Kungl
Tekniska
Högskolan

An Introduction to Algebra and Geometry via Matrix Groups

Lecture Notes — Spring 1995

(First four chapters)

Mats Boij and Dan Laksov

MATRIX GROUPS

MATS BOIJ AND DAN LAKSOV

Contents

Introduction

The purpose of these notes is to introduce, at the undergraduate level, some fundamental notions of algebra and geometry. Throughout we illustrate the theory with examples of matrix groups, and we motivate the introduction of new concepts as tools for distinguishin the most common matrix groups. We are thus, in a natural way, lead to the algebraic structures groups, rings, fields, vector spaces, and Lie algebras, and the geometric structures topological spaces, metric spaces, manifolds, Lie groups, and varieties. The language, methods and spirit of these areas penetrate most parts of mathematics and its applications, and it is of utmost importance that the student encounters these notions at an early stage.

Matrix groups are central in many parts of mathematics and its applications, and the theory of matrix groups is ideal as an introduction to mathematics. On the one hand it is easy to calculate and understand examples, and on the other hand the examples lead to an understanding of the general theoretical framework that incorporates the matrix groups. This is natural historically, as the study of matrix groups has been one of the main forces behind the development of several important branches of mathematics and its applications. It is particularly fascinating how algebra and geometry are intertwined in the study of matrix groups. The unity of algebraic an geometric methods is deeply rooted in mathematics and the matrix groups provide an excellent concrete illustration of this phenomenon.

Throughout we have made an effort to keep the presentation elementary and have included the minimum of the standard material in algebra and geometry that is necessary in the study of matrix groups. As a consequence we have covered less general material than is usually included in more general treatises in algebra and geometry. On the other hand we have included more material of a general nature than is normally presented in specialized expositions of matrix groups. Hopefully we have found an equilibrium that makes the notes enjoyable, and useful, to undergraduate students. There is a vast flora of general textbooks in algebra and geometry that cover the general material of these notes. During the preparation of these notes we have found the books of [1], [2], [3], [5], [6], and [7], of the reference list, useful.

The prerequisites for this course consist of a standard course in linear algebra and calculus. To appreciate the text, mathematical maturity and interest in mathematics is important. We assume that the reader, with a few hints, can fill in details in proofs that are similar to those of the basic courses of linear algebra and calculus. This should cause no difficulties to a student mastering fully the first year courses, and we hope that it is a challenge for the student to rethink earlier courses in a more general setting.

1 Algebraic properties of matrix groups

In this chapter we introduce some important matrix groups, called *classical*, over the complex numbers. We use these groups to motivate the definitions of groups, rings and fields, and to illustrate their properties. It is natural to generalize these matrix groups to general fields. In order to study these *classical groups* over arbitrary fields we discuss the theory of vector spaces over arbitrary fields, and bilinear forms on such vector spaces. We can then define the orthogonal and symplectic group with respect to the bilinear forms. The tools we introduce allow us to determine the generators for the *general linear group*, the *orthogonal group*, the *symplectic group*, and their special subgroups. We then determine the centers of these groups. The center is an invariant that allows us to distinguish several, but not all, of the groups that we have studied.

1.1 Matrix groups over the complex numbers

In this section we define the most common matrix groups with coefficients in the complex numbers. These groups, that are called *classical*, and their properties will serve as a motivation for the material in the remaining part of the book. We assume here that the reader is familiar with the basic notions of linear algebra, like multiplication, determinants, and inverses of matrices.

Let $\mathrm{M}_n(\mathbf{C})$ be the set of all $n \times n$ matrices $A = (a_{ij})$, with complex coordinates a_{ij}. The *identity matrix* which has diagonal coordinates equal to 1 and the remaining coordinates equal to 0 is denoted by I_n and we shall denote the matrix with all coordinates zero by 0, irrespectively of what size it has. Given a matrix $A = (a_{ij})$ in $\mathrm{M}_n(\mathbf{C})$ and a complex number a, we write $aA = (aa_{ij})$ and we denote the determinant of a matrix A in $\mathrm{M}_n(\mathbf{C})$ by $\det A$.

The *transpose* (a_{ji}) of the matrix $A = (a_{ij})$ is denoted by ${}^tA$. We have that

$$\det A = \det {}^tA.$$

Multiplication of two matrices $A = (a_{ij})$ and $B = (b_{ij})$ in $\mathrm{M}_n(\mathbf{C})$ produces a matrix $AB = (\sum_{l=1}^n a_{il}b_{lj})$. The multiplication can be considered as a map

$$\mathrm{M}_n(\mathbf{C}) \times \mathrm{M}_n(\mathbf{C}) \to \mathrm{M}_n(\mathbf{C})$$

from the set of *ordered pairs* (A, B) of matrices in $\mathrm{M}_n(\mathbf{C})$ to $\mathrm{M}_n(\mathbf{C})$, which sends (A, B) to AB. We have the following multiplication rules

$$A = I_nA = AI_n,$$
$$A(BC) = (AB)C,$$
$${}^t(AB) = {}^tB{}^tA,$$

for any three matrices A, B and C of $\mathrm{M}_n(\mathbf{C})$. Moreover, we have that

$$\det AB = \det A \det B.$$

A matrix A in $\mathrm{M}_n(\mathbf{C})$ is called *invertible*, or *non-singular*, if there is a matrix B such that

$$AB = BA = I_n. \tag{1.1.0.1}$$

The matrix B in the expression 1.1.0.1 is uniquely determined. It is called the *inverse* of A and it is denoted by A^{-1}. Since we have that

$$({}^tA)^{-1} = {}^t(A^{-1}),$$

we can write ${}^tA^{-1} = ({}^tA)^{-1} = {}^t(A^{-1})$, without ambiguity.

The subset of $\mathrm{M}_n(\mathbf{C})$ consisting of invertible matrices is denoted by $\mathrm{Gl}_n(\mathbf{C})$, and called the *general linear group*. Given two matrices A and B in $\mathrm{Gl}_n(\mathbf{C})$ we have that the product AB has inverse $B^{-1}A^{-1}$, that is

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Hence the product AB is in $\mathrm{Gl}_n(\mathbf{C})$. Moreover, we have that

$$\det A = \frac{1}{\det A^{-1}}.$$

The subset of $\mathrm{Gl}_n(\mathbf{C})$ consisting of matrices with determinant 1 is denoted by $\mathrm{Sl}_n(\mathbf{C})$ and called the *special linear group*. Given two matrices A and B in $\mathrm{Sl}_n(\mathbf{C})$, it follows from the equation $\det AB = \det A \det B$ that AB is in $\mathrm{Sl}_n(\mathbf{C})$. Moreover, it follows from the equation $\det A^{-1} = (\det A)^{-1}$ that A^{-1} is in $\mathrm{Sl}_n(\mathbf{C})$.

Fix a matrix S in $\mathrm{M}_n(\mathbf{C})$. We shall denote by $\mathrm{G}_S(\mathbf{C})$ the subset of matrices A in $\mathrm{Gl}_n(\mathbf{C})$ that satisfy the relation

$${}^tASA = S.$$

Given two matrices A and B in $\mathrm{G}_S(\mathbf{C})$, we have that AB is in $\mathrm{G}_S(\mathbf{C})$. Indeed, we have that

$${}^t(AB)SAB = {}^tB{}^tASAB = {}^tB({}^tASA)B = {}^tBSB = S.$$

Moreover, we have that A^{-1} is in $\mathrm{G}_S(\mathbf{C})$. Indeed, when multiplying the relation $S = {}^tASA$ to the right by A^{-1} and to the left by ${}^tA^{-1}$, we obtain the equation

$${}^tA^{-1}SA^{-1} = S.$$

When S is in $\mathrm{Gl}_n(\mathbf{C})$ it follows from the equality $\det {}^tA \det S \det A = \det S$ that we have $(\det A)^2 = 1$. Hence $\det A = \pm 1$. In this case we denote the subset of matrices A in $\mathrm{G}_S(\mathbf{C})$ that have determinant equal to 1 by $\mathrm{SG}_S(\mathbf{C})$. As in the case with $\mathrm{Sl}_n(\mathbf{C})$ we have that if A and B are in $\mathrm{SG}_S(\mathbf{C})$, then AB and A^{-1} are both in $\mathrm{SG}_S(\mathbf{C})$.

We have seen that all the sets $\mathrm{Gl}_n(\mathbf{C})$, $\mathrm{Sl}_n(\mathbf{C})$, $\mathrm{G}_S(\mathbf{C})$ and $\mathrm{SG}_S(\mathbf{C})$ share the properties that if A, B and C are elements of the set, then I_n, A^{-1} and AB are also in the set. Clearly we also have that $A = AI_n = I_nA$, $AA^{-1} = A^{-1}A = I_n$ and $A(BC) = (AB)C$, because these relations hold for all elements in $\mathrm{Gl}_n(\mathbf{C})$.

There are two special cases of $\mathrm{G}_S(\mathbf{C})$ and $\mathrm{SG}_S(\mathbf{C})$ that are particularly interesting. The first one is obtained when $S = I_n$. The corresponding groups $\mathrm{G}_S(\mathbf{C})$ and $\mathrm{SG}_S(\mathbf{C})$ are denoted by $\mathrm{O}_n(\mathbf{C})$ and $\mathrm{SO}_n(\mathbf{C})$ and called the *orthogonal group* and *special orthogonal group* respectively. They consist of the elements A in $\mathrm{Gl}_n(\mathbf{C})$ and $\mathrm{Sl}_n(\mathbf{C})$, respectively, such that

$${}^tAA = I_n.$$

To introduce the second special case it is convenient to use the following notation for matrices

Given matrices A, B, C and D of sizes $r\times s$, $r\times(n-s)$, $(n-r)\times s$, and $(n-r)\times(n-s)$, respectively, we denote by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

the $n\times n$ block matrix with A, B, C and D in the upper left, upper right, lower left, and lower right corner, respectively.

