - y = 0$ and this again is equivalent with $x = y$. Therefore f is a monomorphism. □

Note that the above proposition simplifies in a very essential way the verification whether a linear map is a monomorphism or not. By definition $f\colon V \to W$ is a monomorphism if for *every* $x, y \in V$ from $f(x) = f(y)$ it follows that $x = y$. Now using the fact that f is a linear map it it enough to just verify that $f(x) = 0$ implies that $x = 0$!

Proposition 3.13. *Let $f\colon V \to W$ be a linear map of F-vector spaces. Assume that V is an n-dimensional vector space and that $b_1, \ldots, b_n$ is a basis of V. If one sets*

$$u_i := f(b_i)$$

for $1 \leq i \leq n$, then

$$\dim(\operatorname{im} f) = \operatorname{rank}(u_1, \ldots, u_n).$$

PROOF. Due to the linearity of f we have for arbirary elements $a_1, \ldots, a_n \in F$ the equality

$$f\Big(\sum_{i=1}^{n} a_i b_i\Big) = \sum_{i=1}^{n} a_i f(b_i) = \sum_{i=1}^{n} a_i u_i.$$

It follows that $\operatorname{im} f = \operatorname{span}(u_1, \ldots, u_n)$. Thus the claim follows from Proposition 2.38 of Chapter 2. □

Definition 3.14. Let V and W be arbitrary vector spaces (finite or infinite dimensional) over the same field F. Let $f\colon V \to W$ be a linear map. Then the *rank* of f is defined to be

$$\operatorname{rank} f := \dim(\operatorname{im} f).$$

Note that the above definition explicitely allows the case that $\operatorname{rank} f = \infty$. Further we have apparently the following two constraints on the rank of a linear map:

$$\operatorname{rank} f \leq \dim W \qquad \text{and} \qquad \operatorname{rank} f \leq \dim V.$$

If W is a finite dimensional vector space then we have apparently

$$\operatorname{rank} f = \dim W \iff f \text{ is an epimorphism.}$$

If V is a finite dimensional vector space then one can show easily (left as an exercise) that

$$\operatorname{rank} f = \dim V \iff f \text{ is a monomorphism.}$$

We will soon see that there is a connection between the rank of a linear map f and the rank of a matrix A as defined in Chapter 2. This will then make the above definition a very natural one (see Proposition 3.25).

Theorem 3.15 (Dimension Formula for Linear Maps)**.** *Let V and W be vector spaces over the same field F. Assume that V is finite dimensional and let $f\colon V \to W$ be a linear map. Then $\ker f$ and $\operatorname{im} f$ are finite dimensional subspaces of V and W respectively and we have the equality*

$$\dim V = \dim(\operatorname{im} f) + \dim(\ker f). \tag{98}$$

PROOF. First of all note that $\ker f$ is finite dimensional since it is a subspace of a finite dimensional space. Then there exists a linear complement U of $\ker f$ by Theorem 2.40, that is there exists a subspace U of V such that

$$V = U + \ker f \qquad \text{and} \qquad U \cap \ker f = 0.$$

Let $g\colon U \to W$ the restriction of f to the subspace U, that is let g be the linear map

$$g\colon U \to W, v \mapsto g(v) := f(v).$$

Apparently we have for the image and the kernel of g the following relations:

$$\operatorname{im} g = \operatorname{im} f \qquad \text{and} \qquad \ker g = 0.$$

It follows that g defines an isomorphism $U \cong \operatorname{im} f$. Since U is a subspace of the finite dimensional space V it follows that $\operatorname{im} f$ is finite dimensional, too. Furthermore we have due to the second part of Proposition 2.43

$$\begin{aligned} \dim V &= \dim U + \dim(\ker f) \\ &= \dim(\operatorname{im} f) + \dim(\ker f). \end{aligned}$$

□

Corollary. *We can express the dimension formula for linear maps* (98) *also in the form*

$$\operatorname{rank} f = \dim V - \dim(\ker f).$$

□

Note that the above corollary explains why the number $\dim(\ker f)$ is also called the *defect* of the linear map f. Furthermore the proof of the dimension formular verifies also the following useful result:

Proposition 3.16. *Let V and W be arbitrary vector spaces over the field F and let $f\colon V \to W$ be a linear map. If U is a linear complement of $\ker f$ then f maps U isomorphicaly onto $\operatorname{im} f$.* □

Again we will show that linear maps between finite dimensional have a very rigid behavior. We have the following result which is true for finite dimensional vector spaces but *not* for infinite dimensonal vector spaces.

Proposition 3.17. *Let V and W be two vector spaces over the same field F. Assume that V and W have the same finite dimension. If $f\colon V \to W$ is a linear map then the following three statements are equivalent:*

(1) *f is an isomorphism.*
(2) *f is a monomorphism.*
(3) *f is an epimorphism.*

PROOF. Left as an exercise. □

4. The Vector Space $\mathrm{Hom}_F(V,W)$

In this section we will study the structure of linear maps which is essential to the deeper understanding of Linear Algebra. Linear maps are structure preserving maps between vector spaces which map vectors to vectors. But they can also be seen as vectors of suitable vector spaces.

Let V and W vector spaces over the same field F. Then we can consider the set W^V of all maps from V to W. In the same way as we defined in example 5 on page 23 a vector space structure on the set F^I we can define a vector space structure on the set W^V. If $f, g \in W^V$ are two maps $V \to W$, then we define $f+g$ to be the map which is given by

$$(f+g)(v) := f(v) + g(v) \tag{99}$$

for every $v \in V$. Similar if $f \in W^V$ and $a \in F$, then af shall denote the map which is given by

$$(af)(v) := af(v) \tag{100}$$

for every $v \in V$. Apparently $f+g \in W^V$ and $af \in W^V$ and one convinces oneself that with this addition and scalar multiplication W^V becomes indeed a F-vector space.

Proposition 3.18. *Let f, g: $V \to W$ linear maps and $a \in F$. Then both $f+g$ and af are linear maps. In particular the set of all linear maps $V \to W$ is a subspace of W^V.*

Proof. If $u, v \in V$ and $b \in F$, then

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

and

$$\begin{aligned}(f+g)(bv) &= f(bv) + g(bv)\\ &= bf(v) + bg(v)\\ &= b(f(v) + g(v))\\ &= b(f+g)(v)\end{aligned}$$

and therefore $f+g$ is a linear map. Furthermore

$$\begin{aligned}(af)(u+v) &= af(u+v)\\ &= a(f(u) + f(v))\\ &= af(u) + af(v)\\ &= (af)(u) + (af)(v)\end{aligned}$$

and

$$\begin{aligned}(af)(bv) &= af(bv)\\ &= abf(v)\\ &= b(af)(v)\end{aligned}$$

and thus also af is a linear map.

The remaining claim that the set of all linear maps $V \to W$ is a subspace of W^V follows now using the subspace criterion from the above considerations and the fact that there exists always at least the trivial linear map $V \to W$. □

Definition 3.19. Let V and W be vector spaces over the same field F. Then we denote by

$$\mathrm{Hom}_F(V, W)$$

the F-vector space of all linear maps from V to W.

Note that if there is no danger of confusion about the underlying field, then we might ommit the field in the notation and just write $\mathrm{Hom}(V, W)$ instead of $\mathrm{Hom}_F(V, W)$.

We come now to another essential point in this section. Let V, V' and V'' be vector spaces over the same field F. If $f \in \mathrm{Hom}_F(V, V')$ and $g \in \mathrm{Hom}_F(V', V'')$ then the composite map $g \circ f$ is by Proposition 3.3 a linear map

$$g \circ f \colon V \to V'',$$

that is $g \circ f \in \mathrm{Hom}_F(V, V'')$. Thus we get a map

$$\circ \colon \mathrm{Hom}_F(V', V'') \times \mathrm{Hom}_F(V, V') \to \mathrm{Hom}_F(V, V''), (g, f) \mapsto g \circ f.$$

This map assigns each pair (g, f) with $f \in \mathrm{Hom}_F(V, V')$ and $g \in \mathrm{Hom}_F(V', V'')$ the composite map $g \circ f \in \mathrm{Hom}_F(V, V'')$. We say that $g \circ f$ is the *product* of f and g. If there is no danger of confusion we write also gf instead of $g \circ f$:

$$gf := g \circ f. \tag{101}$$

The so defined "multiplication" of linear maps satisfies rules which are very similar to the calculation rules we would expect from a multiplication. First of all the following *distributive laws* hold:

$$g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2 \tag{102}$$

$$(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f \tag{103}$$

where $f, f_1, f_2 \in \mathrm{Hom}_F(V, V')$ and $g, g_1, g_2 \in \mathrm{Hom}_F(V', V'')$. If furthermore $h \in \mathrm{Hom}_F(V'', V''')$, then the following *associative law* holds:

$$h \circ (g \circ f) = (h \circ g) \circ f \tag{104}$$

Moreover the multiplication of linear maps is in the following way compatible with the scalar multiplication. If $a \in F$ is an arbitrary element of the field F, then

$$a(g \circ f) = (ag) \circ f = g \circ (af). \tag{105}$$

For every vector space V we have the identity map $\mathrm{id}_V \colon V \to V$ which is always a linear map and thus $\mathrm{id}_V \in \mathrm{Hom}_F(V, V)$. For every linear map $f \colon V \to V'$ holds:

$$\mathrm{id}_{V'} \circ f = f \circ \mathrm{id}_V . \tag{106}$$

If there is no danger of confusion then we may leave away the index V of id_V and the above relation (106) reads then

$$\mathrm{id} \circ f = f \circ \mathrm{id} . \tag{107}$$

Note that the rules (104) and (106) are true in general for arbitrary maps. The rules (102), (103) and (105) are verified easily if one applies both sides of the equality to an arbitrary element of $x \in V$ and concludes that both sides have the same element as the image. For example (102) is verified as follows:

$$\begin{aligned}(g \circ (f_1 + f_2))(x) &= g((f_1 + f_2)(x)) = g(f_1(x) + f_2(x)) = g(f_1(x)) + g(f_2(x)) \\ &= (g \circ f_1)(x) + (g \circ f_2)(x) = ((g \circ f_1) + (g \circ f_2))(x).\end{aligned}$$

We shall explicitely note that the product of two (linear) maps $f\colon V \to V'$ and $g\colon W \to W'$ is *only* defined if $V' = W$!

Now let us turn to the special case of the vector space $\mathrm{Hom}_F(V,V)$. In this case the the product $fg := f \circ g$ of two elements $f, g \in \mathrm{Hom}_F(V,V)$ is always defined. In addition to the addition "+" on $\mathrm{Hom}_F(V,V)$ we can also define a *multiplication* "$\circ$" on $\mathrm{Hom}_F(V,V)$:

$$\circ\colon \mathrm{Hom}_F(V,V) \times \mathrm{Hom}_F(V,V) \to \mathrm{Hom}_F(V,V), (f,g) \mapsto fg := f \circ g \qquad (108)$$

Proposition 3.20. *Let V be a vector space over the field F. Then the* $\mathrm{Hom}_F(V,V)$ *becomes a ring (with unit) under the addition "+" and the above defined multiplication "$\circ$".*

Proof. Recall that we have defined in the previous chapter on page 21 a ring to be a set together with an addition and multiplication which satisfies all field axioms except (M2) and (M4).

Now the vector space axioms (A1) to (A4) are precisely the field axioms (A1) to (A4). The calculation rule (104) verifies the field axiom (M1). The identity map id_V is the identity element of the multiplication by the calculation rule (106) and this verifies the field axiom (M3). Finally the field axiom (D) is verified by the calculation rule (102) and (103). □

Definition 3.21. An linear map $f\colon V \to V$ is said to be an *endomorphism*. The ring $\mathrm{Hom}_F(V,V)$ is also denoted by

$$\mathrm{End}_F(V) := \mathrm{Hom}_F(V,V)$$

and called the *endomorphism ring* of the F-vector space V.

Note that we may ommit the field F in the above notation in case that there is no danger of confusion, that is we may write $\mathrm{End}(V)$ instead of $\mathrm{End}_F(V)$. Note further that $\mathrm{End}_F(V)$ is in general not commutative nor is $\mathrm{End}_F(V)$ in general zero divisor free.

Note that in Proposition 3.20 and in Definition 3.21 we have yet not paid attention to the fact that we have also a scalar multiplication defined on $\mathrm{End}_F(V) = \mathrm{Hom}_F(V,V)$ for which the calculation rule (105) is satisfies.

Definition 3.22 (F-Algebra)**.** Let F be a field. A ring R with unit which is at the same time also a F-vector space such that

$$a(gf) = (ag)f = g(af)$$

for every $a \in F$, $g, f \in R$, is called an *algebra (with unit) over F* or just an *F-algebra*.

With this definition and since the endomorphism ring $\mathrm{End}_F(V)$ is also a F-vector space which satisfies the calculation rule (105) we can summerize the results of this section in the follwing compact form.

Proposition 3.23. *Let V be a vector space over the field F. Then the endomorphism ring* $\mathrm{End}_F(V)$ *is in a natural way an F-algebra.* □

5. Linear Maps and Matrices

We come now to a central part of Linear Algebra, namely the matrix description of a linear map. We will only consider the case of linear maps between finite dimensional vectorspaces. Therefore let V and W be finite dimensional vector spaces over the same field F, say

$$\dim V = n \qquad \text{and} \qquad \dim W = m$$

for some natural numbers n and m. Furthermore let

$$f\colon V \to W$$

be a linear map from V to W. We fix an ordered basis $B = (b_1, \ldots, b_n)$ of V and a ordered basis $C = (c_1, \ldots, c_m)$ of W. By Proposition 3.5 we know that the linear map f is completely determined by the images

$$a_i := f(b_i) \qquad (i = 1, \ldots, n)$$

of the basis vectors $b_1, \ldots, b_n$ under the map f. We can describe the images $a_1, \ldots, a_n \in W$ as unique linear combinations of the basis vectors $c_1, \ldots, c_m$ of W. We have then

$$f(b_i) = a_i = \sum_{j=1}^{m} a_{ji} c_j \qquad (i = 1, \ldots, n) \tag{109}$$

with some *unique* elements $a_{ji} \in F$. Therefore – after we have fixed a basis B for V and a basis C for W – the linear map $f\colon V \to W$ is completely determined by the matrix

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{110}$$

It is a $m \times n$-matrix with coefficients in the field F. Due to (109) the colums of A are precisely the *coordinate vectors* of $a_1 = f(b_1), \ldots, a_n = f(b_n)$ with respect to the basis $c_1, \ldots, c_m$ of W.

Definition 3.24 (Coordinate Matrix of a Linear Map)**.** Using the assumptions and notation introduced above the $m \times n$-matrix A defined by (109) is called the *coordinate matrix* of the linear map $f\colon V \to W$ with respect to the bases $B = (b_1, \ldots, b_n)$ and $C = (c_1, \ldots, c_m)$ of V and W.

Note that in the previous chapter we did arrange the coordinate coefficients in rows (see page 36). Since we use to write vectors in F^n in columns it is customary to call the matrix A as defined above (and not its transpose ${}^tA$ as we did in the previous chapter) the coordinate matrix of the system $a_1, \ldots, a_n$ with respect of the basis $c_1, \ldots, c_m$. With this changed definition we can say that the coordinate matrix of the linear map $f\colon V \to W$ with respect to the bases B and C of the vector spaces V and W respectively is nothing else than the coordinate matrix of the system of images $f(b_1), \ldots, f(b_n)$ with respect to the basis C of W.

Example. Consider the map $f\colon \mathbb{R}^5 \to \mathbb{R}^4$ given by

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + 3x_2 + 5x_3 + 2x_4 \\ 3x_1 + 9x_2 + 10x_3 + x_4 + 2x_5 \\ 2x_2 + 7x_3 + 3x_4 - x_5 \\ 2x_1 + 8x_2 + 12x_3 + 2x_4 + x_5 \end{pmatrix}$$

This map is apparently linear and its coordinate matrix with respect to the standard bases of $\mathbb{R}^5$ and $\mathbb{R}^4$ is

$$A = \begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 3 & 9 & 10 & 1 & 2 \\ 0 & 2 & 7 & 3 & -1 \\ 2 & 8 & 12 & 2 & 1 \end{pmatrix}$$

Note that sometimes a "too precise" notation is more hindering than helpful. If we denote the coordinate matrix of the linear map f by $c(f)$, then we may express the dependency of the coordinate matrix on the bases B and C in symbols by

$$c_C^B(f). \tag{111}$$

But *much more important* (!) than this notation is that one is *in principle* aware of the dependency of the coordinate matrix on the choice of bases!

Example. Let us return to the previous example. Let B be the standard basis of $\mathbb{R}^5$ and consider the following basis C of $\mathbb{R}^4$:

$$c_1 := \begin{pmatrix} 1 \\ 3 \\ 0 \\ 2 \end{pmatrix}, \quad c_2 := \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad c_3 := \begin{pmatrix} 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad c_4 := \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

Then the coordinate matrix of the linear map f with respect to the bases B and C calculates to

$$A' = \begin{pmatrix} 1 & 3 & 5 & 2 & 0 \\ 0 & 2 & 2 & -2 & 1 \\ 0 & 0 & 5 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

(compare this with the example on page 10 in Chapter 1).

We are now able to explain the connection of the rank of a linear map f as defined in Definition 3.14 and the rank of a matrix as defined in Chapter 2.

Proposition 3.25. *Let V and W be two finite dimensional vector spaces. Let B and C be bases of V and W respectively. Let $f\colon V \to W$ be a linear map and denote by A the coordinate matrix of f with respect to the bases B and C. Then*

$$\operatorname{rank} f = \operatorname{rank} A.$$

Proof. We have

$$\begin{aligned} \operatorname{rank} f &= \dim(\operatorname{im} f) && \text{(Definition 3.14)} \\ &= \operatorname{rank}(f(b_1), \ldots, f(b_n)) && \text{(Proposition 3.13)} \\ &= \operatorname{rank}(a_1, \ldots, a_n) && (a_i = f(b_i) \text{ for } 1 \leq i \leq n) \\ &= \operatorname{rank} A. && \text{(Proposition 2.45 together} \\ & && \text{with Definition 2.50)} \end{aligned}$$

□

We have seen that a linear map $f\colon V \to W$ of an n-dimensional F-vector space V to an m-dimensional F-vector space W is – after choosing bases for V and W – completely determined by a $m \times n$-matrix. In order to be able to describe the connection between linear maps and their coordinate matrices more precisely we shall first study matrices a bit more in general.

Sofar we have just said that a $m \times n$-matrix over a field F is a collection of elements of F arranged into a rectangle with m rows and n columns. We shall give now a more precise definition.

Definition 3.26 (Matrix)**.** Let F be a field and $m, n \geq 1$ some natural numbers. Denote by M and N the sets $M := \{1, \ldots, m\}$ and $N := \{1, \ldots, n\}$. By an *$m \times n$-matrix A over F* we understand a map

$$A\colon M \times N \to F, (i, j) \mapsto a_{i,j}$$

which asigns each pair (i, j) of natural numbers $1 \leq i \leq m$ and $1 \leq j \leq n$ an element $a_{i,j}$ of the field F.

The set of all $m \times n$-matrices over the field F is denoted by

$$F^{m,n} := F^{M \times N}. \tag{112}$$

Note that if there is no danger of confusion, then we may write also a_{ij} instead of $a_{i,j}$.

From (112) follows that the set $F^{m,n}$ obtains in a natural way a F-vector space structure (see Example 5 on page 23 in the previous chapter). The addition and scalar multiplication of $m \times n$-matrices over F are therefore defined *coefficientwise*. That is

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & & & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

and if $c \in F$ then

$$c \cdot \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & & & \vdots \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{pmatrix}$$

If one assigns every $m \times n$-matrix A as in (110) the mn-tuple

$$(a_{11}, \ldots, a_{1n}, a_{21}, \ldots, a_{2n}, \ldots, a_{m1} \ldots, a_{mn}) \in F^{mn}$$

then one ovtains apparently an isomorphism of F-vector spaces. Thus we have

$$F^{m,n} \cong F^{mn}.$$

In particular we have that

$$\dim F^{m,n} = mn.$$

Thus we can now say about the close connection between linear maps of finite dimensional vector spaces and matrices the following.

Theorem 3.27. *Let V and W be finite dimensional vector spaces over the same field F with $\dim V = n$ and $\dim W = m$. Let B be a basis of V and C be a basis of W. If we assign each linear map $f\colon V \to W$ its coordinate matrix $c(f)$ with respect to the bases B and C, then we obtain an isomorphism*

$$c\colon \operatorname{Hom}_F(V, W) \to F^{m,n}$$

of the F-vector space of all linear maps $V \to W$ onto the F-vector space of all $m \times n$-matrices over F.