Let J_m be the matrix in $\mathrm{M}_m(\mathbf{C})$ with 1 on the *antidiagonal*, that is the coordinates a_{ij} with $i+j=m+1$ are 1, and the remaining coordinates 0. Take

$$S=\begin{pmatrix} 0 & J_m \\ -J_m & 0 \end{pmatrix}. \tag{1.1.0.2}$$

The corresponding set $\mathrm{G}_S(\mathbf{C})$ is denoted by $\mathrm{Sp}_{2m}(\mathbf{C})$ and it is called the *symplectic group*. When we write $\mathrm{Sp}_n(\mathbf{C})$, we always assume that n is even.

Remark 1.1.1. In the following it will be important to view the matrix groups $\mathrm{O}_n(\mathbf{C})$, $\mathrm{SO}_n(\mathbf{C})$ and $\mathrm{Sp}_n(\mathbf{C})$ as *automorphisms of bilinear forms*. We shall return to this point of view in Sections 1.7 and 1.8. Here we shall indicate how it is done.

Let $S=(s_{ij})$ be in $\mathrm{M}_n(\mathbf{C})$. Define a map

$$\langle\, ,\rangle\colon \mathbf{C}^n\times\mathbf{C}^n\to\mathbf{C},$$

by

$$\langle (a_1,\dots,a_n),(b_1,\dots,b_n)\rangle = \begin{pmatrix} a_1 & \dots & a_n\end{pmatrix}\begin{pmatrix} s_{11} & \dots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \dots & s_{nn}\end{pmatrix}\begin{pmatrix} b_1 \\ \vdots \\ b_n\end{pmatrix}. \tag{1.1.1.1}$$

The map $\langle\, ,\rangle$ satisfies the following properties:

Given x, y and z in $\mathbf{C}^n$, and a in $\mathbf{C}$, we have that:

(i) $\langle x+y,z\rangle=\langle x,z\rangle+\langle y,z\rangle$.
(ii) $\langle x,y+z\rangle=\langle x,y\rangle+\langle x,z\rangle$.
(iii) $\langle ax,y\rangle=\langle x,ay\rangle=a\langle x,y\rangle$.

We say that $\langle\, ,\rangle$ is a *bilinear form* on $\mathbf{C}^n$. A matrix A in $\mathrm{Gl}_n(\mathbf{C})$ is an *automorphism* of the form if

$$\langle Ax,Ay\rangle=\langle x,y\rangle,\qquad \text{for all pairs } x,\ y \text{ in } \mathbf{C}^n.$$

We have that $\mathrm{G}_S(\mathbf{C})$ consists of all automorphisms of the bilinear form $\langle\, ,\rangle$ defined by S (see Exercise 1.1.4).

Not all the groups given above are different. We have, for example that $\mathrm{Sp}_2(\mathbf{C})=\mathrm{Sl}_2(\mathbf{C})$ (see Exercise 1.1.7). The main theme of these notes is to investigate in which sense they are different. This is done by imposing algebraic and geometric structures on the groups and by associating to these structures invariants that make it possible to distinguish them. As mentioned above we shall determine the centers of the matrix groups of this section, and in this way distinguis between several of them. First we shall however introduce the concept of a group and generalize the matrix groups of this section.

Exercises

1.1.1. Determine the groups $\mathrm{Gl}_1(\mathbf{C})$, $\mathrm{Sl}_1(\mathbf{C})$, $\mathrm{O}_1(\mathbf{C})$ and $\mathrm{SO}_1(\mathbf{C})$.

1.1.2. Show that the inclusions $\mathrm{Sl}_n(\mathbf{C}) \subseteq \mathrm{Gl}_n(\mathbf{C})$ and $\mathrm{SO}_n(\mathbf{C}) \subseteq \mathrm{O}_n(\mathbf{C})$ are proper.

1.1.3. Define the groups $\mathrm{SSp}_2(\mathbf{C})$ and show that $\mathrm{SSp}_2(\mathbf{C}) = \mathrm{Sp}_2(\mathbf{C})$.

1.1.4. Show that the group $\mathrm{G}_S(\mathbf{C})$ is the group of automorphisms of the form $\langle\, , \rangle$, defined by

$$\langle (a_1, \dots, a_n), (b_1, \dots, b_n) \rangle = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix} \begin{pmatrix} s_{11} & \dots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \dots & s_{nn} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

1.1.5. Let $\langle\, , \rangle : \mathbf{C}^n \times \mathbf{C}^n \to \mathbf{C}^n$ be defined as in 1.1.1.1 with $S = I_n$, and let A be a matrix in $\mathrm{M}_n(\mathbf{C})$. Show that ${}^tA$ is the unique matrix B such that $\langle Ax, y \rangle = \langle x, By \rangle$ for all x and y in $\mathbf{C}^n$.

1.1.6. Determine all elements of $\mathrm{O}_2(\mathbf{C})$ and $\mathrm{SO}_2(\mathbf{C})$.

1.1.7. Show that $\mathrm{Sl}_2(\mathbf{C}) = \mathrm{Sp}_2(\mathbf{C})$.

1.2 Groups

We have in Section 1.1 given examples of sets whose elements can be multiplied and the multiplication in all the sets enjoys similar algebraic properties. In this section we shall formalize the essential properties of the multiplication.

A *multiplication* on a set S is a map

$$S \times S \to S$$

from the *Cartesian product* $S \times S$, that is the set of ordered pairs of elements of S, to S. The image of the pair (a, b) we denote by ab.

Definition 1.2.1. A *group* is a set G together with a multiplication that satisfies the following three properties:

(i) (*Associativity*) For any triple a, b, c of elements of G, we have that

$$a(bc) = (ab)c.$$

(ii) (*Identity*) There is an element e in G such that

$$a = ae = ea$$

for all elements a of G.

(iii) (*Inverse*) For each element a of G there is an element b of G such that

$$e = ab = ba.$$

There is only one element in G having the same property (ii) as e. Indeed, if e' were another such element we have that

$$e' = e'e = e.$$

Similarly, given a, there is only one element b in G having the property of (iii). Indeed, if b' were another such element we have that

$$b' = b'e = b'(ab) = (b'a)b = eb = b.$$

Example 1.2.2. We saw in Section 1.1 that all the sets $\mathrm{Gl}_n(\mathbf{C})$, $\mathrm{O}_n(\mathbf{C})$, $\mathrm{Sp}_n(\mathbf{C})$, $\mathrm{Sl}_n(\mathbf{C})$ and $\mathrm{SO}_n(\mathbf{C})$ are groups.

There are many more natural groups than those in the previous example. Here follows some well-known examples.

Example 1.2.3. The integers $\mathbf{Z}$, the rational numbers $\mathbf{Q}$, the real numbers $\mathbf{R}$ and the complex numbers $\mathbf{C}$ are all groups under addition. In these cases we are used to denote the multiplication by the symbol $+$ and the identity by 0.

Example 1.2.4. The non-zero rational, real and complex numbers, $\mathbf{Q}^*$, $\mathbf{R}^*$, and $\mathbf{C}^*$ are groups under multiplication.

Example 1.2.5. Let S be a set. Denote by $\mathfrak{S}_S$ the set of all injective maps of S onto itself. We define the product $\tau\sigma$ of two maps $\sigma\colon S \to S$ and $\tau\colon S \to S$ as the composite map $\tau\sigma\colon S \to S$. With this multiplication $\mathfrak{S}_S$ is a group. The identity is the map that leaves all elements of S fixed, and the inverse of a map τ is the map that sends $\tau(i)$ to i, which exists because τ is injective and onto. When $S = \{1, \dots, n\}$ we write $\mathfrak{S}_S = \mathfrak{S}_n$ and we call $\mathfrak{S}_n$ the *symmetric group* on n letters. It is a group with $n!$ elements.

Definition 1.2.6. A group G is called *abelian* if $ab = ba$ for all pairs of elements a, b of G, and we say that a and b *commute*.

Remark 1.2.7. In abelian groups we shall often, in accordance with Example 1.2.3, denote the multiplication by $+$ and the identity by 0.

Example 1.2.8. The groups of Examples 1.2.3 and 1.2.4 are abelian, while none of the groups in 1.2.2 and 1.2.5 are abelian, when $n > 2$ (see Exercise 1.2.1).

Definition 1.2.9. A *homomorphism* from a group G to a group H is a map

$$\Phi\colon G \to H$$

such that $\Phi(ab) = \Phi(a)\Phi(b)$, for all a, b in G. We can illustrate this rule by the *commutative diagram*

$$\begin{array}{ccc} G \times G & \xrightarrow{\Phi\times\Phi} & H \times H \\ \downarrow & & \downarrow \\ G & \xrightarrow[\Phi]{} & H \end{array}$$

where the vertical maps are the multiplication maps on G and H respectively.

The homomorphism Φ is called an *isomorphism* if it is *surjective*, that is all the elements of H is the image of some element in G, and *injective*, that is if a and b are different elements in G then $\Phi(a)$ and $\Phi(b)$ are different elements in H.

The *kernel* of the homomorphism Φ is the set

$$\ker \Phi = \left\{ a \in G \;\middle|\; \Phi(a) = e_H \right\},$$

and the *image* is the set

$$\operatorname{im} \Phi = \left\{ a \in H \;\middle|\; a = \Phi(b), \quad \text{for some } b \in G \right\}.$$

The kernel and the image of a homomorphism are groups (see Exercise 1.2.3).

Example 1.2.10. The map of groups

$$\det\colon \mathrm{Gl}_n(\mathbf{C}) \to \mathbf{C}^*$$

that sends a matrix to its determinant is a homomorphism of groups because of the formula $\det AB = \det A \det B$. The kernel of this map is $\mathrm{Sl}_n(\mathbf{C})$ and the image is $\mathbf{C}^*$.

Example 1.2.11. The map

$$\Phi\colon \mathrm{Gl}_n(\mathbf{C}) \to \mathrm{Sl}_{n+1}$$

given by

$$\Phi(A) = \begin{pmatrix} (\det A)^{-1} & 0 \\ 0 & A \end{pmatrix}$$

is a homomorphism of groups. Clearly, the homomorphism Φ is injective.

Example 1.2.12. The map $\mathbf{C}^* \to \mathrm{SO}_2(\mathbf{C})$, which sends t to

$$\begin{pmatrix} \frac{1}{2}(t+t^{-1}) & \frac{i}{2}(t-t^{-1}) \\ -\frac{i}{2}(t-t^{-1}) & \frac{1}{2}(t+t^{-1}) \end{pmatrix},$$

is a group homomorphism (see Exercise 1.2.9).

Example 1.2.13. Let

$$\Phi\colon \mathfrak{S}_n \to \mathrm{Gl}_n(\mathbf{C})$$

be the map sending σ to the matrix having coordinates 1 in the position $(\sigma(i), i)$, for $i = 1, \dots, n$, and the remaining coordinates 0. It is clear that Φ is injective.

Let $e_i = (0, \dots, 1, \dots, 0)$ be the $1 \times n$ vector with coordinate 1 in the i'th position, for $i = 1, \dots, n$. We have that

$$\Phi(\sigma)^t e_i = {}^t e_{\sigma(i)}.$$

Consequently we have that

$$\Phi(\tau)\Phi(\sigma)^t e_i = \Phi(\tau)^t e_{\sigma(i)} = {}^t e_{\tau\sigma(i)} = \Phi(\tau\sigma)^t e_i,$$

that is, $\Phi(\tau)\Phi(\sigma) = \Phi(\tau\sigma)$. Thus Φ is a group homomorphism.

The image of Φ consists of matrices with determinant ± 1. We define the map $\mathrm{sign}\colon \mathfrak{S}_n \to \mathbf{C}^*$ by

$$\mathrm{sign}\,\sigma = \det \Phi(\sigma), \qquad \text{for } \sigma \in \mathfrak{S}_n,$$

and obtain from Example 1.2.10 that

$$\mathrm{sign}\,\tau\sigma = \mathrm{sign}\,\tau\,\mathrm{sign}\,\sigma.$$

In other words, the map

$$\mathrm{sign}\colon \mathfrak{S}_n \to \{\pm 1\},$$

into the group with two elements 1 and -1, under multiplication, is a group homomorphism.

Proposition 1.2.14. *Let $\Phi\colon G \to H$ be a homomorphism between the groups G and H, with identity e_G and e_H, respectively. Then*

(i) $\Phi(e_G) = e_H$.
(ii) $\Phi(a^{-1}) = \Phi(a)^{-1}$, *for all* a *in* G.

Proof: We have that

$$e_H = \Phi(e_G)\Phi(e_G)^{-1} = \Phi(e_Ge_G)\Phi(e_G)^{-1} = \Phi(e_G)\Phi(e_G)\Phi(e_G)^{-1} = \Phi(e_G).$$

Moreover, we have that

$$\Phi(a^{-1}) = \Phi(a^{-1})\Phi(a)\Phi(a)^{-1} = \Phi(a^{-1}a)\Phi(a)^{-1} = \Phi(e_G)\Phi(a)^{-1} = e_H\Phi(a)^{-1} = \Phi(a)^{-1}.$$

□

Proposition 1.2.15. *A homomorphism* $\Phi\colon G \to H$ *of groups is injective if and only if the kernel is* $\{e_G\}$. *In other words,* Φ *is injective if and only if* $\Phi(a) = e_H$ *implies that* $a = e_G$.

Proof: If Φ is injective then $\Phi(a) = e_H$ implies $a = e_G$, by the definition of injectivity, and because $\Phi(e_G) = e_H$.

Conversely, assume that $\Phi(a) = e_H$ implies that $a = e_G$. If $\Phi(a) = \Phi(b)$, we have that

$$\Phi(ab^{-1}) = \Phi(a)\Phi(b^{-1}) = \Phi(a)\Phi(b)^{-1} = \Phi(b)\Phi(b)^{-1} = e_H.$$

Hence, $ab^{-1} = e_G$ and $a = ab^{-1}b = e_Gb = b$. □

Definition 1.2.16. A *subgroup* H of a group G is a subset H of G such that for all a and b in H we have that ab and a^{-1} are in H. A subgoup H of G is *normal* if bab^{-1} is in H for all b in G and a in H.

Remark 1.2.17. A subgroup H of G is itself a group. Indeed, the associative law (i) of 1.2.1 holds for all elements of G and thus for all elements of H. By definition the inverse of every element of H is in H and if a is in H then $aa^{-1} = e_G$ is in H, and is the identity element in H too. When H is a subgroup of G we can consider the inclusion of H in G as a map $\varphi\colon H \to G$, which sends an element to itself, that is $\varphi(a) = a$, for all a in H. This map is then a group homomorphism, often called the *inclusion map*.

Exercises

1.2.1. Show that none of the groups $\mathrm{Gl}_n(\mathbf{C})$, $\mathrm{Sl}_n(\mathbf{C})$, $\mathrm{Sp}_n(\mathbf{C})$, $\mathfrak{S}_n$ are abelian when $n > 2$.

1.2.2. Show that the composite map $\Psi\Phi\colon F \to H$ of the two homomorphisms $\Phi\colon F \to G$ and $\Psi\colon G \to H$ of groups is again a homomorphism of groups.

1.2.3. Let $\Phi : G \to H$ be a homomorphism of groups.

(a) Show that the kernel and image of Φ are subgroups of G and H respectively.
(b) Show that the kernel is a normal subgroup of G.

1.2.4. Which of the following maps $\mathrm{Gl}_n(\mathbf{C}) \to \mathrm{Gl}_n(\mathbf{C})$ are group homomorphisms?

(a) $X \to {}^tX$.
(b) $X \to \overline{X}$ *(complex conjugation).*
(c) $X \to {}^tX^{-1}$.
(d) $X \to X^{-1}$.
(e) $X \to A^{-1}XA$, *where* A *is in* $\mathrm{Gl}_n(\mathbf{C})$.

1.2.5. Does the set $\{A \in \mathrm{Gl}_2(\mathbf{R}) | \quad A^2 = I_2\}$ form a subgroup of $\mathrm{Gl}_2(\mathbf{R})$?

1.2.6. Show that for any orthogonal matrix A over the complex numbers, the map $X \to {}^tAXA$ defines a group isomorphism

$$\mathrm{G}_S(\mathbf{C}) \to \mathrm{G}_{{}^tASA}(\mathbf{C}).$$

1.2.7. Show that $\mathrm{Gl}_1(\mathbf{C})$ is isomorphic to $\mathbf{C}^*$, and find all subgroups of these groups with a finite number of elements.

1.2.8. A *Möbius transformation* of the *complex plane* is a map $\mathbf{C} \to \mathbf{C}$ which sends z in $\mathbf{C}$ to $\frac{az+b}{cz+d}$, where a, b, c, d are complex numbers such that $ad - bc \neq 0$.

(a) Show that the Möbius transformations of the complex plane form a group under composition.
(b) Show that this group is isomorphic to one of the groups of Section 1.

1.2.9. Let $\mathbf{C}^*$ be the non-zero elements of $\mathbf{C}$.

(a) Show that the map $\mathbf{C}^* \to \mathrm{SO}_2(\mathbf{C})$, which sends t to
$$\begin{pmatrix} \frac{1}{2}(t+t^{-1}) & \frac{i}{2}(t-t^{-1}) \\ -\frac{i}{2}(t-t^{-1}) & \frac{1}{2}(t+t^{-1}) \end{pmatrix},$$
is a group homomorphism.
(b) Let C be the subgroup of $\mathbf{C}^*$ consisting of complex numbers z such that $|z| = 1$. Is the restriction of Φ to C an isomorphism onto $\mathrm{SO}_2(\mathbf{R}) \subseteq \mathrm{SO}_2(\mathbf{C})$?

1.3 Rings and fields

We have that $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$ and $\mathrm{M}_n(\mathbf{C})$ are abelian groups under addition. They also have a multiplication. The non-zero elements of $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{C}$ form abelian groups with respect to the multiplication, whereas the non-zero elements of $\mathbf{Z}$ and $\mathrm{M}_n(\mathbf{C})$ are not groups under multiplication (see Exercise 1.3.1).

Definition 1.3.1. A *ring* is a set R with an addition and a multiplication, such that R is an abelian group under addition and such that all triples of elements a, b and c of R satisfy the following properties:

(i) (*Distributivity*) $(a+b)c = ab + bc$ and $a(b+c) = ab + ac$.
(ii) (*Identity*) There is an element 1 in R such that $a1 = 1a = a$.
(iii) (*Associativity*) $a(bc) = (ab)c$.