Proof. We known already that the coefficients of the coordinate matrix of a linear map $f\colon V \to W$ describes the linear map f in a unique way. From this follows that c is an injective map. The linearity of the map c is evident. And that the map c is surjective follows from Proposition 3.5. □

Corollary. *If V is a F-vector space of dimension n and if W is a F-vector space of dimension m, then*

$$\dim(\operatorname{Hom}_F(V, W)) = mn. \qquad \square$$

The isomorphism from Theorem 3.27 gives a one-to-one correspondence

$$\mathrm{Hom}_F(V,W) \cong F^{m,n} \tag{113}$$

but since the isomorphism c and therefore the one-to-one correspondence depends essentially on the choice of the bases B and C for the vector spaces V and W this isomorphism is not suitable for identifying linear maps $f\colon V \to W$ with their coordinate matrices. There is general simply no canonical way to choose one set of bases over another one.

The situation is different if one considers linear maps from F^n to F^m. For those vector spaces exists a natural choice for a basis, namely the standard basis of F^n and F^m. Therefore we can and do identify the linear maps $f\colon F^n \to F^m$ with their coordinate matrices $A := c(f)$ with respect to the standard basis of F^n and F^m. Using this identification we get the equality

$$\mathrm{Hom}_F(F^n,F^m) = F^{m,n}. \tag{114}$$

Note the difference between (113) and (114): the first one is an isomorphism of vector spaces, the latter an equality (after identification). The equality of (114) is much stronger than the isomorphism of (113)!

Using this identification we can express the definition of the coordinate matrix of a linear map in the special case of $V = F^n$ and $W = F^m$ in the following way.

Proposition 3.28. *The columns of a $m\times n$-matrix A over F are in turn the images of the canonical unit vectors $e_1,\dots,e_n$ of F^n under the linear map $A\colon F^n \to F^m$.* □

We can visualize the content of Theorem 3.27 with the help of the following "*commutative diagram*"

$$\begin{array}{ccc} V & \xrightarrow{\ f\ } & W \\ {\scriptstyle i_B}\uparrow & & \uparrow{\scriptstyle i_C} \\ F^n & \xrightarrow{\ A\ } & F^m \end{array} \tag{115}$$

where the vertical arrows denote the basis isomorphisms i_B and i_C for the chosen bases B for V and C for W (see page 56). That the diagram (115) *commutes* means that regardless which way one follows the arrows in (115) from F^n to W the corresponding composite maps agree. That is

$$f \circ i_B = i_C \circ A.$$

One proves this by applying both sides of this equality to an arbitrary vector e_i of the standard basis of F^n and one obtains the equality (109).

Note that from (115) follows that

$$A = i_C^{-1} \circ f \circ i_B$$

and

$$f = i_C \circ A \circ i_B^{-1}.$$

and if one use the coordinate isomorphisms c_B and c_C, then

$$A = c_C \circ f \circ c_B^{-1}$$

or if one wants to use the notation introduced by (111) then one has

$$c_C^B(f) = c_C \circ f \circ c_B^{-1}.$$

Now since we identified the vector space of $m \times n$-matrices over a field F with the vector space $\mathrm{Hom}_F(F^n, F^m)$ of all linear maps $F^n \to F^m$ the natural questions arises how to calculate the image of a vector $x \in F^n$ under a given matrix $A = (a_{ij})$. The answer to this question is given in the following result.

Proposition 3.29. *Let $A = (a_{ij})$ be a $m \times n$-matrix over the field F. Then we can describe the linear map A: $F^n \to F^m$ explicitely in the following way: assume that*

$$y = A(x) \qquad \text{with } x \in F^n \text{ and } y \in F^m.$$

Then we can determine the coordinates y_i of y using the coordinates x_j of x by the following formula

$$y_i = \sum_{j=1}^{n} a_{ij} x_j \qquad i = 1, \ldots, m. \tag{116}$$

Proof. Let $e_1, \ldots, e_n$ be the canonical vectors of the standard basis of F^n. By definition we have then

$$A(e_j) = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}, \qquad j = 1, \ldots, n.$$

If we apply the linear map A to the element

$$x = \sum_{j=1}^{n} x_j e_j$$

then we get due to the linearity of A

$$y = A(x) = A\Big(\sum_{j=1}^{n} x_j e_j\Big) = \sum_{j=1}^{n} x_j A(e_j)$$

and thus

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \sum_{j=1}^{n} x_j \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = \sum_{j=1}^{n} \begin{pmatrix} a_{1j} x_j \\ \vdots \\ a_{mj} x_j \end{pmatrix}$$

From this one can then read then the claim. □

Compare (116) with (109) on page 62!

Note that we can write now a linear system of equations in a very compact way. Assume that we have given a non-homogeneous system of m linear equations in n unknown variables

$$x_1 v_1 + \ldots + x_n v_n = b.$$

Let A be the simple coefficient matrix of this system, which is then a $m \times n$-matrix over some field F. Then we can write the system of linear equations in the compact form

$$A(x) = b.$$

Apparently the system is solveable if and only if $b \in \operatorname{im} A$. In case $b = 0$ the solutions M form a linear subspace of F^n, namely $M = \ker A$, and its dimension is $\dim(\ker A)$. In the end of this chapter – see Section 12 on page 88 – we will have a closer look on the content of Chapter 1 in this new view.

6. The Matrix Product

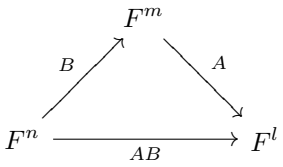
We come now to a complete new point and it will be the first time that we go esentially beyond the content of Chapter 1. Because of (114) we are able to define a *product of matrices* of suitable dimensions.

Definition 3.30 (Matrix Product)**.** Let F be a field and $l, m, n \geq 1$ natural numbers. For $A \in F^{l,m}$ and $B \in F^{m,n}$ we define the *product* AB of A and B to be the composite map

$$AB := A \circ B$$

which is an element of $F^{l,n}$.

Note that the content of the above definition can be visualized with the following commutative diagram:



Again arises as a natural question how we can compute the the coefficents of the product matrix AB using the coeffcients of the matrices A and B. The answer is given by the following theorem.

Theorem 3.31 (Formula for the Matrix Product)**.** *Let $A = (a_{qp}) \in F^{l,m}$ and $B = (b_{rs}) \in F^{m,n}$ be two given matrices. Let $C := AB \in F^{l,n}$ be the product matrix of A and B. Then the coeffcients c_{ij} of the matrix C are given by the formula*

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}, \qquad 1 \leq i \leq l, 1 \leq j \leq n. \tag{117}$$

Proof. Let $e_1, \ldots, e_n$ the vectors of the standard basis of F^n. By definition we have then

$$C(e_j) = \begin{pmatrix} c_{1j} \\ \vdots \\ c_{lj} \end{pmatrix}, \qquad j = 1, \ldots, n. \tag{118}$$

We have to determine the images $C(e_j)$ of e_j under the linear map $C = AB$. By the definition of the matrix product we have

$$C(e_j) = A(B(e_j)).$$

For the $B(e_j)$ we have – analogous to (118) – by definition

$$B(e_j) = \begin{pmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{pmatrix}, \qquad j = 1, \ldots, n. \tag{119}$$

We can now use the result of Proposition 3.29 to determine the coordinates of the images $C(e_j)$ of the vectors $B(e_j)$ under the linear map A. Using (116) on (119) yields indeed

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}, \qquad 1 \leq i \leq l$$

and this is true for $1 \leq j \leq n$. Thus we have proven (117). □

Note that the matrix product AB of A with B is *only* (!) defined if the number of rows of the matrix A is equal to the number of columns of the matrix B.

If $A\colon F^n \to F^m$ is a linear map and

$$y = A(x)$$

with

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in F^n, \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in F^m, \tag{120}$$

then by Proposition 3.29 we have the equalities

$$y_i = \sum_{i=1}^{n} a_{ij} x_k, \qquad i = 1, \ldots, m. \tag{121}$$

One can consider the vectors x and y in (120) as matrices of a special form, namely x can be considered as a $n \times 1$-matrix and y can be considered as a $m \times 1$-matrix. Then the matrix product Ax of the matrices A and x is defined and (121) states – with regard to (117) – that

$$y = Ax.$$

A special case of Theorem 3.31 is the multiplication of a $1 \times m$-matrix (also called *row vector*) with a $m \times 1$-matrix (that is column vector). In this case we have then

$$(a_1, \ldots, a_m) \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \sum_{i=1}^{m} a_i x_i = a_1 x_1 + \ldots + a_m x_m. \tag{122}$$

Thus we can also describe the content of Theorem 3.31 as follows: the coefficient c_{ij} of the matrix $C = AB$ is the product of the i-th row of the matrix A with the j-th column of the matrix B. Here the product of the i-th row and j-th column of A and B is calculated by the formula (122).

Example.

$$\begin{pmatrix} 2 & 6 & 3 \\ 1 & 3 & 5 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \cdot 2 + 6 \cdot 1 + 3 \cdot 3 \\ 1 \cdot 2 + 3 \cdot 1 + 5 \cdot 3 \end{pmatrix} = \begin{pmatrix} 19 \\ 20 \end{pmatrix}$$

Consider the linear map $f\colon V \to W$ of the n-dimensional F-vector space V to the m-dimensional F-vector space W. Let $B = (b_1, \ldots, b_n)$ be a basis of the vector space V and $C = (c_1, \ldots, c_m)$ be a basis of the vector space W. Let

$$x = \sum_{i=1}^{n} x_i b_i$$

be an arbitrary vector of V. Then x has with respect to the basis $b_1, \ldots, b_n$ the coordinate vector

$$\tilde{x} := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in F^n.$$

that is $\tilde{x} = c_B(x) = i_B^{-1}(x)$. Denote then by y the image of x under the linear map f, that is

$$y := fx.$$

Denote by

$$\tilde{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in F^m$$

the coordinate vector of y with respect to the basis $c_1, \ldots, c_m$ of W, that is $\tilde{y} = c_B(y) = i_B^{-1}(y)$. Let $A = (a_{ij})$ the coordinate matrix of f with respect to the bases B and C. Then it follows from the commutative diagram (115) and the above considerations that

$$\tilde{y} = A\tilde{x}.$$

That is for the coordinates hold the equation (121), namely

$$y_i = \sum_{i=1}^{n} a_{ij}x_k, \qquad i = 1, \ldots, m.$$

Apparently we have the following result about the composite map of two linear maps.

Proposition 3.32. *The product – that is the composite map – of linear maps between finite dimensional vector spaces corresponds to the product of their coordinate matrices.*

More precisely: Let V, V' and V'' be finite dimensional vector spaces over the same field F, say

$$\dim V = n, \qquad \dim V' = m \qquad \text{and} \qquad \dim V'' = l.$$

Furthermore let B be a basis of V, C a basis of V' and D a basis of V''. If $f\colon V \to V'$ and $g : V' \to V''$ are linear maps, then for the coordinate matrix A'' of the composite map $g \circ f\colon V \to V''$ with respect to the basis B of V and D of V'' holds the equality

$$A'' = A'A$$

where A is the coordinate matrix of f with respect to the bases B and C and A' is the coordinate matrix of g with respect to the bases C and D. □

We can visualize the content of the above proposition in a nice way with a commutative diagram, namely

$$\begin{array}{ccccc} V & \xrightarrow{f} & V' & \xrightarrow{g} & V'' \\ \uparrow{\scriptstyle i_B} & & \uparrow{\scriptstyle i_C} & & \uparrow{\scriptstyle i_D} \\ F^n & \xrightarrow[A]{} & F^m & \xrightarrow[A']{} & F^l \end{array} \tag{123}$$

Here the vertical arrows denote in turn the basis homomorphisms i_B, i_C and i_D. To be very precise, the previous proposition states that actually the diagram

$$\begin{array}{ccc} V & \xrightarrow{gf} & V' \\ \uparrow{\scriptstyle i_B} & & \uparrow{\scriptstyle i_D} \\ F^n & \xrightarrow[A''=A'A]{} & F^l \end{array} \tag{124}$$

commutes, that is $(gf) \circ i_B = i_D \circ (A'A)$. But the commutativity of this diagram follows apparently from the commutativity of the diagram (123). Formaly we can convince oursel with the following calculation:

$$\begin{aligned}(gf) \circ i_B = (g \circ f) \circ i_B = g \circ (f \circ i_B) = g \circ (i_C \circ A) = (g \circ i_C) \circ A \\ = (i_D \circ A') \circ A = i_D \circ (A' \circ A) = i_D \circ (A'A).\end{aligned}$$

Using the notation introduced in (111) we can express the content of the previous proposition by

$$c_D^B(gf) = c_D^C(g) \cdot c_C^B(f). \tag{125}$$

One can use the follwing help to memorize this:

$$\frac{B}{D} = \frac{\mathscr{C}}{D} \cdot \frac{B}{\mathscr{C}}$$

Now from (114) follows that the calculation rules for linear maps transfer in a natural way to matrices. In particular we have the following rules

$$A(B_1 + B_2) = AB_1 + AB_2 \tag{126}$$

$$(A_1 + A_2)B = A_1B + A_2B \tag{127}$$

$$a(AB) = (aA)B = A(aB) \qquad (a \in F) \tag{128}$$

$$A(BC) = (AB)C \tag{129}$$

in case the matrix product is defined, that is matrices have the right dimensions.

Of course one can verify those rules for matrix multiplicaton just by using the formula for matrix multiplication as given in Theorem 3.31, The rules (126), (127) and (128) are then evident, and even the verification of (129) is not difficult but rather a bit cumbersom and – as said – unnecessary.

Note that the role of the identity map id_V of an n-dimensional F-vector space V is taken in the case of the matrices by the $n \times n$-identity matrix

$$I_n := \begin{pmatrix} 1 & & 0 \\ & 1 & \\ & \ddots & \\ 0 & & 1 \end{pmatrix}, \qquad I_n = \mathrm{id}_{F^n}, \tag{130}$$

whose colums are in turn the elements $e_1, \ldots, e_n$ of the canonical basis of F^n. For every $A \in F^{m,n}$ holds

$$I_m A = A = A I_n. \tag{131}$$

Note that it is often customary to leave away the index n in the notation for the identity matrx I_n, of course only if there is no danger of confusion. Thus (131) can also be written as

$$IA = A = AI. \tag{132}$$

7. The Matrix Description of $\mathrm{End}_F(V)$

Let us turn our attention once again to the case of the linear "self mappings" of an n-dimensional F-vector space V, that is endomorhpsism

$$f\colon V \to V$$

of V and their matrix description.

It follows from Proposition 3.23 that the $n \times n$-matrices

$$F^{n,n} = \mathrm{Hom}_F(F^n, F^n) = \mathrm{End}_F(F^n)$$

form a F-algebra. This F-algebra has numerous applications and this motivates the next definition.

Definition 3.33. Let F be a field and $n \geq 1$ a number. Then F-algebra $F^{n,n}$ is called the *algebra of the $n \times n$-matrices over F* and it is denoted in symbols by

$$M_n(F) := F^{n,n}.$$

Note that the algebra of the $n \times n$-matrices over F is sometimes also called the *(full) matrix ring of degree n over F*. Apparently the identity element of the multiplication in $M_n(F)$ is the $n \times n$-identity matrix I_n. The elements of $M_n(F)$ are also called *square matrices* (of degree n). As a F-vector space $M_n(F)$ has dimension n^2.

We use the notation of Theorem 3.31 but this time we set $W = V$ and $B = C$. In particular this means that we will use *one and the same* basis $B = (b_1, \ldots, b_n)$ for the range and the image for the matrix description of any enodmorphism $f\colon V \to V$. The *coordinate matrix $A = (a_{rs})$ of the endomorphism f with respect to the basis B* is then defined by

$$f(b_i) = \sum_{j=1}^{n} a_{ji} b_j \qquad (i = 1, \ldots, n). \tag{133}$$

Compare this with (109)! As a linear map the matrix A is the endomorphism A of F^n which makes the diagram

$$\begin{array}{ccc} V & \xrightarrow{\ f\ } & V \\ {\scriptstyle c_B}\downarrow & & \downarrow{\scriptstyle c_B} \\ F^n & \xrightarrow[\ A\ ]{} & F^n \end{array}$$

where the vertical arrows are the coordinate isomorphisms $c_B\colon V \to F^n$. Using the notation introduced in (111) we have

$$A = c_B^B(f).$$

But we shall also use the notation

$$A = c_B(f)$$

to denote the coordinate matrix A of the endomorphism $f\colon V \to V$ with respect to the basis B of V. If now f and g are endomorphisms of V which have the coordinate matrices A and A' respectively with respect to the basis B, then the coordinate matrix of the endomorphism $fg = f \circ g$ with respect to the same basis B is given by the product AA'. Using the above introduced notation we get therefore

$$c_B(fg) = c_B(f)c_B(g).$$

This is a straight consequence of Proposition 3.32, but compare the result also with (125)! Furthermore hold

$$c_B(\mathrm{id}_V) = I_n.$$

In order to summarize the matrix description of endomorphisms in a suitable compact way let us define first what we mean by an ismorphism of F-algebras.

Definition 3.34. Let R and S be algebras (with unit) over a field F. A linear map $\varphi\colon R \to S$ is said to be a *homomorphism of F-algebras* if it satisfies the following two conditions:

(1) $\varphi(fg) = \varphi(f)\varphi(g)$ for all $f, g \in R$.
(2) $\varphi(1) = 1$.

If φ is bijective, then φ is called an *isomorphism of F-algebras* (and then also the inverse map $\varphi^{-1}\colon S \to R$ is an isomorphism of F-algebras). Two F-algebras R and S are said to be *isomorphic (as F-algebras)* if there exists an isomorphism $\varphi\colon R \to S$ of F-algebras and in symbols this fact is denoted by $R \cong S$.

Theorem 3.35. *Let V be a n-dimensional vector space over the field F and let $B = (b_1, \ldots, b_n)$ be a basis of V. Then the map*

$$c_B\colon \operatorname{End}_F(V) \to M_n(F),$$

which assigns each endomorphism f of V its coordinate matrix with respect to the basis B of V, is an isomorphism of F-algebras. In particular

$$\operatorname{End}_F(V) \cong M_n(F) \qquad \textit{(as } F\textit{-algebras).}$$

Note that we did know already that $\operatorname{End}_F(V) \cong M_n(F)$ as vector spaces but now we also know that they are also isomorphic as F-algebras. This is a stronger result since a vector space isomorphism between F-algebras is not necessarily a isomorphism of F-algebras!

8. Isomorphisms (Again)

Proposition 3.36. *Let V and W be arbitrary vector spaces over the same field F. Then a linear map $f\colon V \to W$ is an isomorphism if and only if there exists a linear map $g\colon W \to V$ such that*

$$g \circ f = \mathrm{id}_V \qquad \textit{and} \qquad f \circ g = \mathrm{id}_W\,. \tag{134}$$

Proof. "⇒": If $f\colon V \to W$ is an isomorphism, then f is by definition a bijective map. Thus there exists the inverse map $f^{-1}\colon W \to V$ and this map is an isomorphism by Proposition 3.6. If we set $g := f^{-1}$, then apparently (134) is satisfied.

"⇐": If g is a linear map which satisfies (134) then f is apparently bijective and thus an isomorphism. □

For finite dimensional vector spaces we have the following result.

Proposition 3.37. *Let V be a n-dimensional and W an m-dimensional vector space over the same field F. Then for a linear map $f\colon V \to W$ the following statements are equivalent:*

(1) *f is an isomorphism.*
(2) *We have $m = n$ and* $\operatorname{rank} f = n$.
(3) *If A is the coordinate matrix of f with respect to some bases of V and W, then the linear map $A\colon F^n \to F^m$ is a isomorphism.*

Proof. "(1) ⇒ (2)": If f is an isomorphism, then $W = \operatorname{im} f$ and $\dim W = m = n$ and $\operatorname{rank} f = \dim(\operatorname{im} f) = \dim W = n$.

"(2) ⇒ (1)": We have $\operatorname{rank} f = \dim V$ and $\operatorname{rank} f = \dim W$. Then from the notes made after the Definition 3.14 it follows from the first equality that f is a monomorphism and from the latter equality that f is an epimorphism. Thus alltogether we have shown that f is an isomorphism.

"(1) $\Rightarrow$ (3)": From the commutativity of the diagram (115) we have that $A = c_C \circ f \circ i_B$ where i_B is the basis isomorphism with respect to the basis B of V and $c_C = (i_C)^{-1}$ is the coordinate isomorphism of W with respect to the basis C of W. Thus A is a isomorphism since it is the composite of three isomorphism (Proposition 3.6).