Here $a+b$ and ab are the sum and product of a and b. We shall denote by 0 – *zero* – the identity of the addition. When $ab = ba$ for all a and b in R we say that the ring R is *commutative*. The ring R is called a *skew field* when the non-zero elements form a group under multiplication, that is, when every non-zero element has a multiplicative inverse. A commutative skew field is called a *field*.

A subset I of a ring R is called an *ideal* if it is an additive subgroup, and, if for all a in R and b in I, we have that ab is in I.

Remark 1.3.2. From the above axioms one easily verifies that the usual rules for computation by numbers hold. We have, for example, $0a = (0-0)a = 0a - 0a = 0$, and $-1a + a = -1a + 1a = (-1+1)a = 0a = 0$, so that $-1a = -a$.

Example 1.3.3. We have that $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{C}$ are fields.

Example 1.3.4. We have that $\mathbf{F}_2 = \{0, 1\}$, where $1+1 = 0$, is a field, as is $\mathbf{F}_3 = \{0, \pm 1\}$, where $1+1 = -1$.

Example 1.3.5. We have that $\mathbf{Z}$ and $\mathrm{M}_n(\mathbf{C})$ are rings, but not fields.

Example 1.3.6. Let S be a set and R a ring. The set R^S consisting of all maps from S to R forms a ring. We define the sum $f+g$ of two maps $f : S \to R$ and $g : S \to R$ by $(f+g)(x) = f(x) + g(x)$ and the product fg by $(fg)(x) = f(x)g(x)$ (see Exercise 1.3.3).

The following example of rings is extremely important in algebra and geometry.

Example 1.3.7. (Polynomial and formal power series rings.) Let R be a commutative ring. In the previous example we saw how the set $R^{\mathbf{N}}$ of maps $\mathbf{N} \to R$ form a ring, in a natural way. In this example we shall give the group $R^{\mathbf{N}}$ a different multiplicative structure that also makes it into a ring.

For each element a of R we let, by abuse of notation, the map $a\colon \mathbf{N} \to R$ be defined by $a(0) = a$ and $a(i) = 0$ for $i > 0$. In this way we consider R as a subset of $R^{\mathbf{N}}$. Moreover, we define maps

$$x_i\colon \mathbf{N} \to R, \quad \text{for } i = 0, 1, \dots,$$

by $x_i(i) = 1$ and $x_i(j) = 0$, when $i \neq j$. Given elements $r_0, r_1, \dots$ of R we denote by

$$\sum_{i=0}^{\infty} r_i x_i\colon \mathbf{N} \to R$$

the map defined by $(\sum_{i=0}^{\infty} r_i x_i)(j) = r_j$. Clearly all maps $f\colon \mathbf{N} \to R$ can be expressed uniquely as $f = \sum_{i=0}^{\infty} f(i) x_i$. We can now define multiplication of elements f and g of $R^{\mathbf{N}}$ by

$$fg = \sum_{i=0}^{\infty} f(i)x_i \sum_{i=0}^{\infty} g(i)x_i = \sum_{k=0}^{\infty} (\sum_{i+j=k} f(i)g(j)) x_k.$$

It is easy to check that this multiplication, together with the addition of of Example 1.3.6, gives a ring structure on $R^{\mathbf{N}}$ (Exercise 1.3.4). We note that with the given multiplication we have that

$$x_1^i = x_1 \cdots x_1 = x_i.$$

Let $x = x_1$. Every element can thus be uniquely written as a *power series*

$$f = \sum_{i=0}^{\infty} f(i) x^i,$$

and multiplication is similar to that for convergent power series. We denote by $R[[x]]$ the ring $R^{\mathbf{N}}$, with the given multiplication and call it the *ring of power series* in the variable x with coefficients in the ring R.

The subset of $R[[x]]$ consisting of elements $f = \sum_{i=0}^{\infty} f(i)x^i$ such that only a finite number of coefficients $f(i)$ are non-zero forms a ring under the addition and multiplication induced by those on $R[[x]]$. This ring is denoted by $R[x]$ and is called the *ring of polynomials* in the variable x with coefficients in the ring R.

Remark 1.3.8. The advantage of defining polynomial and power series rings with coefficients in a ring is that the construction can be used inductively to define polynomial and power series rings in several variables. Indeed, starting with R we have constructed a polynomial ring $R[x_1]$. Then, starting with $R[x_1]$ we may construct a polynomial ring $R[x_1][x_2]$, which we denote by $R[x_1, x_2]$. In this way we can continue to construct polynomial rings $R[x_1, \dots, x_n]$ in n variables. Similarly, we can define the power series ring $R[[x_1, \dots, x_n]]$ in n variables.

Definition 1.3.9. Let R and S be rings. A map $\Phi\colon R \to S$ is a *ring homomorphism* if, for all pairs a, b of R, we have that:

(i) $\Phi(a + b) = \Phi(a) + \Phi(b)$.
(ii) $\Phi(ab) = \Phi(a)\Phi(b)$.
(iii) $\Phi(1) = 1$.

Remark 1.3.10. Since there need not be any inverses of the elements with respect to multiplication, we have to let $\Phi(1) = 1$ be an axiom, while in a group it follows immediately that a homomorphism has to map the identity element to the identity element.

The *kernel* of a ring homomorphism is the set $\ker \Phi = \{a \in R\colon \Phi(a) = 0\}$, that is, the kernel of the map of additive groups. When R is a subset of S such that the inclusion map is a ring homomorphism and such that $ab = ba$ for all a in R and b in S, we call R a *subring* of S and we say that S is an *algebra over* R, or an R*-algebra*.

Example 1.3.11. We have seen that $\mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R} \subset \mathbf{C}$ is a sequence of subrings, and that the same is true for the sequence $R \subset R[x] \subset R[[x]]$. In particular we have that $R[x]$ and $R[[x]]$ are R-algebras.

Example 1.3.12. Let $\mathrm{M}_2(\mathbf{R})$ be the set of all 2×2 matrices with real coordinates, and let

$$\Phi\colon \mathbf{C} \to \mathrm{M}_2(\mathbf{R})$$

be the map defined by

$$\Phi(z) = \Phi(x+iy) = \begin{pmatrix} x & y \\ -y & x \end{pmatrix}.$$

Then Φ is an injective ring homomorphism (see Exercise 1.3.5).

Example 1.3.13. Let $\mathrm{M}_4(\mathbf{R})$ be the set of 4×4 matrices with real coordinates. Let

$$\mathbf{H} = \left\{ \begin{pmatrix} a & b & c & d \\ -b & a & -d & c \\ -c & d & a & -b \\ -d & -c & b & a \end{pmatrix} \;\middle|\; a,b,c,d \text{ in } \mathbf{R} \right\}.$$

Moreover, let

$$i = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, j = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, k = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}.$$

Every element in $\mathbf{H}$ can be written uniquely in the form $a+ib+jc+kd$, for real numbers a,b,c,d, where we write a instead of aI_4. Consequently the sum of two elements in $\mathbf{H}$ is again in $\mathbf{H}$. We have relations

$$ij = k, jk = i, ki = j, \text{ and } i^2 = j^2 = k^2 = -1. \tag{1.3.13.1}$$

From the relations 1.3.13.1 it follows that the product of two elements in $\mathbf{H}$ is again in $\mathbf{H}$. The set $\mathbf{H}$ with the addition and multiplication induced by the addition and multiplication in $\mathrm{M}_4(\mathbf{R})$ is a ring (see Exercise 1.3.6). Consider $\mathbf{C}$ as the subset $x+iy+j0+k0$ of $\mathbf{H}$. Then $\mathbf{C}$ is a subring of $\mathbf{H}$ (see Exercise 1.3.6).

Every non-zero element $a+ib+jc+kd$ of $\mathbf{H}$ has the inverse $(a^2+b^2+c^2+d^2)^{-1}(a-ib-jc-kd)$. Hence $\mathbf{H}$ is a skew field called the *quaternions*. It is however, not a field (see Exercise 1.3.6).

Example 1.3.14. We have a ring homomorphism $\mathbf{H} \to \mathrm{Gl}_2(\mathbf{C})$ defined by

$$a+ib+jc+kd \longmapsto \begin{pmatrix} a+ib & c+id \\ -c+id & a-ib \end{pmatrix}.$$

This homomorphism sends the subset $\{a+ib+jc+kd \mid a^2+b^2+c^2+d^2 = 1\}$ isomorphically onto $\mathrm{Sp}_2(\mathbf{C})$ (see Exercise 1.3.8).

Example 1.3.15. Let R be a ring. We can define a new ring $R[\varepsilon]$, sometimes called the *ring of dual numbers*, as follows:

As a group $R[\varepsilon]$ is the set $R\times R$ with addition defined by $(a,b)+(c,d) = (a+c, b+d)$. This clearly defines an additive group with zero $(0,0)$. We define a multiplication on $R\times R$ by $(a,b)(c,d) = (ac, ad+bc)$. It is easily checked that $R\times R$ becomes a ring $R[\varepsilon]$ with zero $0 = (0,0)$ and unit $1 = (1,0)$.We define the multiplication of an element a of R with (b,c) by $a(b,c) = (ab, ac)$. Write $\varepsilon = (0,1)$, and identify $\mathbf{R}$ with the subring of $\mathbf{R}[\varepsilon]$ via the homomorphism $\mathbf{R} \to \mathbf{R}[\varepsilon]$ that maps a to $(a,0)$. Every element in $R[\varepsilon]$, can be written uniquely as $(a,b) = a+b\varepsilon$, and the multiplication is given by the multiplication of R and the rule $\varepsilon^2 = 0$.

Example 1.3.16. The kernel of a homomorphism $S \to R$ of rings is an ideal in S (see Exercise 1.3.7).

Remark 1.3.17. Let $\mathbf{K}$ be a field. We write n for the sum $1 + \cdots + 1$ of the unit in $\mathbf{K}$ taken n times. There are two possibilities:

(i) We have that none of the elements n are equal to 0 in $\mathbf{K}$. For each pair of elements m and n of $\mathbf{K}$ we can then define the element $m/n = mn^{-1}$. We can define a map $\mathbf{Q} \to \mathbf{K}$ by sending m/n in $\mathbf{Q}$ to m/n in $\mathbf{K}$. Clearly, this map is injective. In this case we say that $\mathbf{K}$ has *characteristic* 0 and consider $\mathbf{Q}$ as a subfield of $\mathbf{K}$.

(ii) There is an integer n such that n is 0 in $\mathbf{K}$. Since $-n = 0$ if $n = 0$ in $\mathbf{K}$ we can assume that n is positive. Let p be the smallest positive integer such that $p = 0$ in $\mathbf{K}$. Then p is a prime number because if $p = qr$ we have that $p = qr$ in $\mathbf{K}$ and hence $p = 0$ implies that $q = 0$ or $r = 0$, since $\mathbf{K}$ is a field. In this case we obtain a ring homomorphism $\mathbf{Z} \to \mathbf{K}$ with kernel $p\mathbf{Z}$. We say that $\mathbf{K}$ has *characteristic* p.

Example 1.3.18. The group $\{0, 1\}$ with two elements, where $1 + 1 = 0$, is a field of characteristic 2.

Exercises

1.3.1. Show that the non-zero elements of $\mathbf{Z}$ and $\mathrm{M}_n(\mathbf{C})$ are not groups under multiplication.

1.3.2. Show that the only ideals of a field $\mathbf{K}$ are (0) and $\mathbf{K}$.

1.3.3. Show that the set R^S of Example 1.3.6 with the addition and multiplication given there form a ring.

1.3.4. Show that the set $R^{\mathbf{N}}$ with the addition and multiplication given in Example 1.3.7 form a ring.

1.3.5. Show that the map Φ of Example 1.3.12 is a ring homomorphism.

1.3.6. Show that the set $\mathbf{H}$ of Example 1.3.13 is a ring which is not commutative and that $\mathbf{C}$ is a subring via the inclusion of that example.

1.3.7. Prove that the kernel of a ring homomorphism $S \to R$ is an ideal in S.

1.3.8. Show that the map $\mathbf{H} \to \mathrm{Gl}_2(\mathbf{Z})$ of Example 1.3.14 is a homomorphism of rings.

1.3.9. Let S be a set and G a group. We say that G *acts* on S if there is a group homomorphism $\Phi : G \to \mathfrak{S}_S$ to the *symmetric group* of the set S. For x in X and g in G we write $gx = \Phi(g)(x)$. The subset

$$Gx = \{gx \mid g \in G\}$$

of S is called the *orbit* of x, and the subset

$$G_x = \{g \in G \mid gx = x\}$$

of G is called the *stabilizer* of x.

(a) Assume that G has a finite number $|G|$ of elements. Show that

$$|G| = |Gx|\,|G_x|.$$

(b) Let $\mathbf{K}$ be a field with q elements. Compute $|\operatorname{Gl}_n|$, $|Sl_n|$, and $|O_n|$. If you do not succeed proving part (b) for arbitrary n, then try small values of n.

1.4 Matrix groups over arbitrary fields

Most of the theory of matrices that we shall need holds for matrices with coefficients in arbitary fields. The basic results are similar, and and the techniques for proving them are independent of the field. In this section we shall introduce generalizations to arbitrary fields of the *classical* matrix groups of Section 1.1.

Fix a field $\mathbf{K}$. Denote by $\mathrm{M}_{m,n}(\mathbf{K})$ the set of $m \times n$ matrices with coordinates in $\mathbf{K}$, and let $\mathrm{M}_n(\mathbf{K}) = \mathrm{M}_{n,n}(\mathbf{K})$. The *determinant* of a matrix $A = (a_{ij})$ in $\mathrm{M}_n(\mathbf{K})$ is the expression

$$\det A = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sign} \sigma \, a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

For a pair of matrices A, B of $\mathrm{M}_n(\mathbf{K})$ we have that

$$\det(AB) = \det A \det B$$

(see Exercise 1.4.1). Moreover, for each matrix A of $\mathrm{M}_n(\mathbf{K})$, there is an *adjoint matrix* B such that

$$AB = BA = (\det A) I_n$$

(see Exercise 1.4.2). Consequently, when A is *non-singular*, that is $\det A \neq 0$, then A has the inverse $(\det A)^{-1}B$. Hence, the matrices $\mathrm{Gl}_n(\mathbf{K})$ in $\mathrm{M}_n(\mathbf{K})$ with non-zero determinant form a group. Moreover, the subset $\mathrm{Sl}_n(\mathbf{K})$ of $\mathrm{Gl}_n(\mathbf{K})$ consisting of matrices of determinant 1 form a subgroup. These groups are called the *general linear group* respectively the *special linear group* over $\mathbf{K}$.

We have that, for a fixed matrix S in $\mathrm{M}_n(\mathbf{K})$, the subset $\mathrm{G}_S(\mathbf{K})$ of matrices A in $\mathrm{Gl}_n(\mathbf{K})$ such that

$${}^tASA = S$$

form a subgroup of $\mathrm{Gl}_n(\mathbf{K})$, as does the subset $\mathrm{SG}_S(\mathbf{K})$ of $\mathrm{G}_S(\mathbf{K})$ consisting of matrices with determinant 1 (see Exercise 1.4.4). The particular cases when $S = I_n$, that is matrices that satisfy

$${}^tAA = I_n,$$

are denoted by $\mathrm{O}_n(\mathbf{K})$ and $\mathrm{SO}_n(\mathbf{K})$ and called the *orthogonal group* respectively *special orthogonal group* over $\mathbf{K}$.

Remark 1.4.1. As we indicated in 1.1.1 we shall, in Sections 1.7 and 1.8 interpret the orthogonal and symplectic groups in terms of bilinear forms, and we shall see that there are more groups which it is natural to call *orthogonal*.

Finally, let J_m be the matrix in $\mathrm{M}_m(\mathbf{K})$ with 1 on the *antidiagonal*, that is the elements a_{ij} with $i + j = m + 1$ are 1, and the remaining coordinates 0. Take

$$S = \begin{pmatrix} 0 & J_m \\ -J_m & 0 \end{pmatrix}. \tag{1.4.1.1}$$

The corresponding set $\mathrm{G}_S(\mathbf{K})$ is denoted by $\mathrm{Sp}_{2m}(\mathbf{K})$ and is called the *symplectic group* over $\mathbf{K}$. When we write $\mathrm{Sp}_n(\mathbf{K})(\mathbf{K})$ we always assume that n is even.

Exercises

1.4.1. Show that, for a pair of matrices A, B of $\mathrm{M}_n(\mathbf{K})$, we have that

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

1.4.2. For each matrix A of $\mathrm{M}_n(\mathbf{K})$, the *adjoint matrix* B is defined by $B_{ij} = (-1)^{i+j} \det A^{(j,i)}$, where $A^{(i,j)}$ denotes the submatrix of A obtained by deleting the i'th row and the j'th column. Show that B satisfies

$$AB = BA = (\det A)I_n.$$

1.4.3. Let $a_{i1}x_1 + \cdots + a_{in}x_n = b_i$, for $i = 1, \dots, n$ be a system of n equations in the n variables $x_1, \dots, x_n$. Show that if the $n \times n$ matrix $A = (a_{ij})_{i=1,\dots n, j=1,\dots,n}$ is non-singular, then the equations have a unique solution given by $a_i = (-1)^i \frac{\det A_i}{\det A}$, where A_i is the matrix obtained from A by substituting the column ${}^t(b_1, \dots, b_n)$ for the i'th column of A.

1.4.4. Show that, for a fixed matrix S in $\mathrm{M}_n(\mathbf{K})$, the subset $\mathrm{G}_S(\mathbf{K})$ of matrices A in $\mathrm{Gl}_n(\mathbf{K})$ such that

$${}^tASA = S$$

form a subgroup of $\mathrm{Gl}_n(\mathbf{K})$, as does the subset $\mathrm{SG}_S(\mathbf{K})$ of $\mathrm{G}_S(\mathbf{K})$ consisting of matrices with determinant 1.

1.4.5. Determine the 1-dimensional Lorentz group. That is, all matrices A in $\mathrm{M}_n(\mathbf{R})$ such that ${}^tA \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

1.4.6. Let $\mathbf{K} = \mathbf{R}$. Show that $\mathrm{SO}_2(\mathbf{R})$ consists of the matrices $\left(\begin{smallmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{smallmatrix}\right)$. Determine $\mathrm{O}_2(\mathbf{R})$.