"(3) $\Rightarrow$ (1)": Similar as above, but now we can write $f = i_C \circ A \circ c_B$ as the composite of three isomorphisms and therefore f is an isomorphism by Proposition 3.6. □

Now Proposition 3.37 states that a linear map $A\colon F^n \to F^m$ is an isomorphism if and only if $n = m$ and the matrixs A has rank n. In this case there exists the inverse map which we denote by A^{-1}. We are not yet able to express the coefficents of A^{-1} in terms of the coefficents of the matrix A as we need for this a new theoretical approach.[2] But we will soon describe an algorithm based on the Gauss algorithm to calculate the coefficients for the matrix A^{-1} from the matrix A.

Definition 3.38. We say that a matrix $A \in F^{n,n}$ is *invertible* if the linear map $A\colon F^n \to F^n$ is an isomorphism. The *inverse matrix* A^{-1} of an invertible matrix is the inverse linear map $A^{-1}\colon F^n \to F^n$.

Note that apparently the inverse matrix of an invertible matrix is again invertible. Furthermore note that we have from the commutativity of diagram (115) that if $f\colon V \to W$ is an isomorphism, then

$$c_B^C(f^{-1}) = (c_B^C(f))^{-1}.$$

9. Change of Bases

Now we have already many times repeated that the coordinate matrix of a linear map $f\colon V \to W$ depends very much on the chosen bases B and C of the vector spaces V and W respectively. We shall now study this dependency more closely.

Definition 3.39 (Transition Matrix)**.** Let V be an n-dimensional vector space over the field F. Let $B = (b_1, \ldots, b_n)$ and $B' = (b'_1, \ldots, b'_n)$ be two bases of V. Then there exists unique elements $s_{ji} \in F$ such that

$$b'_i = \sum_{j=1}^{n} s_{ji} b_j, \qquad i = 1, \ldots, n. \tag{135}$$

Then the $n \times n$-matrix

$$S := \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1n} \\ s_{21} & s_{22} & \ldots & s_{2n} \\ \vdots & & & \vdots \\ s_{n1} & s_{n2} & \ldots & s_{nn} \end{pmatrix}$$

is called the *transition matrix from B to B'*.

[2] With the help of determinants we will be able to formulate the Cramer rule for matrix inversion.

From Definition 3.24 and (109) we have that the diagram

$$\begin{array}{ccc} V & \xrightarrow{\mathrm{id}_V} & V \\ {\scriptstyle i_{B'}}\uparrow & & \uparrow{\scriptstyle i_B} \\ F^n & \xrightarrow{S} & F^n \end{array} \tag{136}$$

is commutative, where i_B and $i_{B'}$ denote the coordinate isomorphism corresponding to the bases B and B' of the vector space V. Thus S is just the coordinate matrix of the identity map id_V with respect to the bases B' and B (note the order in which the bases are mentioned!). That is

$$S = c_B^{B'}(\mathrm{id}_V). \tag{137}$$

In particular this means by Proposition 3.37 that S is invertible, that is $S\colon F^n \to F^n$ is an isomorphism. Now from the commutativity of the diagram (136) one concludes – since $c_B = i_B^{-1}$ and $c_{B'} = i_{B'}^{-1}$ – that the diagram

$$\begin{array}{ccc} & V & \\ {\scriptstyle c_{B'}}\swarrow & & \searrow{\scriptstyle c_B} \\ F^n & \xrightarrow{S} & F^n \end{array}$$

commutes. Thus for every vector

$$v = \sum_{i=1}^{n} x_i b_i = \sum_{i=1}^{n} x'_i b'_i$$

in V holds $c_B(v) = Sc_{B'}(v)$, that is

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = S \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

So we have for the coordinates of v with respect to bases B and B' the transformation equation

$$x_i = \sum_{j=1}^{n} s_{ij} x'_j, \qquad i = 1, \ldots, n.^3$$

Note further that from the commutativity of the diagram (136) it follows that if S is the transformation matrix from B to B', then S^{-1} is the transformation matrix from B' to B.

Consider still the notations and conditions of Definition 3.39. There exists precisely one linear map

$$f\colon V \to V$$

such that $f(b_i) = b'_i$ for $1 \leq i \leq n$ (Proposition 3.5). It follows from (135) that the transition matrix S of B to B' is at the same time the coordinate matrix of the endomorphism f of V with respect to the basis B, that is

$$S = c_B(f) = c_B^B(f).$$

[3] Note the coefficients s_{ij} in this sum are indeed correct! Compare this with (135) where the coefficients s_{ji} are used!

The isomorphism $g := f^{-1}\colon V \to V$ which now maps the basis B' to the basis B has then – as an endomorphism of V – the coordinate matrix S^{-1} with respect to B, that is

$$S^{-1} = c_B(g) = c_B^B(g).$$

Furthermore we obtain straight from the definition the equalities $c_{B'}^B(f) = I = c_B^{B'}(g)$ and if one applies f to (135) one obtains $S = c_{B'}^{B'}(f)$.

Moreover, the transition matrix from the canonical standart basis $e_1, \ldots, e_n$ of F^n to an arbitrary basis $s_1, \ldots, s_n$ of F^n is apparently the matrix

$$S = (s_1, \ldots, s_n)$$

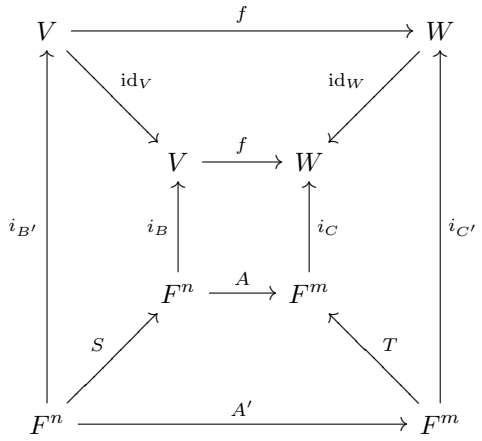
which columns are precisely the vectors $s_1, \ldots, s_n$.

Theorem 3.40 (Transformation of the Coordinate Matrix under Change of Bases)**.** *Let V and W be two finite dimensional vector spaces over the same field F, say $\dim V = n$ and $\dim W = m$. Let B and B' bases of the vector space V and let C and C' be bases of the vector space W. Assume that $f\colon V \to W$ is a linear map which has with respect to the bases B and C the coordinate matrix A, then the coordinate matrix A' of f with respect to the bases B' and C' is given by*

$$A' = T^{-1}AS \tag{138}$$

where S is the transition matrix from B to B' and T is the transition matrix from C to C'.

Proof. Consider the diagram



(139)

We claim that this diagram is commutative. The small inner square of this diagram commutes by the definition of the matrix A. And likewise the big outer square of this diagram commutes by the definition of the matrix A'. The the upper trapezoid is trivially commutative. The left and right trapezoids are the commutative diagrams (136). Now the verification of the commutativity of remaining lower trapezoid is simple:

$$\begin{aligned}
A' &= (i_{C'})^{-1} \circ f \circ i_{B'} && \text{(outer square)}\\
&= T^{-1} \circ i_C^{-1} \circ \mathrm{id}_W \circ f \circ i_{B'} && \text{(right trapezoid)}\\
&= T^{-1} \circ i_C^{-1} \circ f \circ \mathrm{id}_V \circ i_{B'} && \text{(upper trapezoid)}\\
&= T^{-1} \circ i_C^{-1} \circ f \circ i_B \circ S && \text{(left trapezoid)}\\
&= T^{-1} \circ A \circ S && \text{(inner square)}
\end{aligned}$$

□

If one uses the notation (111) we can – using (137) – write the content of Theorem 3.40 in the intuitive form

$$c_{C'}^{B'}(f) = c_{C'}^{C}(\mathrm{id})\, c_{C}^{B}(f)\, c_{B}^{B'}(\mathrm{id}) \tag{140}$$

since $S = c_B^{B'}(\mathrm{id})$ and $T = c_C^{C'}(\mathrm{id})$.

Applying Theorem 3.40 to matrices seen as linear maps we get the next result.

Proposition 3.41. *Let A be a $m \times n$-matrix with coefficients in the field F. Then the coordinate matrix A' of the linear map $A\colon F^n \to F^m$ with respect to arbitrary bases $(s_1, \ldots, s_n)$ and $(t_1, \ldots, t_m)$ of the vector spaces F^n and F^m respectively is given by*

$$A' = T^{-1}AS \tag{141}$$

where S is the $n \times n$-matrix with columns $s_1, \ldots, s_n$ and T is the $m \times m$-matrix with columns $t_1, \ldots, t_m$.

Proof. With respect to the canonical bases of F^n and F^m the coordinate matrix of the linear map $A\colon F^n \to F^m$ is precisely the matrix A. Now the claim of the proposition follows from Theorem 3.40 and the note on page 75 regarding the interpretation of the transition matrix from the standard basis to an arbitrary basis. □

We shall re-state Theorem 3.40 in the special case of an endomorphism and that we consider only coordinate matrices with respect to the same bases for domain and co-domain of the endomorphism.

Theorem 3.42 (Change of Coordinates for Endomorphisms)**.** *Let V be a vector space over the field F with $\dim V = n$. Let B and B' be bases of V. If f has with respect to the basis B the coordinate matrix A, then the coordinate matrix A' of f with respect to the basis B' is given by*

$$A' = S^{-1}AS \tag{142}$$

where S is the transition matrix of B to B'. In particular we have for quadratic matrices the following result.

If A is a $n \times n$-matrix over F, then the coordinate matrix A' of the endomorphism $A\colon F^n \to F^n$ with respect to an arbitrary basis $s_1, \ldots, s_n$ of F^n is

$$A' = S^{-1}AS$$

where S is the $n \times n$-matrix with the columns $s_1, \ldots, s_n$. □

10. Equivalence and Similarity of Matrices

The transformation properties of matrices under change of bases which has been described in Theorem 3.40 and Theorem 3.42 introduce a new, ordering aspect for matrices. For example, Theorem 3.40 states if A and A' are $m \times n$-matrices over the field F such that there exists invertible matrices $S \in F^{n,n}$ and $T \in F^{m,m}$ such that

$$A' = T^{-1}AS \tag{143}$$

then we may see the matrices A and A' to be in a certain way "eqvivalent" to each other. Due to Theorem 3.40 we can regard A and A' to belong to *one and the same* linear map $f\colon V \to W$ (with respect to different bases). Similar is true for quatratic matrices A and A' in $F^{n,n}$. If we have for some invertible $S \in F^{n,n}$ the equality

$$A' = S^{-1}AS \tag{144}$$

then A and A' can be seen as coordinate matrices of *one and the same* endomorphism $f\colon V \to V$ (with respect to different bases of V). Now these considerations motivates the following definition.

Definition 3.43 (Equivalence and Similarity of Matrices)**.** Let F be a field.

(1) If $A, A' \in F^{n,m}$ then we say that A is *equivalent* with A' in $F^{n,m}$, and we denote this fact in symbols by
$$A \sim A',$$
if there exists invertible matrices $S \in F^{n,n}$ and $T \in F^{m,m}$ such that the equality (143) is satisfied.

(2) If $A, A' \in F^{n,n}$ then we say that A is *similar* to A' in $F^{n,n}$, and we denote this fact in symbols by
$$A \approx A',$$
if there exists an invertible matrix $S \in F^{n,n}$ such that the equality (144) is satisfied.

We shall explicitly put the following result on record again.

Proposition 3.44. (1) *Two matrices $A, A' \in F^{m,n}$ are equivalent if and only if there exists a linear map $f\colon V \to W$ between an n-dimensional F-vector space V and an m-dimensional F-vector space W such that both A and A' appear as coordinate matrices of f with respect to suitable bases for V and W.*

(2) *Two matrices $A, A' \in F^{n,n}$ are equivalent if and only if there exists an endomorphism $f\colon V \to V$ of an n-dimensional F-vector space V such that both A and A' appear as coordinate matrices of f with respect to suitable bases for V.*

PROOF. (1) "$\Leftarrow$": If A and A' are coordinate matrices of one and the same linear map f then $A \sim A'$ according to Theorem 3.40.

"$\Rightarrow$": We assume that $A \sim A'$, that is there exists invertible matrices $S \in F^{n,n}$ and $T \in F^{m,m}$ such that $A' = T^{-1}AS$. Let $(s_1, \ldots, s_n)$ be the system of the columns of the matrix S and let $(t_1, \ldots, t_m)$ be the system of the columns of the matrix T. Then according to Proposition 3.41 the linear map $A\colon F^n \to F^m$ has the coordinate matrix A' with respect to the bases $(s_1, \ldots, s_n)$ and $(t_1, \ldots, t_m)$ of F^n and F^m respectively. But the same linear map has the coordinate matrix A with respect to the standard bases of F^n and F^m. Thus there exists a linear map such that both A and A' appear as the coordinate matrix this linear map and this concludes the proof of the first part of the proposition.

(2) The second part of the proposition is proven the same way as the first part. □

Note that the relation "$\sim$" on $F^{m,n}$ which we have introduced in Definition 3.43 is an equivalence relation (see page 54). That is, for every $A, A', A'' \in F^{n,m}$ hold the following three statements.

(1) $A \sim A$ ("reflexivity")
(2) $A \sim A' \Rightarrow A' \sim A$ ("symmetry")
(3) $A \sim A'$ and $A' \sim A'' \Rightarrow A \sim A''$ ("transitivity")

Similarly the similarity relation "$\approx$" on $F^{n,n}$ is an equivalence relation.

Now the natural question is how can one decide easily whether two $m \times n$-matrices A and B over a field F are equivalent. The next result will give a exhaustive answer to this question.

Proposition 3.45. *Let F be a field. Then the following two statements are true.*

(1) *Every matrix $A \in F^{m,n}$ is equivalent to precisely one matrix of the form*

$$\left(\begin{array}{cccccc|c} 1 & 0 & 0 & \cdots & \cdots & 0 & \\ 0 & 1 & 0 & & & \vdots & \\ 0 & 0 & 1 & & & \vdots & 0 \\ \vdots & & & \ddots & & \vdots & \\ \vdots & & & & \ddots & 0 & \\ 0 & 0 & 0 & \cdots & 0 & 1 & \\ \hline & & & 0 & & & 0 \\ & & & & & & \end{array}\right) \in F^{m,n} \tag{145}$$

where the upper left part of this matrix is the $r \times r$-identity matrix and r is a certain natural number $0 \leq r \leq m, n$. In this case $\operatorname{rank} A = r$.

(2) *The matrices A and B of $F^{m,n}$ are equivalent if and only if* $\operatorname{rank} A = \operatorname{rank} B$.

PROOF. (1) Consider the linear map $A\colon F^n \to F^m$. Set $r := \dim(\operatorname{im} A) = \operatorname{rank} A$. Then $\dim(\ker A) = n - r$. Let $b_{r+1}, \ldots, b_r$ be a basis of $\ker A$. By the Basis Extension Theorem 2.36 we can extend this to a basis $b_1, \ldots, b_r$ of F^n. Set $c_i := Ab_i$ for $i = 1, \ldots, r$. Then $c_1, \ldots, c_r$ is a linear independent subset of F^m and again using Theorem 2.36 we can extend this set to a basis $c_1, \ldots, c_m$ of F^m. Then we have

$$A(b_i) = c_i \quad (1 \leq i \leq r) \qquad \text{and} \qquad A(b_i) = 0 \quad (r+1 \leq i \leq n).$$

Thus the coordinate matrix of the linar map A with respect to the above constructed bases for F^n and F^m is precisely of the form (145). It follows that

$$A \sim \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{146}$$

with $r = \operatorname{rank} A$. The uniqueness will follow from the next part.

(2) "$\Rightarrow$": If $A \sim B$ then by Proposition 3.44 there exists a linear map $f\colon V \to W$ such that both matrices A and B appeare as coordinate matrices of f with respect to suitable bases for V and W. Then $\operatorname{rank} A = \operatorname{rank} f = \operatorname{rank} B$ by Proposition 3.25. In particular

$$A \not\sim \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

if $r \neq \operatorname{rank} A$ and this completes the proof of the first part of the proposition.

"$\Leftarrow$": If $r := \operatorname{rank} A = \operatorname{rank} B$, then by the first part it follows that both A and B are equivalent to a matrix of the form (146). But then by the transitivity of the relation "$\sim$" it follows that $A \sim B$. □

Proposition 3.45 classifies all $m \times n$-matrices upto equivalence of matrices: two matrices of $F^{m,n}$ are equivalent if and only if they have the same rank. More over the proposition states, that in every class of matrices which are equivalent to each other exists precisely one matrix of the form (145). Therefore every equivalence class of such matrices can be labeled with a "*representative*" of such a class, which

is outstanding before all other members of this class, namely

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix},$$

a representative which has a very simple form. This simple matrix which represents the class of all $m \times n$-matrices with rank r over the field F is called the *normal form* of this class.

The problem we have discussed above is an example of a classification problem which are encountered frequently in mathematics. Of course classification problems do in general have such a simple solution even if there exists a complete solution to a given problem.

We have classified all $m \times n$-matrices over a field F upto equivalence. A natural question is whether we can classify all quadratic $n \times n$-matrices up to similarity. How would a normal form for such a class look like? But this problem is much harder to solve and in this lecture we will not develope the tools which are needed to give an answer to this problem.

11. The General Linear Group

We begin this section with a small excursion to the world of Algebra. Recall Definition 2.3 where we defined a ring (with unit). If R is a ring then we say that an element $a \in R$ is *invertible* (in R) if there exists a $b \in R$ such that

$$ab = 1 \qquad \textit{and} \qquad ba = 1.$$

We denote the set of all invertible elements of a ring R by $R^\times$, that is we define

$$R^\times := \{a \in R : a \text{ is invertible}\}.$$

The elements of $R^\times$ are called the *units* of the ring R. If a is a unit and $b \in R$ such that $ab = 1$ and $ba = 1$, then we call b the *(multiplicative) inverse* of a. Note that if $a \in R$ is invertible then its multiplicative inverse is uniquely defined. It is customary to denote this unique multiplicative inverse element of a by "a^{-1}".

Examples.
(1) Consider the ring of integers $\mathbb{Z}$. The only two invertible elements in $\mathbb{Z}$ are 1 and -1. Thus $\mathbb{Z}^\times = \{1, -1\}$.
(2) Let F be a field. Then $F^\times = \{x \in F : x \neq 0\}$.
(3) Let V be a F-vector space. Then $\mathrm{End}_F(F)$ is a ring and $\mathrm{End}_F(V)^\times$ is precisely the set of all isomorphisms $f\colon V \to V$ (see Proposition 3.47).

We make the following algebraic observations about $R^\times$: If a and b are units of R, then also their product ab is a unit of R, that is the set $R^\times$ is closed under the multiplication of R. This is due to $ab(b^{-1}a^{-1}) = aa^{-1} = 1$ and likewise $ba(a^{-1}b^{-1}) = bb^{-1} = 1$. In particular $(ab)^{-1} = b^{-1}a^{-1}$ (note the change in the order of a and b!). Furthermore $1 \in R^\times$ and apparently $a^{-1} \in R^\times$ for every $a \in R^\times$. Finally, since the multiplication on R is associative it follows that the multiplication restricted to the units of R is also associative.

A set G together with a law of composition

$$G \times G \mapsto G, (x, y) \mapsto xy$$

which staisfies precisely the above conditions is called in mathematics a *group*. More preciesly we have the following definition.

Definition 3.46 (Group)**.** A *group* is a tuple $G = (G, *)$ consisting of a set G together with a map

$$*\colon G \times G \to G, (x, y) \mapsto x * y$$

(called the *law of composition*) if the follwoing three group axioms are satisfied:

(G1) $(x * y) * z = x * (y * z)$ for every $x, y, z \in G$.
(G2) There exists an element $e \in G$ (called the *identity element* of G) such that $x * e = x$ and $e * x = x$ for all $x \in G$.
(G3) For every $x \in G$ there exists an element $y \in G$ (called the *inverse element* of x) such that $x * y = e$ and $y * x = e$.

If $x * y = y * x$ for every $x, y \in G$, then the group G is called *abelian*[4].

Examples.
(1) The set of integers $\mathbb{Z}$ is an abelian group under the usual addition.
(2) Let F be a field. Then F is an abelian group under the addition. Furthermore $F^\times$ is an abelian group under the multiplication.
(3) Let V be a vector space. Then V is an abelian group under the addition of vectors.
(4) Let R be a ring. Then the set $R^\times$ of all units of R forms a group under the multiplication of R.