1.4.7. Let $\mathbf{K}$ be the field with 2 elements. That is $\mathbf{K} = \{0, 1\}$, with $1+1 = 0$. Determine $\mathrm{Gl}_2(\mathbf{K})$, $\mathrm{Sl}_2(\mathbf{K})$, $\mathrm{O}_2(\mathbf{K})$, $\mathrm{SO}_2(\mathbf{K})$, and $\mathrm{Sp}_2(\mathbf{K})$. Which of these groups are isomorphic?

1.4.8. Determine the 1-dimensional Lorentz group. That is, all matrices A in $\mathrm{M}_n(\mathbf{R})$ such that ${}^tA \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$.

1.5 Generators for groups

Given a group G and elements $\{a_i\}_{i \in I}$ of G. The intersection of all subgroups of G that contain all the elements a_i we denote by G'. The intersection of any family of sugroups of a group G is again a subgroup of G (see Exercise 1.5.1). Consequently we have that G' is a group. We call this group the group *generated* by the elements $\{a_i\}_{i \in I}$ and say that these elements are *generators* of the group G'. The elements of G' can be expressed in an explicit way as follows:

Let G'' be the set of all elements of the form

$$a_{i_1}^{d_1} a_{i_2}^{d_2} \cdots a_{i_m}^{d_m}, \tag{1.5.0.2}$$

for all $m \in \mathbf{N}$, for all sequences $\{i_1, i_2, \dots, i_m\}$ of elements in I and for all sequences $\{d_1, d_2, \dots, d_m\}$ of exponents ± 1. Clearly the set G'' is a subgroup of G. Hence $G'' \subseteq G'$. On the other hand we have that all the element of G'' have to be in any subgroup of G that contains all a_i. Consequently we have that $G' = G''$.

Example 1.5.1. The additive group $\mathbf{Z}$ is generated by the element 1, and the additive group of $\mathbf{Q}$ is generated by all elements of the form $1/p^n$, where $n \in \mathbf{N}$ and p is a prime number.

We shall, in the following, find generators for the groups of Section 1.4.

To find the generators for $\mathrm{Gl}_n(\mathbf{K})$ and $\mathrm{Sl}_n(\mathbf{K})$ we use a well known method of linear algebra often called *Gaussian elemination.* We recall how this is done. Let $E_{ij}(a)$, for $i, j = 1, \dots, n$ and $i \neq j$ be the matrices of $\mathrm{M}_n(\mathbf{K})$ that have 1's on the diagonal, $a \in \mathbf{K}$ in the (i, j)-coordinate, and 0 in all other coordinates. We shall call the matrices $E_{ij}(a)$ the *elementary matrices.* Clearly, $\det E_{ij}(a) = 1$, so $E_{ij}(a)$ is in $\mathrm{Sl}_n(\mathbf{K})$. For every matrix A in $\mathrm{M}_n(\mathbf{K})$ we have that the matrix $E_{ij}(a)A$ is obtained from A by adding a times the j'th row of A to the i'th and leaving the remaining coordinates unchanged. Similarly $AE_{ij}(a)$ is obtained by adding a times the i'th colunm of A to the j'th and leaving the remaining coordinates unchanged. In particular $E_{ij}(a)E_{ij}(-a) = I_n$ so that the inverse of an elementary matrix is again elementary.

Proposition 1.5.2. *The group $\mathrm{Sl}_n(\mathbf{K})$ is generated by the elementary matrices, and the group $\mathrm{Gl}_n(\mathbf{K})$ is generated by the elementary matrices and the matrices of the form*

$$\begin{pmatrix} I_{n-1} & 0 \\ 0 & a \end{pmatrix} \tag{1.5.2.1}$$

with $a \neq 0$ in $\mathbf{K}$.

Proof: Let A be in $\mathrm{Gl}_n(\mathbf{K})$. Not all the entries in the first column are zero. If a_{i1} is not zero for some $i > 1$, we multiply A to the left with $E_{1i}(a_{i1}^{-1}(1 - a_{11}))$ and obtain a matrix whose $(1, 1)$ coordinate is 1. On the other hand, if a_{11} is the only non-zero entry in the first column, we multiply A to the left with $E_{21}(1)E_{21}(a_{11}^{-1}(1 - a_{11}))$, and again obtain a matrix whose $(1, 1)$ coordinate is 1. We can now multiply the resulting matrix, to the right and to the left, with matrices of the form $E_{1i}(a)$, respectively $E_{i1}(a)$, to obtain a matrix of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & A' \end{pmatrix},$$

for some A' in $\mathrm{Gl}_{n-1}(\mathbf{K})$. We can thus use induction on n to reduce the $(n-1) \times (n-1)$ matrix in the lower right corner to a matrix of the form 1.5.2.1, using only elementary matrices of the form $E_{ij}(a)$, with $i, j > 1$.

Thus multiplying the matrix A to the left and to the right with elementary matrices it can be put in the form 1.5.2.1. Multiplying with the inverses of the elementary matrices that appear we obtain the assertion of the proposition for $\mathrm{Gl}_n(\mathbf{K})$. To prove it for $\mathrm{Sl}_n(\mathbf{K})$ we only have to observe that, since the elementary matrices are in $\mathrm{Sl}_n(\mathbf{K})$, we have that the resulting matrix 1.5.2.1 also must be in this group. Consequently, we must have that $a = 1$. □

In order to find generators for the groups $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$ and $\mathrm{Sp}_n(\mathbf{K})$ it is convenient to introduce vector spaces over arbitrary fields and to view the elements of these groups as automorphisms of bilinear forms. We shall do this in Sections 1.6 and 1.7.

Exercises

1.5.1. Show that the intersection of any family of subgroups of a group is again a subgroup.

1.5.2. Write the matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in $\mathrm{Sl}_2(\mathbf{K})$ as a product of elementary matrices.

1.6 Vector spaces

In order to fully understand the nature of the matrix groups that were introduced in Section 1.4, they must be considered as automorphisms of bilinear forms on vector spaces. We shall show how this is done in Section 1.8. In this section we shall recall the relevant properties of vector spaces. The results we need and the methods used are the same for all fields. Consequently we discuss vector spaces over arbitrary fields. We assume that most of the basic theory is well known from a first course in linear algebra, and therefore leave many details to the reader as exercises.

Fix a field $\mathbf{K}$ and Let V be an abelian group. We shall denote the addition in $\mathbf{K}$ and V by $+$ and the zero for the addition by 0. It will be clear from the context in which of the abelian groups $\mathbf{K}$ or V we perform the addition.

Definition 1.6.1. The group V is a *vector space* over $\mathbf{K}$ if there is a map

$$\mathbf{K} \times V \to V,$$

such that, for each pair of elements a, b of $\mathbf{K}$ and x, y of V, the following four properties hold:

(i) $(a+b)x = ax + bx$,
(ii) $a(x+y) = ax + ay$,
(iii) $a(bx) = (ab)x$,
(iv) $1x = x$,

where we denote by ax the image by the element (a, x). We call the elements of $\mathbf{K}$ *scalars* and the elements of V *vectors*.

Remark 1.6.2. From the properties (i)-(iv) we can deduce all the usual rules for manipulation of numbers. For example we have that $0x = (0+0)x = 0x + 0x$. Subtracting $0x$ on both sides, we get that $0x = 0$, where the zero to the left is in K, and the one to the right is in V. Similarly, we have that $a0 = a(0+0) = a0 + a0$. Subtracting $a0$ on both sides, we get that $a0 = 0$. Moreover, we have that $-1x + x = -1x + 1x = (-1+1)x = 0$, such that $-x = -1x$. Thus $-ax = (a(-1))x = a(-1x) = a(-x)$.

The following definition gives the most important example of vector spaces.

Definition 1.6.3. The n'th Cartesian product $\mathbf{K}^n$, considered as an abelian group via coordinatewise addition, that is $x + y = (a_1, \dots, a_n) + (b_1, \dots, b_n) = (a_1 + b_1, \dots, a_n + b_n)$, is a vector space over $\mathbf{K}$ under the multiplication which sends a in $\mathbf{K}$ and $x = (a_1, \dots, a_n)$ to $(aa_1, \dots, aa_n)$. We will denote this vector space by , or sometimes just V^n.

In particular the set $\mathrm{M}_{m,n}(\mathbf{K})$ is a vector space over $\mathbf{K}$. We shall often consider $V_{\mathbf{K}}^n$ as the set $\mathrm{M}_{n,1}$, when we want to operate with an $n \times n$-matrix on $V_{\mathbf{K}}^n$ by multiplication on the left. It will be clear from the context whether the element is considered as an n-tuple or as an $n \times 1$-matrix.

Example 1.6.4. Let V and W be two vector spaces over $\mathbf{K}$. We define a vector space, called the *direct sum* of V and W, and denoted by $V \oplus W$, as follows:

The set $V \oplus W$ is the Cartesian product $V \times W$. We add two elements (x, y) and (x', y') by the rule $(x, y) + (x', y') = (x + x', y + y')$, and multiply by an element a of K by the rule $a(x, y) = (ax, ay)$. It is clear that that $V \oplus W$ becomes a vector space. We write $x + y$ instead of (x, y).

Example 1.6.5. Let V and W be two vector spaces over $\mathbf{K}$. We define a structure as vector space, called the *direct product* of V and W, on $V \times W$ by defining the sum $(x, y) + (x', y')$ of two vectors (x, y) and (x', y') to be $(x + x', y + y')$ and the *scalar product* $a(x, y)$ of an element a of $\mathbf{K}$ with the vector (x, y) to be (ax, ay).

Remark 1.6.6. As we have defined direct sum and direct product, above, there is nothing but the notation that differs, but in principle they are different concepts and we shall distinguish between them.