After this short excurse into the world of Algebra we return to Linear Algebra. Given an arbitrary vector space V over a field F we consider the endomorhismn ring $\mathrm{End}_F(V)$ of V. We make the following observation.

Proposition 3.47. *Let $f\colon V \to V$ be an endomorphism of the vector space V. Then the following two statements are equivalent.*

(1) *f is an isomorphism.*
(2) *f is a unit of the endomorphism ring $\mathrm{End}_F(V)$ of V.*

Proof. By Proposition 3.6 is f precisely then an isomorhphism if there exists a linear map $g\colon V \to V$ such that $g \circ f = \mathrm{id}$ and $f \circ g = \mathrm{id}$. But this is equivalent with f being invertible in $\mathrm{End}_F(V)$. □

Definition 3.48 (General Linear Group)**.** The group of all invertible elements of the endomorphism ring $\mathrm{End}_F(V)$ of the F-vector space V is denoted by

$$\mathrm{GL}_F(V) := \mathrm{End}_F(V)^\times$$

and is called the *General Linear Group* of the F-vector space V. In the special case of $V = F^n$ we set

$$\mathrm{GL}_n(F) := M_n(F)^\times = \mathrm{End}_F(F^n)^\times$$

and this group is called the *General Linear Group (of degree n)* over the field F.

An element $f \in \mathrm{GL}_F(V)$, that is an isomorphism $f\colon V \to V$, is also called an *automorphism*. Therefore the group $\mathrm{GL}_F(V)$ is sometimes also called the *automorphism group* of V.

Proposition 3.49. *Let V be an n-dimensional vector space over the field F. then*

$$\mathrm{GL}_F(V) \cong \mathrm{GL}_n(F) \qquad \textit{(isomorphism of groups).}$$

[4]Named after the Norwegian mathematician Niels Henrik Abel, 1802–1829

PROOF. Note that we say that two groups G and G' are isomorphic (as groups) if there exists a group isomorphism $f\colon G \to G'$, that is a bijective map such that $f(xy) = f(x)f(y)$ for all $x, y \in G$.

Now let B be any basis of V. Then we know by Theorem 3.35 that the coordinate isomorphism

$$c_B\colon \operatorname{End}_F(V) \to M_n(F)$$

is an isomorphism of F-algebras. In particular c_B maps isomorphism of V to invertible matrices of F. Apparently c_B induces then an isomorphism of groups $\mathrm{GL}_F(V) \to \mathrm{GL}_n(F)$. □

Because of the above result it is now possible to restrict our attention without any loss of generality to the groups $\mathrm{GL}_n(F)$ when studying the general linear groups of finite dimensional F-vector spaces.

Again we tie up with Chapter 1: Let A be a $m \times n$-matrix over a field F with columns $v_1, \ldots, v_n$, that is the v_i $(1 \leq i \leq n)$ are all vectors of F^m. We consider an elementary column transformation of type I. Let us denote by $U_{ij}(a)$ precisely the elementary transformation

$$(\ldots, v_i, \ldots, v_j, \ldots) \mapsto (\ldots, v_i, \ldots, v_j + av_i, \ldots)$$

of the system of vectors $(v_1, \ldots, v_n)$ which replaces the vector v_j in $(v_1, \ldots, v_n)$ by the vector $v_j + av_i$ $(i \neq j$ and $a \in F)$. Consider on the other hand the linear map

$$T\colon F^n \to F^n$$

defined by the equations

$$Te_j = e_j + ae_i, \qquad Te_k = e_k \quad (\text{for } k \neq j), \tag{147}$$

where $e_1, \ldots, e_n$ denotes as usuall the standard basis of F^n. Then we have for the composite map (that is the matrix product) AT by definition $ATe_j = A(e_j + ae_i) = Ae_j + aAe_i = v_j + av_i$ and for $k \neq j$ we get $ATe_k = Ae_k = v_k$. That is we have

$$(AT)e_j = v_j + av_i, \qquad (AT)e_k = v_k \quad (\text{for } k \neq j). \tag{148}$$

Now recall the following (compare this with Proposition 3.28): the columns of a $m \times n$-matrix C over F are in turn the images of $e_1, \ldots, e_n$ under the linear map $C\colon F^n \to F^m$. From (148) it follows that the elementary column transformation $U_{ij}(a)$ is obtained by multiplying the matrix A from *right* with the $n \times n$-matrix

$$T_{ij}^{(n)}(a) := \begin{pmatrix} 1 & & & & \vdots & & \\ & 1 & & & \vdots & 0 & \\ \cdots & \cdots & \ddots & \cdots & a & \cdots & \cdots \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & 0 & & & \vdots & \ddots & \\ & & & & \vdots & & 1 \end{pmatrix} \tag{149}$$

which has on the main diagonal only ones "1" and otherwise only zeros except the entry in the i-th row and j-th column where we have the element a. That is, if the coefficients of the matrix $T_{ij}^{(n)}(a)$ are denoted by t_{kl} then we have

$$t_{kl} = \begin{cases} 1 & \text{if } k = l, \\ a & \text{if } k = i \text{ and } l = j, \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 3.50 (Interpretation of Elementary Row and Column Transformations of Type I with the Help of Matrix Multiplications)**.** *Let A be an $m \times n$-matrix over a field F. Denote by $(v_1, \ldots, v_n)$ the system of its columns and denote by $(u_1, \ldots, u_m)$ the system of its rows. An elementary column transformation $U_{ij}(a)$ of A – that is replacing of the j-th column v_j of A by $v_j + av_i$ – is obtained by multiplying the matrix A with the special matrix $T_{ij}^{(n)}(a)$ from the* right*. Likewise the effect of multiplying the matrix A with $T_{ij}^{(m)}(a)$ from the* left *is the elementary row transformation $U_{ji}(a)$, that is replacing the i-th row of A by $u_i + au_j$.*

Proof. We have already shown that the proposition is true for column transformations. The second part of the proposition one verifies by calculating in a similar way the matrix $T_{ij}^{(m)}(a)A$. □

Definition 3.51 (Elementary Matrices)**.** The set of matrices $T_{ij}(a)$ of the form (149) for $i \neq j$ and $a \in F$ are called $n \times n$*-elementary matrices* over F.

One obtains directly from the definion (147) the following result about elementary matrices.

Proposition 3.52. *The follwoing two calculation rules for the elementary $n \times n$-matrices over a field F are satisfied:*

$$T_{ij}(a+b) = T_{ij}(a)T_{ij}(b) \qquad (a, b \in F) \tag{150}$$

$$T_{ij}(0) = I \tag{151}$$

In particular every elementary $n \times n$-matrices is invertible, that is $T_{ij}(a) \in \mathrm{GL}_n(F)$ for every $1 \leq i \neq j \leq n$ and $a \in F$. For the inverse of an elementary $n \times n$-matrix we have the simple formula

$$T_{ij}(a)^{-1} = T_{ij}(-a). \tag{152}$$

□

If one applies successively elementary column transformation to a matrix A then by Proposition 3.50 this corrsponds to successively multiplying A from the *right* by certain elementary matrices $T_1, \ldots, T_r$:

$$((AT_1)T_2)\cdots)T_r = A(T_1T_2\cdots T_r). \tag{153}$$

Similarly successively elementary row transformations are corresponds by Proposition 3.50 to successively multiplying A from the *left* by certain elementary matrices $T_1, \ldots, T_r$:

$$T_r(\cdots T_2(T_1A)) = (T_r \cdots T_1)A. \tag{154}$$

In order to describe the repeated application of elementary column and row transformations of type I in a better way we make the following definition.

Definition 3.53 (The Special Linear Group)**.** The set

$$\mathrm{SL}_n(F)$$

of all possible products of elementary $n \times n$-matrices over the field F is called the *special linear group (of degree n)* over F.

Proposition 3.54. *The special linear group $\mathrm{SL}_n(F)$ is a subgroup of the general linear group $\mathrm{GL}_n(F)$.*

PROOF. Note that a subset H of a group G is called a *subgroup* if it is a group under the law of composition of G.[5]

The matrix product of two elements $T_1T_2\cdots T_r$ and $T_1'T_2'\cdots T_{r'}'$ is apparently again an element of $\mathrm{SL}_n(F)$. Thus the matrix product defines a law of composition on $\mathrm{SL}_n(F)$. It is clear that this law of composition inherits the associativity from the associativity of the matrix product. Due to (151) we have $I \in \mathrm{SL}_n(F)$ for the identity matrix. Finally it is follows from 152 and the definition of $\mathrm{SL}_n(F)$ that the inverse element of $T_1T_2\cdots T_r$ is given by

$$(T_1T_2\cdots T_r)^{-1} = T_r^{-1}\cdots T_2^{-1}T_1^{-1}$$

and it is therefore also an element of $\mathrm{SL}_n(F)$. Therefore $\mathrm{SL}_n(F)$ is a group under the matrix product and thus a subgroup of $\mathrm{GL}_n(F)$. □

Let $A \in \mathrm{GL}_n(F)$ be an invertible matrix over the field F. This means that

$$\operatorname{rank} A = n. \tag{155}$$

We try solve the task to transform the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

into a matrix of a as simple form as possible, and this only by using the fact (155) and row transformations of type I. To avoid triviality we may assume that $n \geq 2$.

Since the column rank of A is equal to $\operatorname{rank} A = n$ we have that the first row of A is not empty. Thus we can achive by using a suitable row transformation of type I that the second entry of the first column of A – that is a_{21} – is different from 0. Then by adding $a_{21}^{-1}(1 - a_{11})$-times the second row to the first row we get a matrix A' with

$$a_{11}' = 1.$$

Now we can eliminate the coefficients in the first column below the element a_{11}' by suitable row transformations and get a matrix A'' of the form

$$A'' = \left(\begin{array}{c|ccc} 1 & * & \dots & * \\ \hline 0 & & & \\ \vdots & & C & \\ 0 & & & \end{array}\right) \tag{156}$$

with a $(n-1) \times (n-1)$-matrix C which must have rank $n-1$. If $n \geq 3$, then we can apply the above described algorithm to the matrix C. If one takes into account that the row transformations of A'' which corresponds to the row transformations applied to the matrix C do not affect the zero entries in the first row of (156) we see that we can transform the matrix A into a matrix of the form

$$\begin{pmatrix} 1 & & & & \\ & 1 & & * & \\ & & \ddots & & \\ & & & \ddots & \\ & 0 & & 1 & \\ & & & & d \end{pmatrix}, \qquad (d \neq 0). \tag{157}$$

[5]Compare the definition of a subgroup with the definition of a linear subspace, see Definition 2.5 on page 23.

The main diagonal of this matrix consist only of ones except the last entry, which is an element $d \in F$ from which we know only that it is different from zero. The entries in this matrix below the main diagonal are all zero. Apparently we can – only with the help of row transformations of type I – transform this matrix into the following form

$$D_n(d) := \begin{pmatrix} 1 & & & & \\ & 1 & & 0 & \\ & & \ddots & & \\ & & & \ddots & \\ & 0 & & 1 & \\ & & & & d \end{pmatrix}, \qquad (d \neq 0), \tag{158}$$

which is a diagnoal matrix with only ones on the diagonal except that the last element of the diagonal is an element $d \in F$ from which we only know that it is different from zero. We have that $D_n(d) \in \mathrm{GL}_n(F)$. Thus we have shown:

Theorem 3.55. *If A is an invertible $n \times n$-matrix over a field F, then we can transform this matrix by row transformations of type I into a diagonal matrix of the special form* (158) *for some non-zero $d \in F$.*

In other words, if $A \in \mathrm{GL}_n(F)$, then there exists a $S \in \mathrm{SL}_n(F)$ and a non-zero $d \in F$ such that

$$A = SD_n(d). \tag{159}$$

were $D_n(d)$ is a diagonal matrix of $\mathrm{GL}_n(V)$ of the special form (158).

Proof. We have already shown the first part of the theorem. Now from the first part and Proposition 3.50 follows that there exist elementary matrices $T_1, \ldots, T_r$ such that

$$T_r \cdots T_1 A = D_n(d).$$

Since all elementary matrices are invertible we get then

$$A = T_1^{-1} \cdots T_r^{-1} D_n(d)$$

and we see that (159) is satisfied with $S := T_1^{-1} \cdots T_r^{-1} \in \mathrm{SL}_n(F)$. □

Note that we can of course transform an invertible $n \times n$-matrix using elementary *column* transformations of type I to a matrix of the special form (158). That is, we get similar to the above result a decomposition

$$A = D_n(d')S' \tag{160}$$

of A with $S' \in \mathrm{SL}_n(F)$ and a non-zero $d' \in F$.

The previous theorem shows that the matrices of the special linear group $\mathrm{SL}_n(F)$ are by far not as special as one would imagine on the first sight. The elements of $\mathrm{SL}_n(F)$ are – upto a simple factor of the simple form $D_n(d)$ – already all elements of $\mathrm{GL}_n(F)$![6]

Now a natural uniqueness question arises from the existence of the decomposition (159). And further, does in (159) and (160) hold $d = d'$. If we could give a positive answer to those questions, then we could assign in a non-trivial way to every invertible matrix A over F a non-zero element $d = d(A) \in F$ as an invariant. In other words: according to Theorem 3.55 we know that we can transform every invertible $n \times n$-matrix A over F – only by using elementary row transformations of type I – into the form $D_n(d)$ for some non-zero element $d \in F$. Ineviteable we

[6]The mathematical precise result is that the general linear group is isomorphic (as groups) to a semi-direct product of the special linear group and the multiplicative group $F^\times$ of the field F, in symbols $\mathrm{GL}_n(F) \cong \mathrm{SL}_n(F) \rtimes F^\times$.

have to ask ourself what kind of character this number has. Does it only depend on the original matrix A? Does always appear the same number d regardless how we transform the matrix A with the help of elementary transformations of type I into the form (158)? Let us formulate this question more precisely in the following way.

Problem 3. Is it possible to assign every invertible $n \times n$-matrix over a field F an non-zero element $d = d(A)$ of F such that this number does not change under a row transformation of type I and such that we assign this way to the (invertible) matrix $D_n(d)$ precisely the element d?

The next theorem gives the following complete answer to this problem.

Theorem 3.56. *For every element $A \in \mathrm{GL}_n(F)$ there exists precisely one decomposition of the form*

$$A = SD_n(d)$$

with $S \in \mathrm{SL}_n(F)$ and a non-zero $d \in F$. Likewise the decomposition in (160) *is unique and we have the equality $d = d'$.*[7]

But we are not yet able to proof this theorem because we will need more theoretical concepts. A direct proof of Theorem 3.56 would be desireable but the claim of this theorem is not evident. We have a similar situation as we had in the end of Chapter 1 with Problem 1. There we were not able to answer this problem before we introduced a new concept, namely the rank of a matrix in Chapter 2. Similar we will need to find a new, suitable invariant for matrices to answer this problem. This will lead to the concept of *determinants* of a $n \times n$-matrix. If we have introduced this new concept, then the proof of Theorem 3.56 – and with it the solotion of Problem 3 – will turn out to be very simple.[8]

So far we have only shown how row and column transformations of type I are described using matrices. Now we shall complete the list of elementary transformations by showing how an elementary row or column transformation of type II and III are described with the help of matrix multiplication.

Proposition 3.57. *Let A be a $m \times n$-matrix over a field F. Then the multiplication of the matrix A from right with a diagonal matrix of the form*

$$D_i^{(n)}(a) := \begin{pmatrix} 1 & & & \vdots & & & \\ & \ddots & & \vdots & & 0 & \\ & & 1 & \vdots & & & \\ \cdots & \cdots & \cdots & a & \cdots & \cdots & \cdots \\ & & & \vdots & 1 & & \\ & 0 & & \vdots & & \ddots & \\ & & & \vdots & & & 1 \end{pmatrix} \tag{161}$$

(the only non-zero entries are on the main diagonal and are all equal to 1 except the entry at position (i,i) which is equal to $a \in F$) performs the following elementary column transformation of type III: replacing the i-th column v_i of A by av_i, $a \in F$.

[7] Note that we do not have necessarily the equality $S = S'$!

[8] See page 98 where we will finally carry out the proof of Theorem 3.56.

Similarly the same row transformation will be performed, if A is multiplied from the left with the matrix $D_i^{(m)}$.

Finally, an elementary column or row transformation of type II – that is switching the i-th column (or row) with the j-th column (or row) of the matrix A – is obtained by multiplying the matrix A from right (or left) with a matrix of the form

$$R_{ij}^{(n)} := \begin{pmatrix} \ddots & & \vdots & & & & \vdots & & \\ & & \vdots & & & & \vdots & & \\ & 1 & \vdots & & & & \vdots & & \\ & & 0 & \cdots & \cdots & \cdots & 1 & & \\ & & \vdots & 1 & & & \vdots & & \\ & & \vdots & & \ddots & & \vdots & & \\ & & \vdots & & & 1 & \vdots & & \\ & & 1 & \cdots & \cdots & \cdots & 0 & & \\ & & \vdots & & & & \vdots & 1 & \\ & & \vdots & & & & \vdots & & \ddots \end{pmatrix} \tag{162}$$

which is derived from the $n \times n$-identity matrix by switching the i-th with the j-th column.

PROOF. The proof of this proposition is left as an exercise. □

It is apparent that claims of the type

> *"A $m \times n$-matrix A over F can be transformed with the help of certain elementary transformations to a matrix A'."*

can be formulated using suitable products of matrices of the form (149), (161) and (162).

Recal for example Theorem 2.48 from the previous chapter. It states that we can transform any matrix A of rank r with the help of elementary column and row transformations to a matrix A' of the form

$$A' = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

We can rewrite this theorem in our new language in the following way:

Theorem 3.58. *Let A be an $m \times n$-matrix over the field F. Then there exists matrices $P \in \mathrm{GL}_m(F)$ and $Q \in \mathrm{GL}_n(F)$ such that the equality*

$$PAQ = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \tag{163}$$

holds. Here $r = \operatorname{rank} A$ and I_r denotes the $r \times r$-identity matrix.

PROOF. From Theorem 2.48 we know that we can transform the matrix A using elementary row and column transformations into the desired form. Thus there exists invertible matrices $P_1, \ldots, P_s$ and $Q_1, \ldots, Q_t$ of the form (149), (161) and (162) such that

$$P_s \cdots P_2 P_1 A Q_1 Q_2 \cdots Q_t = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \tag{164}$$

Therefore (163) holds with $P := P_s \cdots P_2 P_1$ and $Q := Q_1 Q_2 \cdots Q_t$. The claim that $r = \operatorname{rank} A$ follows from the observation that the rank of a matrix is invariant under elementary row and column transformations and that the matrix of the right hand side of (163) has apparently rank r. □

Note that the content of Theorem 3.58 is nothing else than what we have said before in Proposition 3.45 even though we gave there a complete different kind of proof. But now the theorem tells also how one can establish the equivalence

$$A \sim \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, \qquad r = \operatorname{rank} A$$

of matrices of $F^{m,n}$ in an effective way, namely by applying suitable elementary column and row transformations using the recipe from Chapter 1.

Now consider the special case of quadratic $n \times n$-matrices. If one obtains – after the algorithm treminates – the $n \times n$-identity matrix I then $\operatorname{rank} A = n$ and A is an invertible matrix. In this way we can not only verify whether a given $n \times n$-matrix A is invertible or not, but we also get tool to compute the inverse matrix A^{-1} for a given invertible matrix A. If (164) is satisfied with $r = n$, then

$$P_s \cdots P_2 P_1 A Q_1 Q_2 \cdots Q_t = I.$$

Therefore we get[9]

$$\begin{aligned} A &= P_1^{-1} P_2^{-1} \cdots P_s^{-1} I Q_t^{-1} \cdots Q_2^{-1} Q_1^{-1} \\ &= P_1^{-1} P_2^{-1} \cdots P_s^{-1} Q_t^{-1} \cdots Q_2^{-1} Q_1^{-1} \\ &= (Q_1 Q_2 \cdots Q_t P_s \cdots P_2 P_1)^{-1}. \end{aligned}$$

and thus we have for the inverse matrix A^{-1} of the matrix A the equality

$$\begin{aligned} A^{-1} &= \big((Q_1 Q_2 \cdots Q_t P_s \cdots P_2 P_1)^{-1}\big)^{-1} \\ &= Q_1 Q_2 \cdots Q_t P_s \cdots P_2 P_1. \end{aligned} \tag{165}$$

Now in the case that A is invertible Theorem 3.55 states that actually elementary row transformations are already enough to achive the transformation of A to I. That is we can in this case assume with out any loss of generality that

$$P_s \cdots P_2 P_1 A = I \tag{166}$$

where $P_1, \ldots, P_{s-1}$ are elementary matrices and $P_s = D_n(d^{-1})$ is a diagonal matrix of the form (158).[10] Then (165) is just

$$\begin{aligned} A^{-1} &= P_s \cdots P_2 P_1 \\ &= P_s \cdots P_2 P_1 I. \end{aligned} \tag{167}$$

If one interpretes the equality (167) by means of elementary row transformations of matrices we get the following result.

Proposition 3.59 (Calulating the Inverse Matrix)**.** *Let A be an $n \times n$-matrix over the field F. If one can transform the matrix A with elementary row transformations to the identity matrix I then A is invertible. If one applies the same row transformations in the same order to the identity matrix I, then the indentity matrix transforms to the inverse A^{-1} of A.* □

Example. We want to determine the invers matrix of the matrix

$$A := \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

[9] Using amongst othere the following easily to verify calculation rules for elements g, h of a group G: $(gh)^{-1} = h^{-1} g^{-1}$ and $(g^{-1})^{-1} = g$. The proof of these rules are left as an exercise.

[10] Note that the value of s and the matrices P_i in (166) are not necessarily the same as in (165).

Therefore we write the matrix A and the identity matrix I next to each other and write blow this what we obtain by succesively applying the same elementary row transformations to both matrices:

$$\begin{array}{cccc|cccc}
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\hline
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & -1 & -1 & -1 & 1 & 0 & 0 \\
0 & -1 & 0 & -1 & -1 & 0 & 1 & 0 \\
0 & -1 & -1 & 0 & -1 & 0 & 0 & 1 \\
\hline
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & -1 & 0 & -1 & -1 & 0 & 1 & 0 \\
0 & -1 & -1 & 0 & -1 & 0 & 0 & 1 \\
0 & 0 & -1 & -1 & -1 & 1 & 0 & 0 \\
\hline
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 \\
0 & 0 & -1 & -1 & -1 & 1 & 0 & 0 \\
\hline
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & -2 & -1 & 1 & 1 & -1 \\
\hline
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & -1 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\
\hline
 & & \vdots & & & & \vdots & \\
\hline
1 & 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
0 & 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
0 & 0 & 1 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\
0 & 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}
\end{array}$$

This means that the matrix A is indeed invertible and its inverse A^{-1} is:

$$A^{-1} = \begin{pmatrix}
-\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\
\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\
\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}
\end{pmatrix}$$

12. Application to Systems of Linear Equations (Again)

Recall that in a note on page 66 we have already mentioned that we can see Chapter 1 also in the view of linear maps. We return swiftly to this topic. Therefore consider the system of linear equations

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\vdots \quad\quad \vdots \quad\quad \vdots \quad\quad \vdots & \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned} \tag{168}$$

of m equations in the n unknown variables $x_1, \ldots, x_n$ over a field F. Then to find all solution to this system of equations is equivalent to find all solutions x to the equation

$$Ax = b \tag{169}$$

where A is the simple coefficient matrix of (168) and $b \in F^m$ the vector with the coefficients $b_1, \ldots, b_m$.

To find all solution to the homogeneous part of (168) is equivalent to find all solutions to

$$Ax = 0. \tag{170}$$

Thus we see that the set of all solutions to the homogeneous part of (168) is precisely the kernel of the linear map A: $F^n \to F^m$, that is in the notation Chapter 1 we have

$$M_0 = \ker A,$$

in particular – since we know that $\ker A$ is a linear subspace of F^n – the set of all solutions M_0 is a subspace of F^n which is the content of Proposition 1.3 in Chapter 1.

Apparently the set M of all solutions to (169) is given by

$$M = x' + M_0$$

where $x' \in F^n$ is some element such that $Ax' = b$ is satisfied, which is Proposition 1.2.

Now (169) is solvable if and only if

$$b \in \operatorname{im} A \tag{171}$$

and in this case the solution is unique if and only if the linear map A is a monomorphism which is equivalent with

$$\ker A = 0. \tag{172}$$

Proposition 1.7 of Chapter 1 is a consequence of the dimension formula for linear maps: By this formula we have

$$\dim F^n = \dim(\operatorname{im} A) + \dim(\ker A)$$

and thus

$$\begin{aligned} \dim(\ker A) &= \dim F^n - \dim(\operatorname{im} A) \\ &\geq \dim F^n - \dim F^m \\ &= n - m \end{aligned}$$

where the inequalty is due to $\dim(\operatorname{im} A) \leq \dim F^m$. Thus $\dim(\ker A) > 0$ if $n > m$ and in this case there must exists a non-trivial solution to (170) in this case.

In the special case of $m = n$ we know from the dimension formula for linear maps that the linear map A: $F^n \to F^n$ is an epimorphism if and only if it is a monomorphism. Thus $b \in \operatorname{im} A$ for every $b \in F^n$ if and only if $\ker A = 0$, that is, if and only if the homogeneous equation (170) has only the trivial solution. This proves Proposition 1.8 in Chapter 1.

CHAPTER 4

Determinants

1. The Concept of a Determinant Function

In this section F denotes always a field. While we have studied in Section 11 of the previous chapter the general linear group of a vector space we have encountered in a natural way the problem (see Problem 3 on page 85) whether there exists a function

$$d\colon\ \mathrm{GL}_n(F) \to F$$

with the following two properties:

(1) If $A, A' \in \mathrm{GL}_n(F)$ are two matrices and A' is obtained from A by an elementary column transformation of type I, then

$$d(A') = d(A).$$

(2) For every matrix $D_n(a) \in \mathrm{GL}_n(F)$ of the form (158) we have

$$d(D_n(a)) = a.$$

One can show – the proof is left as an exercise – that the above two properties are equivalent with the following three properties:

(1) If $A, A' \in \mathrm{GL}_n(F)$ are two matrices and A' is obtained from A by an elementary column transformation of type I, then

$$d(A') = d(A).$$

(2) If $A, A' \in \mathrm{GL}_n(F)$ are two matrices and A' is obtained from A by multiplying a column with a non-zero element $a \in F$, then

$$d(A') = ad(A).$$

(3) For the identity matrix $I \in \mathrm{GL}_n(F)$ holds

$$d(I) = 1.$$

Now these considerations motivate the following definition of a *determinant function*. (Note that we include in this definition also non-invertible matrices.)

Definition 4.1 (Determinant Function)**.** A map

$$d\colon\ F^{n,n} \to F$$

is called a *determinant function* if it satisfies the following three properties:

(1) If $A, A' \in F^{n,n}$ are two matrices and A' is obtained from A by replacing the i-th column with the sum of the i-th and j-th column ($1 \leq i, j \leq n$, $i \neq j$), then

$$d(A') = d(A).$$

(2) If $A, A' \in F^{n,n}$ are two matrices and A' is obtained from A by multiplying a column with a non-zero element $a \in F$, then

$$d(A') = ad(A).$$

(3) For the identity matrix I holds

$$d(I) = 1.$$

Note that if one identifies a $n \times n$-matrix A over F with system $v_1, \ldots, v_n$ of its columns then a determinant function is nothing else then a map which maps n-tuples of vectors of F^n to the field F, in symbols

$$d\colon (F^n)^n \to F,$$

and which satisfies with the following three properties:

(1) For every system $v_1, \ldots, v_n$ of vectors of F^n and every $1 \leq i, j \leq n$ with $i \neq j$ holds

$$d(v_1, \ldots, v_{j-1}, v_j + v_i, v_{j+1}, \ldots, v_n) = d(v_1, \ldots, v_n).$$

(2) For every system $v_1, \ldots, v_n$ of vectors of F^n, every $a \in F$ and every $1 \leq j \leq n$ holds

$$d(v_1, \ldots, v_{j-1}, av_j, v_{j+1}, \ldots, v_n) = ad(v_1, \ldots, v_n).$$

(3) For the canonical basis $e_1, \ldots, e_n$ of F^n holds

$$d(e_1, \ldots, e_n) = 1.$$

Depending on which point of view is more convenient we will consider in the following a determinant function either to be a map from all $n \times n$-matrices to the field F or we will consider it as a function of all n-tuples of vectors of F^n.

Proposition 4.2. *Let A and A' be $n \times n$ matrices over F and let $d\colon F^{n,n} \to F$ be a determinant function. Then the following three statements are true:*

(1) *If A' is obtained from A by adding the a times the j-th column of A to the i-th column of A ($a \in F$, $i \neq j$), then*

$$d(A') = d(A). \tag{173}$$

(2) *If A' is obtained from A by exchanging two columns, then*

$$d(A') = -d(A). \tag{174}$$

(3) *If A' is obtained from A by multiplying the i-th column by $a \in F$, then*

$$d(A') = ad(A). \tag{175}$$

PROOF. We denote in this proof the colums of the matrix A by $v_1, \ldots, v_n$.

(1) In order to avoid triviality we may assume that $a \neq 0$. Then

$$\begin{aligned} d(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) &= a^{-1} d(v_1, \ldots, v_i, \ldots, av_j, \ldots, v_n) \\ &= a^{-1} d(v_1, \ldots, v_i + av_j, \ldots, av_j, \ldots, v_n) \\ &= d(v_1, \ldots, v_i + av_j, \ldots, v_j, \ldots, v_n). \end{aligned}$$

(2) We have

$$\begin{aligned} d(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) &= d(v_1, \ldots, v_i + v_j, \ldots, v_j, \ldots, v_n) \\ &= d(v_1, \ldots, v_i + v_j, \ldots, v_j - (v_j + v_i), \ldots, v_n) \\ &= d(v_1, \ldots, v_i + v_j, \ldots, -v_i, \ldots, v_n) \\ &= d(v_1, \ldots, v_j, \ldots, -v_i, \ldots, v_n) \\ &= -d(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n). \end{aligned}$$

(3) This is just the second property of a determinant function. □

In the mathematical language there is another common notation for invertible and non-invertible matrices:

Definition 4.3. Let A be a $n \times n$-matrix over the field F. Then A is called *singular* if A is not invertible. Likewise A is called *non-singular* (or *regular*) if it is invertible, that is if $A \in \mathrm{GL}_n(F)$.

Proposition 4.4. *Let $d\colon F^{n,n} \to F$ be a determinant function and $A \in F^{n,n}$ a matrix. Then A is singular if and only if $d(A) = 0$.*

PROOF. "⇒": We know that a matrix is singular if it doesn't have full rank, that is if $\operatorname{rank} A < n$. This is the case if and only if there exists one colum v_i which is a linear combination of the remaining $n-1$ columns. Without any loss of generality we may assume that

$$v_1 = a_2 v_2 + \ldots + a_n v_n$$

for some elements $a_2, \ldots, a_n$.

$$\begin{aligned} d(v_1, \ldots, v_n) &= d(v_1 - a_2 v_2 - \ldots - a_n v_n, v_2, \ldots, v_n) \\ &= d(0, v_2, \ldots, v_n) \\ &= 0 \end{aligned}$$

where the last equality follows from the second property of a determinant function.

"⇐": We assume that A is non-singular. Then Theorem 3.55 states that we can write

$$A = SD_n(a)$$

for some $S \in \mathrm{SL}_n(F)$ and a non-zero $a \in F$ where $D_n(a)$ is the diagonal matrix of the special form (158). But this means that A is derived from the identity matrix by multiplying the last column with a and furthermore only elementary transformations of type I. Thus $d(A) = a$ and since $a \neq 0$ it follows that $d(A) \neq 0$. Thus necessarily A must be singular if $d(A) = 0$. □

Note that we still do not know whether determinant functions exist! In order to find an answer to the question whether determinant functions exists in general we proceed in a way which is common to existence probelms in mathematics: often one studies the hypothetical properties of an object – which is postulated to exists – in detail until one is able to actually prove the existence of the object in question or until one gathers enough evidence which rules out the possibility that the object in question can exist. So let us continue with the studies.

Let $d\colon F^{n,n} \to F$ and $d'\colon F^{n,n} \to F$ be two determinant functions. Then we know already that $d(A) = 0 = d'(A)$ if $A \in F^{n,n}$ is a singular matrix. On the other hand we have seen in the previous proof that if $A = SD_n(a)$, then $d(A) = a$ and for the same reason $d'(A) = a$, too. We can summarize this observation in the following result.

Proposition 4.5. *Let $d\colon F^{n,n} \to F$ and $d'\colon F^{n,n} \to F$ are two determinant functions, then $d(A) = d'(A)$ for every $A \in F^{n,n}$. In other words there exists at most one determinant function on $F^{n,n}$.* □

Proposition 4.6. *A determinant function $d\colon F^{n,n} \to F$ is "linear in every column", that is for every $1 \leq i \leq n$ holds*

$$\begin{aligned} &d(v_1, \ldots, v_{i-1}, av_i + bw_i, v_{i+1}, \ldots, v_n) = \\ &\quad = ad(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n) + bd(v_1, \ldots, v_{i-1}, w_i, v_{i+1}, \ldots, v_n) \end{aligned} \tag{176}$$

for all $v_1, \ldots, v_i, w_i, \ldots, v_n$ in F^n and all $a, b \in F$.

PROOF. Note first that due to (175) it is enough to verify the claim for the special case $a = b = 1$. Due to (174) we can furthermore assume with out any loss of generality that $i = 1$.

Since the vector space F^n has dimension n it follows that the system

$$v_1, w_1, v_2, \ldots, v_n$$

of $n+1$ vectors in F^n is linear dependent. Therefore there exists a non-trivial linear combination

$$cv_1 + c'w_1 + c_2v_2 + \ldots + c_nv_n = 0 \tag{177}$$

of the zero vector with $c, c', c_2, \ldots, c_n \in F$. Now either $c = c' = 0$ or not.

Assume first that $c = c' = 0$. Then (177) is a non-trivial linear combination of the form

$$c_2v_2 + \ldots + c_nv_n = 0$$

and thus $\operatorname{rank}(v_2, \ldots, v_n) < n - 1$. But then all the terms in (176) are equal to 0 and therefore the equality holds (compare with Proposition 4.4).

Thus it remains to verify the case that $c \neq 0$ or $c' \neq 0$. It is enough to verify one of the two possibilities, say $c \neq 0$. Without any loss of generality we may then also assume that in this case $c = -1$. Thus from (177) follows that

$$v_1 = c'w_1 + c_2v_2 + \ldots + c_nv_n.$$

Using this and repeatedly (173) we get

$$\begin{aligned}
d(v_1 + w_1, v_2, \ldots, v_n) &= d(c'w_1 + c_2v_2 + \ldots + c_nv_n + w_1, v_2, \ldots, v_n) \\
&= d((c'+1)w_1 + c_2v_2 + \ldots + c_nv_n, v_2, \ldots, v_n) \\
&= d((c'+1)w_1 + c_2v_2 + \ldots + c_{n-1}v_{n-1}, v_2, \ldots, v_n) \\
&= \ldots = \\
&= d((c'+1)w_1 + c_2v_2, v_2, \ldots, v_n) \\
&= d((c'+1)w_1, v_2, \ldots, v_n) \\
&= (c'+1)d(w_1, v_2, \ldots, v_n) \\
&= c'd(w_1, v_2, \ldots, v_n) + d(w_1, v_2, \ldots, v_n)
\end{aligned}$$

and on the other hand

$$\begin{aligned}
d(v_1, v_2, \ldots, v_n) &= d(c'w_1 + c_2v_2 + \ldots + c_nv_n, v_2, \ldots, v_n) \\
&= d(c'w_1 + c_2v_2 + \ldots + c_{n-1}v_{n-1}, v_2, \ldots, v_n) \\
&= \ldots = \\
&= d(c'w_1 + c_2v_2, v_2, \ldots, v_n) \\
&= d(c'w_1, v_2, \ldots, v_n) \\
&= c'd(w_1, v_2, \ldots, v_n).
\end{aligned}$$

Combining these two equations we get that also in this case

$$d(v_1 + w_1, v_2, \ldots, v_n) = d(v_1, v_2, \ldots, v_n) + d(w_1, v_2, \ldots, v_n)$$

is true. □

2. Proof of Existence and Expansion of a Determinant with Respect to a Row

Theorem 4.7 (Existence of a Determinant)**.** *Let F be a field. Then for every natural number $n \geq 1$ exists precisely one determinant function $d\colon F^{n,n} \to F$. We denote this function by* det *or more precise* $\det_n$.

PROOF. We know already that if a determinant function exists, then it is unique. We did prove this in Proposition 4.5. Thus it remains to verify the the existence of a determinant function. We will carry out this proof by induction with respect to n.

"$n = 1$": If $A = (a)$ is a 1×1-matrix, then $\det_1(A) := a$ defines apparently the determinant function.

"$n - 1 \Rightarrow n$": We assume that $n > 1$ and that we have a determinant function $\det_{n-1}\colon F^{n-1,n-1} \to F$ is given. We want to construct a map $d\colon F^{n,n} \to F$ using $\det_{n-1}$. For every $n \times n$-matrix $A = (a_{ij})$ over F define

$$d(A) := \sum_{j=1}^{n} (-1)^{n-j} a_{nj} \det_{n-1}(A_{nj}). \tag{178}$$

Here A_{nj} denotes the $(n-1) \times (n-1)$-matrix which is derived from A by leaving away the n-th row and j-th column. We need to verify that the so defined map is indeed a determinant function. Thus we need to verify all the three properties of a determinant function from Definition 4.1, one by one.

(1) Assume that the matrix $A' = (a'_{rs})$ is obtained from the matrix $A = (a_{rs})$ by replacing the i-th column v_i of A by $v_i + v_k$ where v_k ist the k-th column of A and $i \leq k$. Then

$$a'_{ni} = a_{ni} + a_{nk} \qquad \text{and} \qquad a'_{nj} = a_{nj} \quad \text{for } j \neq i.$$

Furthermore we have $A'_{ni} = A_{ni}$ and if $j \neq i, k$ then

$$\det_{n-1}(A'_{nj}) = \det_{n-1}(A_{nj}).$$

Finally it follows from Proposition 4.6 that $\det_{n-1}(A'_{nk})$ can be written as

$$\det_{n-1}(A'_{nk}) = \det_{n-1}(A_{nk}) + \det_{n-1}(B)$$

where B is a matrix which is derived from the matrix A_{in} by shifting the k-th row past $|k - i| - 1$ rows. Every time v_k is shifted by one row the sign of the determinant gathers an additional factor of -1. Thus we get

$$\det_{n-1}(B) = (-1)^{k-i-1} \det_{n-1}(A_{ni})$$

(note that $(-1)^{|k-i|-1} = (-1)^{k-i-1}$) and alltogether

$$\det_{n-1}(A'_{nk}) = \det_{n-1}(A_{nk}) + (-1)^{k-i-1} \det_{n-1}(A_{ni}).$$

Collecting this information and using (178) we get

$$\begin{aligned} d(A') - d(A) &= (-1)^{n-i}\big[a'_{ni} \det_{n-1}(A'_{ni}) - a_{ni} \det_{n-1}(A_{ni})\big] \\ &\quad + (-1)^{n-k}\big[a'_{nk} \det_{n-1}(A'_{nk}) - a_{nk} \det_{n-1}(A_{nk})\big] \\ &= (-1)^{n-i} a_{nk} \det_{n-1}(A_{ni}) + (-1)^{n-k} a_{nk} \det_{n-1}(A_{ni}) = 0 \end{aligned}$$

and therefore we have indeed $d(A') = d(A)$.

(2) Assume that the matrix $A' = (a'_{rs})$ is obtained from $A = (a_{rs})$ by multiplying the i-th column with a number $a \in F$. Then

$$a'_{ni} = aa_{ni} \qquad \text{and} \quad a'_{nj} = a_{nj} \quad \text{for } j \neq i.$$

Furthermore we have $A'_{ni} = A_{ni}$ and if $j \neq i$ then

$$\det_{n-1}(A'_{nj}) = a \det_{n-1}(A_{nj})$$

because for $j \neq i$ the matrix A'_{nj} is obtained from A_{nj} by multiplying a column of A_{nj} with the element a. Using this information we obtain from (178) indeed the desired equality $d(A') = ad(A)$:

$$\begin{aligned} d(A') &= \sum_{j=1}^{n} (-1)^{n-j} a'_{nj} \det_{n-1}(A'_{nj}) \\ &= a \sum_{j=1}^{n} (-1)^{n-j} a_{nj} \det_{n-1}(A_{nj}) = ad(A). \end{aligned}$$

(3) If $A = (a_{rs})$ is the $n \times n$-identity matrix I_n over F, then $a_{n,1} = a_{n,2} = \ldots = a_{n,n-1} = 0$ and $a_{nn} = 1$. Furthermore $A_{nn} = I_{n-1}$. Therefore $d(I_n) = \det_{n-1}(I_{n-1}) = 1$ by (178).