Definition 1.6.7. Let V be a vector space over $\mathbf{K}$. A set of vectors $\{x_i\}_{i \in I}$ *generates* V if all elements x in V can be written in the form

$$x = a_1 x_{i_1} + \cdots + a_n x_{i_n},$$

for some indices $i_1, \dots, i_n$ of I and elements $a_1, \dots, a_n$ of $\mathbf{K}$. The vectors $\{x_i\}_{i \in I}$ are *linearly independent* over $\mathbf{K}$ if there is no relation of the form

$$a_1 x_{i_1} + \cdots + a_n x_{i_n} = 0,$$

where $i_1, \dots, i_n$ in I, and $a_1, \dots, a_n$ are elements in $\mathbf{K}$, that are not all zero.

The space V is *finitely generated* if there is a set of generators with finitely many elements. A set of generators consisting of linearly independent elements is called a *basis* for V.

Example 1.6.8. The vectors $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ form a basis for the space $V^n_{\mathbf{K}}$, called the *standard basis.*

The following is the main result about generators and linear independence in finitely generated vector spaces:

Theorem 1.6.9. *Let V be a vector space that has generators $x_1, \dots, x_m$. Then any set of linearly independent elements contains at most m elements. Moreover, given a (possibly empty) subset $x_{i_1}, \dots, x_{i_r}$ of $x_1, \dots, x_m$, consisting of linearly independent elements of V, then it can be extended to a subset $\{x_{i_1}, \dots, x_{i_n}\}$ of $\{x_1, \dots, x_m\}$ that is a basis of V.*

Proof: First consider the case $m = 1$. Assume that $y_1, \dots, y_n$ are linearly independent vectors in V, where $n > 1$. Then we have that $y_1 = a_1 x_1$ and $y_2 = a_2 x_1$ for two nonzero elements a_1 and a_2 of $\mathbf{K}$. We obtain that $a_2 y_1 - a_1 y_2 = a_2 a_1 x_1 - a_1 a_2 x_1 = 0$, which contradicts the linear independence of y_1 and y_2. Hence any set of linearly independent vectors in V contains at most one element.

Consequently, we can proceed by induction on m. Assume that the first part of the theorem holds for $m - 1$. Let $y_1, \dots, y_n$ be linearly independent vectors in V. We shall show that $n \leq m$. Assume, to the contrary, that $n > m$. To obtain a contradiction we only need to consider the vectors $y_1, \dots, y_{m+1}$, that is, we consider the case $n = m+1$. Since the x_i generate V we have that

$$y_i = \sum_{j=1}^{m} a_{ij} x_j,$$

for $i = 1, \dots, m+1$ and for some a_{ij}. Since the y_i are nonzero, there is an a_{ij} which is nonzero, for each i. Renumbering the x_j if necessary, we may assume that $a_{m+1,m} \neq 0$. The vectors

$$y_i' = y_i - \frac{a_{i,m}}{a_{m+1,m}} y_{m+1} = \sum_{j=1}^{m-1} (a_{ij} - \frac{a_{i,m}}{a_{m+1,m}} a_{m+1,j}) x_j, \quad \text{for } i = 1, \dots, m$$

are in the vector space W spanned by $x_1, x_2, \dots, x_{m-1}$. Hence by the induction hypothesis we have that any set of linearly independent vectors in W contains at most $m-1$ elements.

However, $y_1', y_2', \dots, y_m'$ are linearly independent because, if $\sum_{i=1}^m b_i y_i' = 0$, for some b_i, not all zero, we get that $\sum_{i=1}^m b_i (y_i - (a_{i,m}/a_{m+1,m}) y_{m+1}) = 0$. This implies that $\sum_{i=1}^m b_i y_i - (\sum_{i=1}^m b_i a_{i,m}/a_{m+1,m}) y_{m+1} = 0$, which contradicts the linear independence of $y_1, \dots, y_{m+1}$. Thus we have a contradiction to the assumption that $n > m$, which proves the first part of the theorem.

For the second part, denote by W the vector space generated by the linearly independent vectors $x_{i_1}, \dots, x_{i_r}$. If $V = W$ we have finished. If not, there is a vector $x_{i_{r+1}}$ among $x_1, \dots, x_n$ which is not in W. Then the vectors $x_{i_1}, \dots, x_{i_{r+1}}$ are linearly independent, because if we have a linear dependence $a_1 x_{i_1} + \cdots + a_{r+1} x_{i_{r+1}} = 0$, then $a_{r+1} \neq 0$, since the first r vectors are linearly independent. Consequently, we obtain that $x_{i_{r+1}} = -(a_1/a_{r+1}) x_{i_1} - \cdots - (a_r/a_{r+1}) x_{i_r}$, which contradicts the choise of $x_{i_{r+1}}$ outside of W. If the vector space generated by $x_{i_1}, \dots, x_{i_{r+1}}$ is V we have finished the proof. If not we can pick a vector $x_{i_{r+2}}$ among $x_1, \dots, x_m$ such that $x_{i_1}, \dots, x_{i_{r+2}}$ are linearly independent. In this way we can continue until we find a subset $\{x_{i_1}, \dots, x_{i_n}\}$ of $\{x_1, \dots, x_m\}$ that is a basis of V. We have proved the second part of the theorem. □

It follows from Theorem 1.6.9, that when V is finitely generated, the smallest number of generators is equal to the largest number of linearly independent elements. This number is called the *dimension* of V, and denoted $\dim_{\mathbf{K}} V$. It also follows from the theorem that every finite dimensional vector space has a basis, and that all bases have the same number, $\dim_{\mathbf{K}} V$, of elements (see Exercise 1.6.2).

Definition 1.6.10. Let V and W be two vector spaces over $\mathbf{K}$. A map

$$\Phi \colon V \to W$$

is $\mathbf{K}$-*linear*, or simply *linear* if, for all elements a in $\mathbf{K}$ and all pairs x, y of elements of V, we have that:

(i) $\Phi(x+y) = \Phi(x) + \Phi(y)$.
(ii) $\Phi(ax) = a\Phi(x)$.

A linear map is an *isomorphism* if it is injective and surjective.

Example 1.6.11. Let $V_{\mathbf{K}}^n$ and $V_{\mathbf{K}}^m$ be the vector spaces of Example 1.6.3, and let $A = (a_{ij})$ be an $m \times n$ matrix. The map $A\colon V_{\mathbf{K}}^n \to V_{\mathbf{K}}^m$, which sends $(a_1, \dots, a_n)$ to $A^t(a_1, \dots, a_n)$ is linear.

Let U, V and W be vector spaces over $\mathbf{K}$ and let $\Phi\colon U \to V$ and $\Psi\colon V \to W$ be $\mathbf{K}$-linear maps. Then the composite map $\Psi\Phi\colon U \to W$ is a linear map (see Exercise 1.6.3).

Definition 1.6.12. Let $\Phi\colon V \to W$ be a linear map between vector spaces over $\mathbf{K}$. The *kernel* of Φ is

$$\ker \Phi = \left\{ x \in V \;\middle|\; \Phi(x) = 0 \right\},$$

and the *image* is

$$\operatorname{im} \Phi = \left\{ \Phi(x) \;\middle|\; x \in V \right\}.$$

Hence the kernel and image are the same as for maps of abelian groups.

When V is a subset of W and Φ is the inclusion, we say that V is a *subspace* of W. The image of a map $\Phi\colon V \to W$ is a subspace of W and the kernel a subspace of V.

Let U and V be two subspaces of a vector space W. If every vector z in W can be written uniquely as $x + y$, with x in U and y in V we say that W is the *direct sum* of U and V, and write .

Lemma 1.6.13. *Let V be a finite dimensional vector space and let $\Phi\colon V \to W$ a linear map to a vector space W. Then $\ker \Phi$ and $\operatorname{im} \Phi$ are both finite dimensional and*

$$\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} \ker \Phi + \dim_{\mathbf{K}} \operatorname{im} \Phi.$$

In particular, if $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} W$, then Φ is injective, or surjective, if and only if Φ is an isomorphism.

Proof: It follows from Theorem 1.6.9 that $\ker \Phi$ is finite dimensional. Since V is generated by a finite number of elements, the same is true for $\operatorname{im} \Phi$. Hence $\operatorname{im} \Phi$ is finite dimensional. Choose a basis $x_1, \dots, x_r$ of $\ker \Phi$, and a basis $y_1, \dots, y_s$ of $\operatorname{im} \Phi$. Moreover, choose elements $x_{r+1}, \dots, x_{r+s}$ of V such that $\Phi(x_{r+i}) = y_i$ for $i = 1, \dots, s$.

Then $x_1, \dots, x_{r+s}$ is a basis of V. In fact, we have that the vectors $x_1, \dots, x_{r+s}$ are linearly indepenedent because if $a_1x_1 + \dots + a_{r+s}x_{r+s} = 0$ then $0 = \Phi(a_1x_1 + \dots + a_{r+s}x_{r+s}) = a_{r+1}y_1 + \dots + a_{r+s}y_s$, and hence $a_{r+1} = \dots = a_{r+s} = 0$ since $y_1, \dots, y_s$ are linearly independent. Then $a_1x_1 + \dots + a_rx_r = 0$, so that $a_1 = \dots = a_r = 0$ because $x_1, \dots, x_r$ are linearly independent. We also have that the vectors $x_1, \dots, x_{r+s}$ generate V because for x in V we have that $\Phi(x) = b_1y_1 + \dots + b_sy_s$ for some elements $b_1, \dots, b_s$ in $\mathbf{K}$. But then $x - b_1x_{r+1} - \dots - b_sx_{r+s}$ is in $\ker \Phi$ and can consequently be written as $c_1x_1 + \dots + c_rx_{r+s}$ for some elements $c_1, \dots, c_r$ in $\mathbf{K}$. Hence we have that $x = c_1x_1 + \dots + c_rx_r + b_1x_{r+1} + \dots + b_sx_{r+s}$.