Thus the map $d\colon F^{n,n} \to F$ defined by (178) satisfies all properties of a determinant function. Therefore we define $\det_n := d$. And this concludes the induction step "$n-1 \Rightarrow n$". □

Now after this lengthish technical proof we know that for every field F and every $n \geq 1$ there exists a unique map $F^{n,n} \to F$ satisfying the properties of a determinant function as defined in Definition 4.1, namely $\det_n$.

Definition 4.8. Let F be a field and let det be the unique determinant function

$$\det\colon F^{n,n} \to F.$$

If $A \in F^{n,n}$ then the number $\det(A) \in F$ is called the *determinant* of A. The determinant of A is also denoted by $|A|$, that is

$$\begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix} := \det(A).$$

But what kind of map is this? The answer will be given by the following

Proposition 4.9. *Let F be a field and $n \geq 1$. Let $A = (a_{rs})$ be a $n \times n$ matrix over F. Then $\det_n(A)$ is a polynomial of degree n in the n^2 variables $a_{11}, \ldots, a_{nn}$.*

Proof. We prove this claim again by induction with respect to n.

"$n = 1$": Apparently $\det_1(A) = a_{11}$ is a polynomial of degree 1 in the single variable a_{11}.

"$n-1 \Rightarrow n$": We assume that $\det_{n-1}$ is a polynomial of degree $n-1$ in $(n-1)^2$ variables. Then from (4.1),

$$\det_n(A) = \sum_{j=1}^{n} (-1)^{n-j} \underbrace{a_{nj} \det_{n-1}(A_{nj})}_{\text{polynomial of degree } n},$$

follows that $\det_{n-1}$ is apparently a polynomial of degree n in the n^2 variables $a_{11}, \ldots, a_{nn}$. This proves the induction step. □

The first three determinant functions are then explicitly the following:

$$\begin{aligned}
\det_1(A) &= a_{11} \\
\det_2(A) &= a_{11}a_{22} - a_{12}a_{21} \\
\det_3(A) &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\
&\qquad - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}
\end{aligned}$$

Note that the number of terms in those polynomials is increasing drastically with n. For $n = 1$ we have one summand, for $n = 2$ we have $2 = 1 \cdot 2$ summands, for $n = 3$ we have $6 = 1 \cdot 2 \cdot 3$ summands. The general rule is that $\det_n(A)$ is a polynomial with $n! = 1 \cdot 2 \cdot 3 \cdots n$, that is *$n$-factorial* summands. For example $\det_{10}(A)$ is a polynomial with 3 628 800 summands in 100 variables! Thus it becomes apparent that it is difficult to memorize those polynomials and that we need to develope calculation methodes to simplify the calculation of a determinant in the concrete case.

One way to simplify the calculation can be obtained from the equation (178) which can also be written as

$$\det(A) = \sum_{j=1}^{n} (-1)^{n+j} a_{nj} \det(A_{nj})$$

with the help of the equality $(-1)^{n-j} = (-1)^{n+j}$. If the last row of the matrix A contains many coefficients which are equal to 0, then the right hand side of the above equation contains only few summands. In the extreme case that *only one* coefficient of the last row of A is different from zero, say only the j-th coeffient $a_{nj} \neq 0$, then the above equation simplifies to

$$\det(A) = (-1)^{n+j} a_{nj} \det(A_{nj}).$$

Since A_{nj} is only a $(n-1) \times (n-1)$-matrix it is apparent that the right hand side of this equation is easier to calculate than the left hand side.

But can we gain a similar advantage if another row of the matrix A – say the i-th row – is sparsely populated with non-zero coefficients? The answer to this question is: yes, we can! Assume that A' is the $n \times n$-matrix whichi is obtained from A by shifting the i-th row downwards past the $n - i$ rows below it. This is done by $n - i$-times exchanging two adjacent rows and thus we have the equality $\det(A') = (-1)^{n-i} \det(A)$. Therefore

$$\begin{aligned}
\det(A) &= (-1)^{i-n} \det(A') \\
&= \sum_{j=1}^{n} (-1)^{i+j} a'_{nj} \det(A'_{nj})
\end{aligned}$$

and after using the apparent equalities $a'_{nj} = a_{ij}$ and $A'_{nj} = A_{ij}$ we get

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

Thus we have obtained the following general formula to "*expand the determinant of A along the i-th row*".

Proposition 4.10. *Assume that A is an $n \times n$-matrix over the field F. Then we have for every $1 \leq i \leq n$ the equality*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \tag{179}$$

where A_{ij} denotes the $(n-1) \times (n-1)$-matrix which is obtained from A by leaving away the i-th row and j-th column.

Example. Consider the following calculation:

$$\begin{vmatrix} 0 & 10 & 4 & 0 \\ 2 & 3 & 1 & 5 \\ 0 & 0 & 3 & 0 \\ 4 & 6 & 2 & 3 \end{vmatrix} = 3 \begin{vmatrix} 0 & 10 & 0 \\ 2 & 3 & 5 \\ 4 & 6 & 3 \end{vmatrix} = 30 \begin{vmatrix} 2 & 5 \\ 4 & 3 \end{vmatrix} = 30(2 \cdot 3 - 5 \cdot 4) = -420.$$

Here we have expanded the determinant in the first step along the 3-rd row and then in the next step we have expanded it along the 1-st row. Notice how fast the determinant shrinks in size. Only in the last step there was no way to gain simplicity by applying previous proposition.

Before we begin to study the determinant function more closely we shall give now the **Answer to Problem 3** (which has been stated on page 85) by finally verifying Theorem 3.56.

Proof of Theorem 3.56. Let $A \in \mathrm{GL}_n(F)$. Then by Theorem 3.55 there exists a decomposition of A of the form

$$A = D_n(a)S \tag{180}$$

with $S \in \mathrm{SL}_n(F)$ and $a \in F^\times$. We want to show that this decomposition is unique. Assume therefore that there exists another decomposition of A of this form, say

$$A = D_n(a')S'$$

with $S' \in \mathrm{SL}_n(F)$ and $a' \in F^\times$. Then

$$D_n(a)S = D_n$$

and it follows

$$D_n(a) = D_n(a')S'S^{-1}$$

Since $\mathrm{SL}_n(F)$ is a group it we have that $S'S^{-1} \in \mathrm{SL}_n(F)$ and thus the above equation means that $D_n(a)$ is obtained from $D_n(a')$ by elementary column transformations of type I. Therefore necessarily $\det D_n(a) = \det D_n(a')$. Since $D_n(a)$ is obtained from the identity matrix I by multiplying the last column by a it follows that $\det D_n(a) = a \det I = a$. Likewise $\det D_n(a') = a'$ and thus we must have that $a = a'$ and therefore $D_n(a) = D_n(a')$. Multiplying the previous equation from the left with $D_n(a)^{-1}$ and from the right with S yields then the equality

$$S = S'.$$

Therefore the decomposition (180) is uniue.

We still have to prove the second claim of Theorem 3.56, namely if have a decomposition of A of the form

$$A = S'D_n(a') \tag{181}$$

then necessarily $a' = a$ where a is the element of $F^\times$ which appeared in (180). Interpreting (180) in terms of column transformations we see that A is obtained

from the identity matrix I by applying elementray column transformations of type I and then by multiplying the the last column by a'. Therefore

$$\det A = a' \det I = a'.$$

Similarly, if we interprete (181) in terms of column transformations it follows that $\det A = a$ since A is obtained from the identity matrix by multiplying the last column by a and then applying column transformations of type I only. Therefore $a = a'$. □

3. Elementary Properties of a Determinant

Proposition 4.11. *A $n \times n$-matrix A over the vield F is invertible if and only if* $\det A \neq 0$. *In other words:*

$$A \text{ is singular} \iff \det A = 0.$$

Proof. This is precisely Proposition 4.4. □

Proposition 4.12. *The determinant function is a* multiplicative map *in the following sense: for every $A, B \in M_n(F)$ we have the equality*

$$\det AB = \det A \det B. \tag{182}$$

Proof. Consider first the case that $\det B = 0$. Then the right hand side of (182) is equal to 0. We have to show that in this case also the left side is equal to 0. Since $\det B = 0$ it follows from the previous proposition that B is singular. In particular this means that $\operatorname{rank} B < n$. But then also $\operatorname{rank} AB = \dim(\operatorname{im} AB) \leq \dim(\operatorname{im} B) = \operatorname{rank} B < n$ and thus the matrix AB is singular, too. But that implies that the left hand side of (182) is equal to 0. Therefore in this case the equation (182) is satisfied.

Thus we can consider the remanining case that B is invertible. By Theorem 3.55 there exists a $S \in \mathrm{SL}_n(F)$ and a non-zero $b \in F$ such that $B = SD_n(b)$. We get the equality

$$AB = ASD_n(b).$$

This means, the matrix AB is obtained from A by applying first several column transformations of type I and finally by multiplying the last column by the element b. Thus

$$\det AB = b \det A.$$

Since apparently $\det B = b$ we obtain from this the desired equality. □

Corollary. *If $A, B \in M_n(F)$ then we have the equality*

$$\det AB = \det BA.$$

If moreover A is invertible then we have the equality

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

Proof. The first claim is evident. If A is invertible then it follows from Proposition 4.12 that $\det A \det A^{-1} = \det AA^{-1} = \det I = 1$ and thus the claim is evident. □

Note that the content of Proposition 4.12 can be interpreted in the language of Algebra as follows: the determinant function is a "*group homomorphism*"[1]

$$\det\colon \mathrm{GL}_n(F) \to F^\times.$$

[1] A map $f\colon G \to G'$ of groups is called a group homomorphism if $f(xy) = f(x)f(y)$ for every $g, h \in G$. Thus a isomorphism of groups is a bijective group homomorphism.

where $F^\times$ is the multiplicative group of all non-zero (and thus inverible) elements of the field F.

Proposition 4.13 (Characterisation of the Special Linear Group)**.** *The special linear group* $\mathrm{SL}_n(F)$ *consists precisely of all* $n\times n$*-matrices* A *over* F *with* $\det A = 1$*. That is*

$$\mathrm{SL}_n(F) = \{A \in M_n(F) : \det A = 1\}.$$

PROOF. "$\Rightarrow$": Let $S \in \mathrm{SL}_n(F)$. Then S can be obtained from the identity matrix I by applying column transformations of type I only. Thus $\det S = \det I = 1$.

"$\Leftarrow$": Assume that $A \in M_n(F)$ with $\det A = 1$. By Proposion 4.11 we know that $A \in \mathrm{GL}_n(F)$. Thus it follows by Theorem 3.55 that there exists a $S \in \mathrm{SL}_n(F)$ and a non-zero $c \in F$ such that

$$A = SD_n(c)$$

and Proposition 4.12 follows that $\det A = \det S \det D_n(c)$. Since $S \in \mathrm{SL}_n(F)$ it follows that $\det S = 1$. Thus also $\det D_n(c) = 1$. On the other hand we have apparently $\det D_n(c) = c \det I = c$ and thus necessarily $c = 1$. Therefore $D_n(c) = I$ and it follows that $A = S$ is an element of $\mathrm{SL}_n(F)$. □

Note that sofar we have introduced the concept of a determinant only by looking at the columns of a matrix and only by considering elementary column transformations. Definition 4.1 is in no way symmetric with respect to columns and rows. Assume that $d\colon M_n(F) \to F$ is a function which satisfies the analogous conditions as in Definition 4.1 but just with respect to rows. We shall call such a function for the moment a *row determinant function*. Likewise we shall mean for the moment by a *column determinant function* a function which satisfies the conditions of Definition 4.1.

Furthermore, if $d\colon M_n(F) \to F$ is an arbitrary function then we shall denote by ${}^t d$ the function

$${}^t d\colon M_n(F) \to F, A \mapsto {}^t d(A) := d({}^t A)$$

where ${}^t A$ denotes the transposed matrix of A (see page 45). It is apparent that d is a row (column) determinant function if and only if ${}^t d$ is a column (row) determinant function. Thus everything we have said about so far about a column determinant function remains valid for row determinant functions. Furthermore it is a consequence of Theorem 4.7 that there exists precisely one row determinant function $d\colon M_n(F) \to F$, namely ${}^t\det$.

Theorem 4.14. *For every* $A \in M_n(F)$ *holds the equality*

$$\det A = \det {}^t A.$$

PROOF. We will prove this by showing that the column determinant function det is also a row determinant function. It follows then by Proposition 4.5 that $\det = {}^t\det$.

(1) Let $A, A' \in M_n(F)$ and assume that A' is obtained from A by a row transformation of type I, that is there exists a $S \in \mathrm{SL}_n(F)$ such that $A' = SA$. Then

$$\det A' = \det SA = \det S \det A$$

where the last equality is due Proposition 4.12. Since $S \in \mathrm{SL}_n(F)$ we have that $\det S = 1$ and thus $\det A' = \det A$.

(2) Let $A, A' \in M_n(F)$ and assume that A' is obtained from A by multiplying the i-th row by $a \in F$. That is $A' = D_i^{(n)}(a)A$ where $D_i^{(n)}(a)$ is the diagonal matrix (161). Then

$$\det A' = \det D_i^{(n)}(a)A = \det D_i^{(n)}(a) \det A$$

where the last equality is due Proposition 4.12. Since $\det D_i^{(n)}(a) = a$ we have then $\det A' = a \det A$.

(3) Clearly $\det I = 1$ is satisfied. □

Thus we have seen that we actually need not to distinguish between column and row determinant functions because they are actually the very same functions. Thus we shall abolish this notation again and we will from now on speak the *determinant function* or just *determinant*

$$\det\colon M_n(F) \to F.$$

As an consequence of the previous result we get a variation of the Proposition 4.10, namely how to expand a determinant with respect to an arbirary column. Precisely this is the following

Proposition 4.15. *Assume that A is an $n \times n$-matrix over the field F. Then we have for every $1 \leq i \leq n$ the equality*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ji} \det(A_{ji}) \tag{183}$$

where A_{ij} denotes the $(n-1) \times (n-1)$-matrix which is obtained from A by leaving away the i-th row and j-th column.

Proof. Denote by $A' = (a'_{ij})$ the transposed matrix of A. Then $a'_{ij} = a_{ji}$ and $A'_{ij} = A_{ji}$. Using this information and Proposition 4.10 we get

$$\begin{aligned} \det A &= \det {}^t A \\ &= \sum_{j=1}^{n} (-1)^{i+j} a'_{ij} \det(A'_{ij}) \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{ji} \det(A_{ji}) \end{aligned}$$

□

We can enrich the collection of identities given in Proposition 4.10 and 4.15 by two more identities which are less important but which then combined give a compact formula for matrix inversion.

Lemma 4.16. *Let $A \in M_n(F)$. Then we have for any $1 \leq i, k \leq n$ with $i \neq k$ the following two equalities:*

$$\sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{kj}) = 0 \tag{184}$$

$$\sum_{j=1}^{n} (-1)^{i+j} a_{ji} \det(A_{jk}) = 0 \tag{185}$$

Proof. Assume that A' is the matrix which is obtained from A by replacing the k-th row of A by the i-row of A. Then the system of vectors obtained from the rows of A' is linear dependent and thus $\det A' = 0$. If one expands A' along the k-th row one obtains precisely (184). The equality (185) is verified in a similar way. □

Definition 4.17 (Complimentary Matrix)**.** Let $A \in M_n(F)$. Then the *complimentary matrix* $\tilde{A}$ to A is defined to be the $n \times n$-matrix which has the coefficients

$$\tilde{a}_{ij} := (-1)^{i+j} \det_{n-1}(A_{ji}). \tag{186}$$

Note the twist in the indices i and j on both sides of the equation!

Theorem 4.18 (Cramer's Rule for the Complementary Matrix)**.** *For any $A \in M_n(F)$ we have the two equalities*

$$A\tilde{A} = \det(A)\, I \qquad \textit{and} \qquad \tilde{A}A = \det(A)\, I$$

Proof. Using the formula for the matrix product (see Theorem 3.31) we get that the coefficient in the i-th row and k-th column of the matrix $C := A\tilde{A}$ is given by

$$\begin{aligned} c_{ik} &= \sum_{j=1}^{n} a_{ij}\tilde{a}_{jk} \\ &= \sum_{j=1}^{n} a_{ij}(-1)^{j+k} \det_{n-1}(A_{kj}) \\ &= \sum_{j=1}^{n} (-1)^{j+k} a_{ij} \det_{n-1}(A_{kj}) \\ &= \begin{cases} \det A & \text{if } i = k, \\ 0 & \text{if } i \neq k, \end{cases} \end{aligned}$$

where the first case is due to Proposition 4.10 and the remaining case due to Lemma 4.16. But this means that $C = \det(A)\, I$.

The second equality is shown in a similar way or follows straight from the fact that $M_n(F)$ is a ring. □

Theorem 4.19 (Cramer's Rule for the Matrix Inversion)**.** *Let $A \in \mathrm{GL}_n(F)$. Assume that A' is the inverse matrix of A. Then the coefficients a'_{ij} of A' are given by the formula*

$$a'_{ij} = \frac{1}{\det A}(-1)^{i+j} \det(A_{ji}) \tag{187}$$

Proof. If A is invertible then $\det A \neq 0$ by Proposition 4.11. Thus we obtain from

$$\det(A) I = A\tilde{A}$$

the equation

$$A^{-1} = \frac{1}{\det A}\tilde{A}.$$

For $A^{-1} = (a'_{ij})$ this states then precisely (187). □

Note that the importance of the Cramer's rule for matrix inversion does lie so much in the ability to actually calculate the inverse of a matrix but rather in the theoretical content of it. Consider the case that $F = \mathbb{R}$ or $F = \mathbb{C}$. Since we know that the determinant is a polynomial in the coefficients of the matrix we know that it is a continuous function. Thus (187) states that the coefficients of the inverse matrix A^{-1} depend in a continuous way on the coefficients of the matrix A.

Example. Assume that the matrix

$$A := \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with coefficients in the field F is invertible. Then $\det A = ac - bd \neq 0$ and we have by above theorem the equality

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

In praticular if $A \in \mathrm{SL}_n(F)$ then $\det A = 1$ and we get the even more simple formula

$$A^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Let us turn to an special case where A is a square matrix over F of the form

$$A := \begin{pmatrix} B & C \\ 0 & B' \end{pmatrix}$$

where $B \in M_m(F)$, $B' \in M_n(F)$ and C is an arbitrary $m \times n$-matrix over F. If A is singular then it follows that either B, B' or both are singular, too. Thus trivialy

$$\det A = \det B \det B' \tag{188}$$

since both sides of this equation are equal to 0. We want to show that this equation is even true in the case that A is regular[2].

In the case that A is regular it follows that necessarily B and B' are regular, too (why?). Thus we can transform the matrix A by applying only elementary row transformations of type I to the first m rows into a matrix of the form

$$A' := \begin{pmatrix} D_m(b) & C' \\ 0 & B' \end{pmatrix}$$

with $b = \det B$ where C' is a $m \times n$-matrix over F. Applying elementary row transformations of type I to the last n rows we can transform the matrix A' into a matrix of the form

$$A'' := \begin{pmatrix} D_m(b) & C' \\ 0 & D_n(b') \end{pmatrix}$$

with $b' = \det B'$. Finally by adding suitable multiples of the last n rows to the first m rows we can transform A'' to a matrix of the form

$$A''' := \begin{pmatrix} D_m(b) & 0 \\ 0 & D_n(b') \end{pmatrix}.$$

Now A''' is a diagonal matrix which is obtained from the identity matrix I by multiplying the n-th and the last column by the elements b and b'. Therefore $\det A''' = bb' \det I = \det B \det B'$. Thus we have seen that we can transform the matrix A to a matrix A''' by using elementary row transformations of type I only. Thus $\det A = \det A'''$ and it follows that (188) is also satisfied in the case that A is a regular matrix.

---

[2]Recall that a square matrix is called regular if it is not singular which is by definition equivalent with being invertible, see Definition 4.3.

Using induction we obtain then the following general result:

Theorem 4.20. *Let $B_1, \ldots, B_r$ be arbitrary square matrices over the same field F. Then we have the equality*

$$\begin{vmatrix} B_1 & & & * \\ & B_2 & & \\ & & \ddots & \\ 0 & & & B_r \end{vmatrix} = \prod_{i=1}^{r} |B_i| \tag{189}$$

□

Note that the symbol $\prod$ is the comon way to define an arbitrary product in the same way as $\sum$ is used denote a sum. That is the right hand side of (189) is just the compact form to denote the product $\det B_1 \cdots \det B_r$.

Example. In particula the result of the previous theorem is true in case the B_i are all 1×1-matrixes. In this case we get the following simple rule: assume that A is a *upper triangular matrix*, that is

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ & a_{22} & \ldots & a_{2n} \\ & & \ddots & \vdots \\ 0 & & & a_{nn} \end{pmatrix}.$$

Then $\det A = a_{11} \cdots a_{nn}$ is the product of the elements on the diagonal.

Let us conclude this section with a note about one possible use of the determinant function. Recall that in Section 10 of the previous chapter we introduced the concept of similarity of square matrices: two matrices $A, A' \in M_n(F)$ are called similar if there exists a regular $n \times n$-matrix S over F such that

$$A' = S^{-1}AS.$$

Thus in this case we have

$$\begin{aligned} \det A' &= \det S^{-1}AS \\ &= \det AS^{-1}S \\ &= \det A. \end{aligned}$$

where the second equality is due to the corollary to Proposition 4.12 and the last equality is due to $S^{-1}S = I$ is the identity matrix. Thus similar matrices have always the same determinant. Thus we have a convenient way to verify whether two matrices are *not* (!) similar to each other. For example the real matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

cannot be similar since the first matrix has determinant $+1$ and the latter matrix has determinant -1. So these two matrices cannot represent one and the same endomorphism $f\colon \mathbb{R}^2 \to R^2$. It would be much more difficult to show this without the concept of a determinant.

Note that the converse is not true in general: if two $n \times n$-matrices over the same field have the same determinant they might still be not similar to each other!

4. The Leibniz Formula for Determinants

Recall the first three determinat functions which

$$\det_1(A) = a_{11}$$

$$\det_2(A) = a_{11}a_{22} - a_{12}a_{21}$$

$$\det_3(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

(see page 97). These polynomials share a great amount of symmetries and the recursive formuly (178) hints that even the following determinant functions share similar symmetries. The natural question arises whether there is a compact formula for the determinant for an arbitrary $n \times n$-matrix which exposes this symmetries. And the answer is: yes! In this section – which will be the last one about determinants in general – we will derive this formula which is known as the *Leibniz formula for determinants*.[3] But before we can state this formula we need to take again an excursion into the realms of Algebra. The aim of this excursion is to study the behaviour of $\det A$ under arbitrary permutations of the columns of A.

Assume that M is an arbitrary non-empty set. We shall denote by

$$S_M$$

the set of all *bijective* maps $\sigma\colon M \to M$. Such a bijective map shall be called a *permutation* of M. We make the following observations about this set. If $\sigma, \tau \in S$, then the composite map

$$\sigma \circ \tau\colon M \to M, x \mapsto \sigma(\tau(x))$$

is again a bijective map and thus an element of S_M. Therefore we obtain a multilication on the set S_M. The product $\sigma\tau$ of two elements of $\sigma, \tau \in S_M$ is defined to be the composite map $\sigma \circ \tau$.[4] Forming the composite map is an associative operation. Apparently the identity map on M is contained in S_M and $\sigma \circ \mathrm{id} = \sigma$ and $\mathrm{id} \circ \sigma = \sigma$ for every $\sigma \in S_M$. Thus $\mathrm{id} \in S_M$ is the identity element for this product and is also called the *trivial permutation*. Furthermore the inverse map σ^{-1} is defined for every $\sigma \in S_M$ and is itself a bijective map $\sigma^{-1}\colon M \to M$ and thus an element of S_M. Clearly σ^{-1} is the inverse element of σ with respect to the above defined product on S_M, that is $\sigma \circ \sigma^{-1} = \mathrm{id}$ and $\sigma^{-1} \circ \sigma = \mathrm{id}$. Therefore S_M satisfies with this product all the properties required by a group (see Definition 3.46). This group is called the *group of permutations of* M.

As a warning, note that in general the group S_M is not abelian![5] Therefore we have to be carefull when switching the order of elements in a product of permutations as does mostly leave the result not unchanged.

We are interested in the group of permutations in the following special case:

Definition 4.21 (Symmetric Group of n Elements)**.** Let $n \geq 1$ be a natural number and denote by N the set of the natural numbers $1 \leq i \leq n$, that is $N := \{1, \ldots, n\}$. Then the group of permutations of N is called the *symmetric group of n elements* and is denoted by

$$S_n := S_N.$$

---

[3] Named after the German polymath Gottfried Wilhelm von Leibniz, 1646–1761.

[4] Compare this with the definiton of the multiplication of two endomorphisms of a vector space, that is 3.21. Though $\mathrm{End}_F(V)$ is not a group under this multiplication the construction is of a very similar kind. But then again the restriction of this multiplication to the set of all automorphisms yields a group, namely the $\mathrm{GL}_F(V)$.

[5] One can show that the group S_M is abelian if and only if the set M has not more 3 elements.

We have to make some basic observations about the group S_n. Amongst all the permutations of the set $N = \{1, \ldots, n\}$ there are certain simple perumtations. This leads to the next definition.

Definition 4.22. Let $\tau \in S_n$. If there exists natural numbers $1 \leq i, j \leq n$, $i \neq j$ such that

$$\tau(i) = j \text{ and } \tau(j) = i$$

and such that $\tau(k) = k$ for any other integer $1 \leq k \leq n$ then τ is called a *transposition.*

That is, a transposition of the set N switches two numbers of N and leaves all the other elements of N unchanged. We shall in the following denote by $\tau_{i,j}$ the transposition which maps $i \mapsto j$ and $j \mapsto i$. Note that there exist no transposition in S_1 as we need at least two different elements in N.

The next result about transpositions is apparent:

Lemma 4.23. *For any transposition $\tau \in S_n$ holds $\tau^2 = \mathrm{id}$.* □

We will use this result without further reference. Another very intuitive result is the following. It basically states that any finite collection objects can be ordered by exchanging switching finitely many times suitable objects (you might want to try this with a deck of cards). Precisely formulatet this is the following

Lemma 4.24. *Let $\sigma \in S_n$ be a non-trivial permutation. Then σ can be written as a finite product of transpositions, that is there exists a decomposition*

$$\sigma = \tau_1 \cdots \tau_k$$

where the τ_i $(1 \leq i \leq k)$ are all transpositions of the set N.

Proof. We proof the result by induction with respect to $n \geq 2$.

"$n = 2$": There exists only one non-trivial permutation $\sigma \in S_n$ and this permutation is equal to $\tau_{1,2}$. Thus the claim of the lemma is apparently satisfied.

"$n - 1 \Rightarrow n$": Thus we assume that $n > 2$ and that the lemma is prooven for all $n' \leq n - 1$. Denote by N' the set $\{1, \ldots, n - 1\}$. Let $\sigma \in S_n$. Then either $\sigma(n) = n$ or $\sigma(n) \neq n$.

Consider the first case, that is $\sigma(n) = n$. Then the restriction of σ to the set N' is an element of S_{n-1} and by induction follows that

$$\sigma = \tau_1 \cdots \tau_k$$

is a product of transpositions of N'. But then the same decomposition is also a product of transpositions of N. And this proves the claim of the lemma in the case that $\sigma(n) = n$.

Thus it still remains to consider the case that $\sigma(n) \neq n$. Set $i := \sigma(n)$ and furthermore set $\sigma' := \tau_{i,n}\sigma$. Then σ' is a permutation of N such that $\sigma'(n) = \tau_{i,n}(\sigma(n)) = \tau_{i,n}(i) = n$. If σ' is the trivial permutation then $\sigma = \tau_{i,n}$ and the claim of the lemma is proven. Thus we may assume that σ' is not the trivial permutation and we can apply the considerations of the first case. We get a decomposition $\sigma' = \tau_1 \cdots \tau_k$ of σ' into transpositions. But then due to $\tau_{i,n}\tau_{i,n} = \mathrm{id}$ we get that $\sigma = \tau_{i,n}\tau_{i,n}\sigma = \tau_{i,n}\sigma' = \tau_{i,n}\tau_1 \cdots \tau_k$ is a decomposition of σ into transpositions of N. And this completes the remaining case of the induction step. □

In the words of an Algebraist the above lemma states that the group S_n $(n \geq 2)$ is *generated* by the transpositions of the set N.

Note that the above lemma does not state that the decomposition of a permutation into transpositions is unique. The lemma just states that always such a

decomposition exists. But even though the decomposition of a permutation σ is not unique there is a property which is uniquely defined: it will turn out – and this is not a trivial result – is that for a given permutation $\sigma \in S_n$ the number of transpositions needed for its decomposition into transpositions is either always even or always odd. We will see this soon. But before we are heading towards this result we shall still state the following not to difficult observation.

Lemma 4.25. *The symmetric group of n elements S_n has $n! = 1 \cdot 2 \cdots (n-1) \cdot n$ elements.*

PROOF. Left as an exercise to the reader. □

We begin to return from our excursion to the real of Algebra back to Linear Algebra . Let F be a field and $n \geq 1$ a fixed integer $n \geq 1$. For any $\sigma \in S_n$ we know that there exists a unique linear map $P\colon F^n \to F^n$ defined by the equations

$$P_\sigma e_j = e_{\sigma(j)}, \qquad 1 \leq j \leq n, \tag{190}$$

where $(e_1, \ldots, e_n)$ denotes canonical standard basis of F^n (Proposition 3.5). Then the columns of the matrix P_σ are precisely the vectors

$$P_\sigma = (e_{\sigma(1)}, \ldots, e_{\sigma(n)}), \tag{191}$$

since the columns of a matrix are precisely the images of the vectors of the canonical standard basis (Proposition 3.28). The matrix P_σ is therefore obtained from the identity matrix by permutation of the columns.

Definition 4.26. Let $\sigma \in S_n$. Then the matrix P_σ as defined above is called the *permutation matrix* belonging to the permutation σ.

In particular the special matrix R_{ij} as introduced in (162) in the previous chapter is then nothing else then the permutation matrix belonging to the transposition τ_{ij}.

Straight from the definition (190) follows that we have the equality

$$P_\sigma P_\tau = P_{\sigma\tau} \tag{192}$$

for every $\sigma, \tau \in S_n$. Therefore we get that the set of all permuation matrices forms a group which we shall for the moment denote by $P_n(F)$. It is a subgroup of the genral linear group $\mathrm{GL}_n(F)$. It follows from (192) that

$$P\colon S_n \to P_n(F), \sigma \mapsto P_\sigma$$

is a homomorphism of groups and it is apparent that this homomorphism is actually an isomorphism of groups. Thus the symmetric group S_n is isomorphic to $P_n(F)$ as groups.

Now let $A \in M_n(F)$ be an arbitrary $n \times n$-matrix. Denote by $v_1, \ldots, v_n$ the system of column vectors of A. Furthermore denote for any $\sigma \in S_n$ by A_σ the matrix for which the system of column vectors is precisely $v_{\sigma(1)}, \ldots, v_{\sigma(n)}$, that is we define

$$A_\sigma := (v_{\sigma(1)}, \ldots, v_{\sigma(n)}).$$

Note that in particular $I_\sigma = P_\sigma$ due to (191). Due to (190) we have then

$$A_\sigma = AP_\sigma. \tag{193}$$

Thus we can interprete the permutation of the columns of the matrix A by a multiplication of A from the *right* with a permutation matrix. From (193) follows then

$$\det A_\sigma = \det A \det P_\sigma. \tag{194}$$

Thus in order to determine $\det A_\sigma$ we must calculate $\det P_\sigma$. In the particular case of a transposition τ we know from (174) that

$$\det P_\tau = -1.$$

For an arbitrary non-trivial permutation $\sigma \in S_n$ we know from Lemma 4.24 that there exists a decomposition of σ into transpositions t_i and therefore it follows that we get

$$\det P_\sigma = \det P_{\tau_1} \cdots \det P_{\tau_k} = (-1)^k.$$

Since the left hand side of this equation depends only on P_σ (and therefore σ) but since it is independent of the decomposition it follows that the number k is independently of the decomposition either even or odd.[6]

Definition 4.27. Let F be a field with $\operatorname{char} F \neq 2$. Then the function

$$\operatorname{sgn}\colon S_n \to \{+1, -1\}, \sigma \mapsto \operatorname{sgn}(\sigma) := \det P_\sigma \tag{195}$$

which assigns each element of the symmetric group of n elements S_n either the number 1 or the number -1 is called the *sign* function. If $\operatorname{sgn}(\sigma) = +1$ then we say that σ is an *even* permutation and if $\operatorname{sgn}(\sigma) = -1$ we say that σ is an *odd* permutation.

Note that whether a permutation $\sigma \in S_n$ is even or odd is independent of the choice of the field F. It is purely a property of the permutation σ. Moreover nothing hinders us to use the definition of a sign function also in the case that $\operatorname{char} F = 2$. But since the symbols $+1$ and -1 denote in this case the same element of F we have that this function does not carry much (actually any) information when $\operatorname{char} F = 2$. It is then just the constant 1 function.

Note that with this notation we can now write the equation (194) as

$$\det A_\sigma = \operatorname{sgn}\sigma \det A.$$

We return now to the actual aim of this section, namely to derive a closed expression for the determinant of a $n \times n$-matrix

$$A = (a_{ij})$$

over a field F. Let us denote by $v_1, \ldots, v_n$ the system of columns of A. Then we have

$$\begin{aligned} \det A &= \det(v_1, \ldots, v_n) \\ &= \det\Big(\sum_{i=1}^n a_{i1}e_i, \ldots, \sum_{i=1}^n a_{in}e_i\Big) \end{aligned}$$

and using the multilinearity of the determinant (Proposition 4.6) we get

$$= \sum_{(i_1,\ldots,i_n)\in N^n} a_{i_1,1} \cdots a_{i_n,n} \det(e_{i_1}, \ldots, e_{e_n}). \tag{196}$$

Here the last sum is taken over all n-tuples $(i_1, \ldots, i_n)$ of elements in N. If in such an n-tuple two numbers coincide then the rank of the system $e_{i_1}, \ldots, e_{e_n}$ is strictly less than n and therefore $\det(e_{i_1}, \ldots, e_{e_n}) = 0$. Thus we actually only taket the sum in (196) only over all n-tuples $(i_1, \ldots, i_n)$ for which hold

$$\{i_1, \ldots, i_n\} = \{1, \ldots, n\}.^7 \tag{197}$$

[6] Note that we silently ignored the case that $\operatorname{char} F = 2$. But this does not influence the following discussion.

[7] Note that this is an equality of *sets*!

Now if $(i_1, \ldots, i_n)$ is such an n-tuple, then $\sigma(k) := i_k$ defines an bijective map $\sigma\colon N \to N$ and is therefore an element of S_n. Vice versa, if $\sigma \in S_n$, then $(\sigma(1), \ldots, \sigma(n))$ is apparently an n-tuple which satisfies (197). Thus we get from the equality (196) the formula

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n}. \tag{198}$$

Theorem 4.28 (The Leibniz Formula for Determinants)**.** *Let $A = (a_{ij})$ be an arbitrary $n \times n$-matrix over the field F. Then*

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}. \tag{199}$$

PROOF. The formula (199) follows from (198) due to $\det A = \det {}^t A$. □

Notice the impressive beauty and simplicity of the Leibniz formula (199). Yet it is not important for the explicite computation of a determinant. To compute a determinant it the properties and results from Section 1 and Section 3 of this chapter are the useful ones.

The importance of the Leibniz formula lies in the theoretical insight it provides. It gives an explicite relationship between the coefficients of the matrix A and the value of its determinant $\det A$:

For example, if $F = \mathbb{R}$ or $F = \mathbb{C}$, then the determinant turns out to be a continuous function in the coefficients of the matrix A, because after all it is a polynomial and polynomials are continuous. Or in case that A is a matrix with only integer coefficients, then the determinant must be an interger aswell.

Appendix

APPENDIX A

Some Terminology about Sets and Maps

1. Sets

We shall collect in this section a few mathematical definitions about sets. We will restrict ourself to what is know as *naive set theory* (and will not touch the more complicated issues about axiomatic set theory).

Definition A.1. A *set* A is a well defined collection of objects. The objects of a set are called *elements* of the set.

If x is an element of the set A, then we denote this fact in symbols by $x \in A$. If x is not an element of the set A, then we denote this fact by $x \notin A$. If A and B are sets, then the sets are *equal* if they contain precisely the same elements. The set which does not contain any elements is called the *empty set* and denoted in symbols by $\emptyset$.

Note that the above definition means that a set A is *charactericed* by its elements. In order to show that two sets A and B are equal we have to show *always* (!) two seperate things:

(1) For every $x \in A$ holds $x \in B$.
(2) For every $x \in B$ holds $x \in A$.

Note that the above statements are really two different statements. If the first is true, then we say that A is a *subset* of B and vice versa, if the second statement is true, then we say that B is a subset of A. That is:

Definition A.2. Let A and B be sets. Then we say A is a *subset* of B if for every $x \in A$ holds $x \in B$. We denote this fact in symbols by $A \subset B$.

Thus $A = B$ if and onyl if $A \subset B$ and $B \subset A$. Note that the empty set is subset of every set.

We may describe sets by listing there elements in curly bracets. That is, the empty set is given by

$$\{\}$$

and the set containing all integers from 1 to 5 is given by

$$\{1, 2, 3, 4, 5\}.$$

We might also list the elements of a set by its property. For example we may write the set $\mathbb{N}$ of all natural numbers is

$$\mathbb{N} := \{i \in \mathbb{Z} : i \geq 0\},$$

which reads $\mathbb{N}$ is by definition the set of all integers i for which hold $i \geq 0$. This is just a few examples, also other, similar notations are possible.

We can construct new sets out of given sets. The most common constructs are the following:

Definition A.3. Let A and B be sets. Then the their *union* $A \cup B$ is the set

$$A \cup B := \{x : x \in A \text{ or } x \in B\}$$

and their *intersection* $A \cap B$ is the set

$$A \cap B := \{x : x \in A \text{ and } x \in B\}.$$

The *(set theoretic) difference* $A \setminus B$ of A and B is the set

$$A \setminus B := \{x : x \in A \text{ and } x \notin B\}$$

Note that always $A \cup B = B \cup A$ and $A \cap B = B \cap A$ but in general it is *not* (!) true that $A \setminus B = B \setminus A$.

Another common way to construct new sets out of given ones is to form the cartesian product.

Definition A.4. Let A and B sets. Then the *cartesian product* $A \times B$ is the set of all pairs (a, b) with $a \in A$ and $b \in B$, that is

$$A \times B := \{(a, b) : a \in A \text{ and } b \in B\}.$$

Note that $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$. Note that in general $A \times B \neq B \times A$. Furthermore it is always true that $A \times \emptyset = \emptyset$ and $\emptyset \times A = \emptyset$. Cartesian products with more factors are defined in a smimilar way.

If $n \geq 1$ is a natural number then the n-fold cartesian product of a set A is usual denoted by A^n, that is

$$A^n := \underbrace{A \times \ldots \times A}_{n\text{-times}}.$$

Elements in A^n are called n-tuples. We set $A^0 := \emptyset$. Apparently $A^1 = A$ and $A^2 = A \times A$. Note that this explains the notation F^n in Chapter 1 and Chapter 2.

2. Maps

We shall collect a few basic mathematical definitions about maps in general, not specific to Linear Algebra.

Definition A.5. Let A and B two sets. Then a *map* (or *function*) f from A to the set B,

$$f\colon A \to B, x \mapsto f(x),$$

is a rule which assigns each element $x \in A$ precisely one element $f(x) \in B$. The set A is called the *domain* of f and the set B is called the *range* (or *co-domain*) of f.

Two maps $f\colon A \to B$ and $g\colon C \to D$ are equal if $A = C$, $B = D$ and for every $x \in A$ holds $f(x) = g(x)$.[1]

If $x \in A$ then the element $f(x) \in B$ is the *image of x under f*. If $y \in B$ then the *pre-image of y under f* is the set $f^{-1}(y) := \{x \in A : f(x) = y\}$ which is a subset (and not an element!) of A. Similarly if $A' \subset A$, then the *image of A under f* is the set

$$f(A) := \{f(x) : x \in A\}$$

[1] One might relax the definition of "equal maps" by leaving away the requirement $B = D$, depending on the situation.

and this is a subset of the set B. On the other hand, if $B' \subset B$, then by the *pre-image* of B' under f we mean the set

$$f^{-1}(B') := \{x \in A : f(x) \in B'\}$$

and this is a subset of the set A

We may classify maps by the number of elements in the pre-images of elements in the range:

Definition A.6. Let $f\colon A \to B$ be a map. Then we say that f is an *injective* map[2] if the following statement is true:

$$\text{The pre-image } f^{-1}(y) \text{ contains } \textit{at most } 1 \text{ element for every } y \in B. \tag{200}$$

Likewise we say that f is a *surjective* map[3] if the following statement is true:

$$\text{The pre-image } f^{-1}(y) \text{ contains } \textit{at least } 1 \text{ element for every } y \in B. \tag{201}$$

A *bijective* map is a map which is both injective and surjective.

Note that the statement (200) is equivalent with

$$\text{For every } x, y \in A \text{ follows from } f(x) = f(y) \text{ that } x = y.$$

And similarly the statement (201) is equivalent with

$$f(A) = B.$$

Note further that every injective map $f\colon A \to B$ defines always bijective map $A \to f(A)$.

We can construct new maps from given ones:

Definition A.7. Let $f\colon A \to B$ be a map and $A' \subset A$. Then the *restriction* $f|A'$ *of* f *to the set* A' is the map

$$f|A'\colon A' \to B, x \mapsto f(x).$$

Furthermore, if C is another set and $g\colon B \to C$ is another map, then by the *composite map* $g \circ f$ of f and g we mean the map

$$g \circ f\colon A \to C, x \mapsto g(f(x)).$$

Note that the operation of forming composite maps is associative. That is, if $h\colon C \to D$ is another map, then

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

For every set A there exists the map

$$\mathrm{id}_A : A \to A, x \mapsto x$$

which is called the *identity map* of the set A. If there is no danger of confusion then the identity map of the set A is denoted just by id. If $A' \subset A$, then the map

$$i\colon A' \to A, x \mapsto x$$

is called the *inclusion of* A' *into* A. This map is the restriction of the identity map of A to the subset A' and this map is always injective.

If $f\colon A \to B$ is a bijective map, then for every $y \in B$ the pre-image $f^{-1}(y)$ contains precisely one element. We can define a map

$$f^{-1}\colon B \to A$$

---

[2] An injective map is also know as an *one-to-one* map in the literature.

[3] A surjective map is also know as a map *onto* in the literature.

which maps every $y \in B$ to precisely this one element $x \in A$ for which holds $f(x) = y$. This map is called the *inverse map* of the bijective map f. Note that the inverse map is only defined for bijective maps!

Apparently if f is a bijective map, then its inverse map f^{-1} is again a bijective map. The inverse map $(f^{-1})^{-1}$ of the inverse map f^{-1} is again equal to f. If $f\colon A \to B$ and $g\colon B \to C$ are bijective maps, then $g \circ f$ is a bijective map, too, and we have the equality

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}.$$

It is useful to know that a map $f\colon A \to B$ is a bijectvie map if and only if there exist a map $g\colon B \to A$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$. In this case $g = f^{-1}$.

APPENDIX B

Fields with Positive Characteristic

In Chapter 2, when we defined the algebraic structure of a field, we have only given exapmles of fields with characteristic 0. In this appendix we shall give some concrete examples of fields with positive characteristic. We will construct fields with only finite many elements.

Define for a given number $m > 0$ the set $\mathbb{F}_m$ to be

$$\mathbb{F}_m := \{0, 1, 2, \ldots, m-1\}.$$

That is, the set $\mathbb{F}_m$ consists of exactly m elements. We shall define an addition and a multiplication on $\mathbb{F}_m$ which makes the this set into a ring and we will see under which conditions this ring is also a field.

We need a result basic result from number theory: the *division algorithm for integers*. This result states that if $m > 0$ is a given positive integer and x an arbitrary integer, then there exists *unique* integers q and r such that

$$x = qm + r \qquad \text{and} \qquad 0 \leq r < m. \tag{202}$$

We say that r is the *reminder* of the division of x by m and we denote this reminder in the following by

$$r_m(x).$$

Now we define the addition and the multiplication in $\mathbb{F}_m$ as follows:

$$x + y := r_m(x + y) \qquad \text{and} \qquad x \cdot y := r_m(xy) \tag{203}$$

for every $x, y \in \mathbb{F}_m$. The so defined addition and multiplication is called the addition and multiplication *modulo* m. We shall state without a proof the following result.

Proposition B.1. *Let $m > 0$ be a positive integer. Then:*

(1) *The set $\mathbb{F}_m$ together with the addition and multiplication modulo m as defined in* (203) *is a commutative ring with unit.*
(2) *The ring $\mathbb{F}_m$ is a field if and only if m is a prime number.* □

Apparently we have for any prime number p the equality $\mathrm{char}(\mathbb{F}_p) = p$. On the other hand one has the following result which gives a strong constraint on the characteristic of a field.

Proposition B.2. *Let F be a field with* $\mathrm{char}(F) > 0$. *Then* $\mathrm{char}(F) = p$ *is a prime number.*

Proof. We only need to show that if $\mathrm{char}(F) = k$ is not a prime, then F cannot be a field. Assume therefore that $k = mn$ for some $1 < m, n < k$. Then $x := me \neq 0$ and $y := ne \neq 0$. On the other hand

$$xy = (me)(ne) = (mn)e = ke = 0$$

but this contradicts to (39). Thus F is not a field if k is not a prime. Thus if F is a field with $\mathrm{char}(F) > 0$ then $\mathrm{char}(F)$ is necessarily a prime number. □

APPENDIX C

Zorn's Lemma and the Existence of a Basis

Recall that we have proven the existence of a basis for a vector space V only in the case that V had been a finitely generated vector space. In this case the proof was even relatively easy. If one wants to prove the existence of a basis in the general case where V is not necessarily finitely generated one needs much heavier machinery to conquer the problem.

The key to prove the existence of a basis for vector spaces which are not finitely generated is *Zorn's Lemma*[1], which is a result of set theory and equivalent to the Axiom of Choice.

Before we can state Zorn's Lemma we need some definitions.

Definition C.1. Let X be a set. A relation "$\leq$" is said to be a *partial order* on X if the following three conditions hold:

(1) The relation "$\leq$" is *reflexive*, that is $x \leq x$ for every $x \in X$.
(2) The relation "$\leq$" is *antisymmetric*, that is for every $x, y \in X$ it follows from $x \leq y$ and $y \leq x$ that $x = y$.
(3) The relation "$\leq$" is *transitive*, that is for every $x, y, z \in X$ it follows from $x \leq y$ and $y \leq z$ that $x \leq z$.

A partial order on X is called a *total order* if $x \leq y$ or $y \leq x$ for every $x, y \in X$, that is any two elements of X can be compared in either or the other way.

Let $A \subset X$ be a subset of a partialy ordered set. Then $m \in X$ is an *upper bound* of A if $x \leq m$ for every $x \in A$. An element $m \in X$ is called a *maximal element* if $m \leq x$ implies $m = x$ for every $x \in X$.

Example. Let X be an arbitrary set. Then the subset relation "$\subset$" defines a partial order on the set of subsets of X.

Zorn's Lemma. *Every non-empty partially ordered set in which every chain (that is totally ordered subset) has an upper bound contains at least one maximal element.* □

Let V be an arbitrary vector space over a field F. Denote by $\mathscr{L}$ the set of all linear independent subsets of V. Then $\mathscr{L}$ is a partial ordered set under the inclusion relation "$\subset$". We want to show that $\mathscr{L}$ satisifies the conditions required by Zorn's Lemma.

First of all $\mathscr{L}$ is not empty since the empty set $\emptyset$ is a linear independent subset of V (see the examples to Definition 2.16) and thus $\emptyset \in \mathscr{L}$.

Let $\mathscr{C} \subset \mathscr{L}$ be a subset which is totaly ordered by the inclusion relation "$\subset$". We claim that its union

$$C := \bigcup_{M \in \mathscr{C}} M$$

[1]Named after the American algebraist, group theorist and numelrical analyst Max August Zorn, 1906–1993.

is linear independent subset. Therefore choose pairwise distinct vectors $v_1, \ldots, v_n \in C$ and assume that $a_1, \ldots, a_n \in F$ are numbers such that

$$a_1 v_1 + \ldots + a_n v_n = 0 \qquad (*)$$

is a linear combination of the zero vector. Now there exists $M_1, \ldots, M_n \in \mathscr{C}$ such that $v_i \in M_i$ for $i = 1, \ldots, n$. Since $\mathscr{C}$ is by assumption totaly ordered by incluseion there exists a $1 \leq i_0 \leq n$ such that $M_i \subset M_{i_0}$ for every $i = 1, \ldots, n$. In particular this means that $(*)$ is a linear combination of the zero vector by vectors of the is linearly independent set M_{i_0}. Thus $a_1 = \ldots = a_n = 0$. Now since $v_1, \ldots, v_n$ had been arbitrary, but pairwise distinct vectors in C this shows that C is a linearly independent subset of V and therefore an element of $\mathscr{L}$ and by construction for every $M \in \mathscr{C}$ holds $M \subset C$. Thus $\mathscr{C}$ has an upper bound.

Since $\mathscr{C}$ was an arbitrary chain of $\mathscr{L}$ this shows that every chain in $\mathscr{L}$ has an upper bound. Thus the can apply Zorn's Lemma which states that there exists a maximal element $B \in \mathscr{L}$.

Then B is a maximal linear independent subset of V and thus by Proposition 2.22 it follows that B is a basis of V. This proves Theorem 2.23 of Chapter 2:

Theorem (Existence of a Basis)**.** *Every vector space has a basis.* □

The interesting issue about this theorem is that it is so strong that it can be proven to be equivalent with Zorn's Lemma (see [**Bla84**]). If we *require* that every vector space has a basis, then it follows that the Axiom of Choice must hold and thus Zorn's Lemma holds, too. In this sense the existence of a basis is a very deep and strong result of Linear Algebra.

Note the beauty of the above theorem! It takes only as little as six words we can state a theorem which relates to the very foundation of mathematics.

Observe that with a slightly addapted proof we can verify that the basis extension theorem – that is Theorem 2.36 – holds true in the infinite dimensional case, too. Precisely we have the following result.

Theorem (Basis Extension Theorem)**.** *Let N be a linear independent subset of a vector space V. Then N can be extendet to a basis of V, that is there exists a basis M of V such that $N \subset M$.* □

APPENDIX D

A Summary of Some Algebraic Structures.

Semigroup: A *semigroup* $H = (H, *)$ is a non-empty set H together with a binary operation

$$* : H \times H \to H, (x, y) \mapsto x * y$$

which satisfies the associativity law. Note that a semigroup is not required to hava an idenitity element. A semigroup H is said to be *commutative* if $x * y = y * x$ for every $x, y \in H$.

Monoid: A *monoid* $H = (H, *)$ is a semigroup which has a identity element. An element $e \in H$ is said to be an identity element if $x*e = x$ and $e*x = x$ for every $x \in H$. The identity element of a monoid is always unique. A monoid is called *commutative* if it is a commutative semigroup.

Group: A *group* $G = (G, *)$ is a monoid where every element of G has an inverse element. If $x \in G$, then $y \in G$ is an inverse element of x if $x*y = e$ and $y * x = e$ where e is the identity element of G. An inverse element is always unique, therefore one can speak of "the" inverse elment of an element $x \in G$. A group is called *abelian* if it is a commutative monoid. (See page 80.)

Ring: A *ring* (with unit) $R = (R, +, \cdot)$ is a set R together with two binary operations

$$+ : R \times R \to R, (x, y) \mapsto x + y, \qquad \text{(addition)}$$
$$\cdot : R \times R \to R, (x, y) \mapsto xy \qquad \text{(multiplication)}$$

satisfying the following three conditions:
(1) $(R, +)$ is an abelian group.
(2) $(R, \cdot)$ is an monoid.
(3) $x(y + z) = xy + xz$ for every $x, y, z \in R$ (distributive law).
The identity of the addition is usually denoted by "0" and the identity of the multiplication is usually denoted by "1". In order to avoid triviality one usually requires that $0 \neq 1$, that is the identity element of the addition and the identity element of the multiplication are different elements. A ring where the multiplication is commutative is called a *commutative* ring. A *zero divisor* of R is an element $0 \neq x \in R$ such that $xy = 0$ for some $0 \neq y \in R$. A ring which does not have zero divisors is called a *regular* ring. (See page 21.)

Field: A *field* $F = (F, +, \cdot)$ is a commutative ring such that $(F^\bullet, \cdot)$ is a group (where $F^\bullet := F \setminus \{0\}$). That is, in a field the multiplication is commutative and every non-zero element has a multiplicative inverse. (See page 19.)

Vector Space: A *vector space* $V = (V, +, \cdot)$ *over a field* F is a set F with two maps

$$+\colon V \times V \to V, (x, y) \mapsto x + y, \qquad \text{(addition)}$$
$$\cdot\colon F \times V \to V, (a, x) \mapsto ax \qquad \text{(scalar multiplication)}$$

satisfying the following conditions:

(1) $(F, +)$ is an abelian group.
(2) $(ab)x = a(bx)$ for every $a, b \in F$ and $x \in V$.
(3) $1x = x$ for every $x \in V$.
(4) $a(x + y) = ax + ay$ for every $a \in F$ and $x, y \in V$.
(5) $(a + b)x = ax + bx$ for every $a, b \in F$ and $x \in V$.

Elements of V are called *vectors* and elements of F are called *scalars.* (See page 21.)

Algebra: A *algebra (with unit)* V *over a field* F (or F-algebra) is a vector space over the field F with a multiplication

$$\cdot\colon V \times V \to V, (x, y) \mapsto xy$$

satisfying the following conditions:

(1) V together with the addition of vectors and the above defined multiplication on V is a ring.
(2) For every $x, y \in V$ and $a \in F$ holds

$$a(xy) = (ax)y = x(ay).$$

Note that we use by abuse of notation the very same symbol for the scalar multiplication and the multiplication on V! (See page 61.)

Examples.

(1) The set of *natural numbers*

$$\mathbb{N} := \{0, 1, 2, 3, \ldots\}$$

is a commutative monoid under addition and under multiplication.

(2) The set of strictly positive natural numbers

$$\mathbb{N}^+ := \{1, 2, 3, \ldots\}$$

is a commutative semigroup under addition and a commutative monoid under multiplication.

(3) The set of *integers*

$$\mathbb{Z} := \{0, \pm 1, \pm 2, \pm 3, \ldots\}$$

is a abelian group under addition and a commutative monoid under multiplication. In particular $(\mathbb{Z}, +, \cdot)$ is a commutative, regular ring.

(4) The set of all *rational numbers*

$$\mathbb{Q} := \{\tfrac{r}{s} : r \in \mathbb{Z} \text{ and } s \in \mathbb{N}^+\}$$

is a field under the addition and multiplication.

(5) The set of all *real numbers* $\mathbb{R}$ is a field under the additon and multiplication.

(6) The set of all *complex numbers*

$$\mathbb{C} := \{a + ib : a, b \in \mathbb{R}\}$$

is a field under the multiplication of complex numbers (note that for the *imaginary unit* i holds the relation $i^2 = -1$).

(7) If V is a F-vector space, then the *endomorphism ring*

$$\mathrm{End}_F(V) := \{f : f\colon V \to V \text{ is a linear map}\}$$

is a F-algebra in a natural way (see page 61).

(8) In particular the *algebra of the $n \times n$-matrices over a field F*

$$M_n(F) := F^{n,n}$$

is a F-algebra (see page 71).

(9) The *general linear group*

$$\mathrm{GL}(V) := \{f \in \mathrm{End}_F(V) : f \text{ is an isomorphism}\}$$

of the F vector space V is a group (see page 80).

(10) In particular the set of all invertible $n \times n$-matrices over F

$$\mathrm{GL}_n(F) := \mathrm{GL}_F(F^n)$$

is a group, called the *general linear group of degree n.*

APPENDIX E

About the Concept of a Rank

In Section 5 we defined the important concept of a rank. The rank of a system of vectors is a typical example of how a simple but well crafted definition can result into a very fruitful concept in mathematics. This appendix shall recollect the different variants of the rank and point out some aspects of the concept.

Recall Definition 2.25 were we defined what we mean by the rank of the finite system of vectors $u_1, \ldots, u_m$ is r, namely:

(1) There exist a linear independent subset of the set $\{u_1, \ldots, u_m\}$ which consists of exactly r vectors.
(2) Any subset of $\{u_1, \ldots, u_m\}$ which consists of $r+1$ vectors is linear dependent.

We obtained straight away some simple results about the rank: the rank of a system of m vectors is always bounded by the number m, that is

$$\operatorname{rank}(u_1, \ldots, u_m) \leq m,$$

and equality holds if and only if $u_1, \ldots, u_m$ are m pairwise distinct vectors which form a linar independent set. Another simple result has been that the rank of a system of vectors is not decreasing if new vectors are added, that is

$$\operatorname{rank}(u_1, \ldots, u_{m+1}) \geq \operatorname{rank}(u_1, \ldots, u_m),$$

and that the rank does not increase if u_{m+1} is a linear combination of the vectors $u_1, \ldots, u_m$ (Proposition 2.27).

The rank of a system of vectors is invariant under elementary transformations (Definition 2.29 and Proposition 2.30). And Theorem 2.31 described an algorithm how to compute the rank of a finite system of vectors using the Gauss Algorithm.

The invariance of the dimension (Theorem 2.34) is the first non-trivial result which we was obtained using the concept of a rank. The key to the proof of this result is Proposition 2.38 which states that the rank of a finite system of vectors is bounded by the dimension of the vector space, that is

$$\operatorname{rank}(u_1, \ldots, u_m) \leq \dim V.$$

The so obtained theorem about the invariance of the dimension enabled us to define the dimension of a vector space in a proper way. Proposition 2.38 gives intuitive interpretation of the rank of a system of vectors, namely

$$\operatorname{rank}(u_1, \ldots, u_m) = \dim \operatorname{span}(u_1, \ldots, u_m).$$

Since the columns and rows of an $m \times n$-matrix over a field F can be seen as elements of the vector space F^m and F^n the concept of a rank extends in a natural way to matrices. We defined – see Definition 2.44 – the column rank of a $m \times n$-matrix A to be the rank of the n vectors $u_1, \ldots, u_n$ obtained from the colums of the matrix A, that is

$$\operatorname{rank}_{\mathrm{c}}(A) := \operatorname{rank}(u_1, \ldots, u_n).$$

Similarly we define the row rank of A to be the rang of the m vectors $v_1, \ldots, v_m$ obtained from the rows of the matrix A, that is

$$\operatorname{rank_r}(A) := \operatorname{rank}(v_1, \ldots, v_m).$$

It turns out – see Theorem 2.49 – that the row and the column rank are equal for every matrix A, that is

$$\operatorname{rank_r} A = \operatorname{rank_c} A.$$

Thus we can define the rank of a matrix A to be either of those two numbers, that is

$$\operatorname{rank} A := \operatorname{rank_r}(A)$$

which is done in Definition 2.50. Using the interpretation given in Proposition 2.38 this means that the rank of a matrix A is equal to the dimension of the vector space spanned by the vectors obtained from the columns of A and that this number is equal to the dimension to the dimension of the vector space spanned by the vectors obtained from the rows of A.

In Chapter 3 the rank of a linar map $f\colon V \to W$ is introduced in Definition 3.14 as the dimension of image space of f, that is

$$\operatorname{rank} f := \dim(\operatorname{im} f).$$

Note that this allows the possibility of $\operatorname{rank} f = \infty$. In the case that V an W are finite dimensional vector spaces Proposition 3.25 states the following relation between the rank of matrices and the rank of a linar map: for any coordinate matrix A of the linear map f we have the equality

$$\operatorname{rank} f = \operatorname{rank} A.$$

Thus the rank of a linear map turns out to be a very natural definition.

From the rank of a linear map $f\colon V \to W$ we can draw some basic conclusions. For example if V is a finite dimensional then f is a monomorphism if and only if $\operatorname{rank} f = \dim V$. Likewise, if W is finite dimensional then f is an epimorphism if and only if $\operatorname{rank} f = \dim W$. And from this follows that f is an isomorphism of finite dimensional vector spaces of same dimension n if and only if $\operatorname{rank} f = n$.

From the dimension formula for linear maps – that is Theorem 3.15 – we get in the case that V is finite dimensional the following interpretation for the rank of a linear map:

$$\operatorname{rank} f = \dim V - \dim(\ker f)$$

and this explains why the dimension of the kernel of a linear map is also called the defect of a linear map. In particular a linear map is a monomorphism if it has defect 0.

Index

Bibliography

[Bla84] Andreas Blass, *Existence of bases implies the axiom of choice*, Contemporary Mathematics (1984), no. 31, 31–33.

[Lan66] Serge Lang, *Linear algebra*, Addison–Wesley, 1966.

[Lan67] ———, *Algebraic structures*, Addison–Wesley, 1967.

[Lor92] Falko Lorenz, *Lineare algebra 1*, 3rd ed., Wissenschaftsverlag, 1992.