Since $x_1, \dots, x_{r+s}$ is a basis of V we have that $\dim_{\mathbf{K}}(V) = r + s$, and since $r = \dim_{\mathbf{K}} \ker \Phi$ and $s = \dim_{\mathbf{K}} \operatorname{im} \Phi$ by definition, we have proved the first part of the Lemma.

Assume that $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} W$. If Φ is injective then $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} \operatorname{im} \Phi$. Hence $\dim_{\mathbf{K}} \operatorname{im} \Phi = W$, such that Φ is surjective. Conversely, if Φ is surjective, that is, we have that $\operatorname{im} \Phi = W$ we have that $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} \ker \Phi + \dim_{\mathbf{K}} W$, and hence that $\dim_{\mathbf{K}} \ker \Phi = 0$. Consequently $\ker \Phi = 0$, and Φ is injective. □

1.6.14. We denote by $\mathrm{Hom}_{\mathbf{K}}(V, W)$ the set of all linear maps between the vector spaces V and W. The sum of two linear maps Φ and Ψ and the product of a linear map by a scalar a are defined by

$$(\Phi + \Psi)(x) = \Phi(x) + \Psi(x),$$

and

$$(a\Phi)(x) = a\Phi(x).$$

With these operations we have that $\mathrm{Hom}_{\mathbf{K}}(V, W)$ is a vector space (see Exercise 1.6.4).

The case when $W = \mathbf{K}$ is particularly important. In this case we denote the vector space $\mathrm{Hom}_{\mathbf{K}}(V, \mathbf{K})$ by $\check{V}$ and call this space the *dual space* of V.

We denote the space $\mathrm{Hom}_{\mathbf{K}}(V, V)$ by $\mathrm{M}(V)$ and the subset consisting of isomorphisms by $\mathrm{Gl}(V)$. Moreover we define the product of two elements Φ and Ψ of $\mathrm{Gl}(V)$ to be the composite map $\Phi\Psi$. With this product we have that $\mathrm{Gl}(V)$ is a group. We call $\mathrm{Gl}(V)$ the *general linear group* of V.

1.6.15. Let $\{v_i\}_{i\in I}$ be a basis for V. A linear map $\Phi\colon V \to W$ is uniquely determined by its values $\Phi(v_i)$ on the basis for $i \in I$. Conversely, given vectors $\{w_i\}_{i\in I}$ in W, then there is a unique linear map $\Psi\colon V \to W$ such that $\Psi(v_i) = w_i$, for $i \in I$. We have, for $x = a_1 v_{i_1} + \cdots + a_n v_{i_n}$, that $\Psi(x) = a_1 w_{i_1} + \cdots + a_n w_{i_n}$ (see Exercise 1.6.5).

In particular, let

$$\check{v}_i\colon V \to \mathbf{K}$$

be the linear map defined by $\check{v}_i(v_i) = 1$ and $\check{v}_i(v_j) = 0$, for $i \neq j$. The vectors $\{\check{v}_i\}_{i\in I}$ are linearly independent, and if V is finite dimensional, they span $\check{V}$, and we say that $\{\check{v}_i\}_{i\in I}$ is the *dual basis* of $\{v_i\}_{i\in I}$. In particular, when V is finite dimensional, we obtain that $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} \check{V}$ (see Exercise 1.6.5).

Remark 1.6.16. Let $v_1, \dots, v_n$ be a basis for V. Then we obtain a canonical isomorphism of vector spaces

$$\Psi\colon V \to V_{\mathbf{K}}^n$$

defined by $\Psi(a_1 v_1 + \cdots + a_n v_n) = (a_1, \dots, a_n)$. Hence every finite dimensional vector space is isomorphic to some space $V_{\mathbf{K}}^n$. This explains the importance of the spaces $V_{\mathbf{K}}^n$.

1.6.17. Let $v_1, \dots, v_n$ be a basis for the vector space V, and let $w_1, \dots, w_m$ be a basis for the vector space W. A linear map $\Phi\colon V \to W$ determines uniquely an $m \times n$-matrix $A = (a_{ij})$ in $\mathrm{M}_{m,n}(\mathbf{K})$ by the formula

$$\Phi(v_i) = a_{1i} w_1 + \cdots + a_{mi} w_m, \quad \text{for } i = 1, \dots, n.$$

Conversely, every matrix in $\mathrm{M}_{m,n}(\mathbf{K})$ determines uniquely a linear map $V \to W$, by the same formula. That is, we have a bijective correspondence

$$\mathrm{Hom}_{\mathbf{K}}(V, W) \to \mathrm{M}_{m,n}(\mathbf{K}). \tag{1.6.17.1}$$

The map 1.6.17.1 is an isomorphism of vector spaces.

Let $\Theta\colon W \to V_{\mathbf{K}}^m$ be the isomorphism corresponding to the basis $w_1, \dots, w_m$ of W. Then, if A is the matrix corresponding to a linear map $\Phi\colon V \to W$, we have the *commutative diagram*

$$\begin{array}{ccc} V & \xrightarrow{\Phi} & W \\ {\scriptstyle\Psi}\downarrow & & \downarrow{\scriptstyle\Theta} \\ V_{\mathbf{K}}^n & \xrightarrow{\Theta\Phi\Psi^{-1}} & V_{\mathbf{K}}^m, \end{array} \qquad (1.6.17.2)$$

where the lower map $\Theta\Phi\Psi^{-1}$ is given by the matrix A. That is, it sends ${}^t(a_1, \dots, a_n)$ to $A^t(a_1, \dots, a_n)$.

Remark 1.6.18. When we relate the linear maps to their expression as matrices with respect to given bases the notation becomes confusing. Indeed, it is natural to consider the vectors of $V_{\mathbf{K}}^n$ as $n \times 1$-matrices. However, if $\Phi\colon V_{\mathbf{K}}^n \to V_{\mathbf{K}}^m$ is a linear map, and A its associated matrix with respect to the standard bases, we have that $\Phi(a_1, \dots, a_n) = (b_1, \dots, b_m)$, if and only if $A^t(a_1, \dots, a_n) = {}^t(b_1, \dots, b_m)$. Hence, to use the *functional notation*, and avoid the monstrous $(b_1, \dots, b_m) = {}^t(A^t(a_1, \dots, a_n)) = (a_1, \dots, a_n)^t A$, we often consider $V_{\mathbf{K}}^n$ as $\mathrm{M}_{n,1}(\mathbf{K})$ as mentioned in 1.6.3. The above is one argument for using the notation $(x)f$ for the value of a function f at an element x. Another reason is that the latter notation looks better when we take composition of functions.

Let B and C be the invertible matrices that represent Ψ respectively Θ with respect to the given bases of V and W, respectively, and the standard basis of $V_{\mathbf{K}}^n$. Then Φ is expressed by CAB^{-1} with respect to the bases $v_1, \dots, v_n$ and $w_1, \dots, w_m$ of V respectively W. In particular, when $V = W$ and $v_i = w_i$ we have that Φ is expressed by BAB^{-1}. Consequently $\det A$ is independent of the choise of basis for V and we can define $\det \Phi$ to be $\det A = \det(BAB^{-1})$.

Definition 1.6.19. The subset of $\mathrm{Gl}(V)$ consisting of linear maps with determinant 1 is clearly a subgroup. This group is called the *special linear group* of V and is denoted by $\mathrm{Sl}(V)$.

In order to define the orthogonal and symplectic groups in this *coordinate free form* we shall introduce *bilinear maps* on vector spaces.

Exercises

1.6.1. Show that in example 1.6.11 we have that $\ker A$ consists of all solutions $(a_1, \dots, a_n)$ to the equations $a_{i1}x_1 + \cdots + a_{in}x_n = 0$, for $i = 1, \dots, n$, in the n variables $x_1, \dots, x_n$, and the image is the subspace of $V_{\mathbf{K}}^n$ generated by the columns ${}^t(a_{1j}, \dots, a_{mj})$ of A, for $j = 1, \dots, n$.

1.6.2. Let V be a finite dimensional vector space over $\mathbf{K}$. Prove that V has a basis and that the following numbers are equal

(a) The smallest number of generators of V.
(b) The largest number of linearly independent elements in V.
(c) The number of elements of any basis of V.

1.6.3. Prove that if U, V and W are vector spaces over K and that $\Phi\colon U \to V$ and $\Psi\colon V \to W$ are $\mathbf{K}$-linear maps. Then the composite map $\Psi\Phi\colon U \to W$ is a $\mathbf{K}$-linear map.

1.6.4. Show that $\operatorname{Hom}_{\mathbf{K}}(V, W)$ is a vector spaces with the addition and scalar multiplication given in 1.6.14

1.6.5. Let V be a finite dimensional vector space with basis $\{v_i\}_{i\in I}$ and let W be another vector space.

(a) Show that a linear map $\Phi\colon V \to W$ is uniquely determined by the images $\Phi(v_i)$, for $i \in I$.
(b) Let $\{w_i\}_{i\in I}$ be elements of W. Show that there is a uniqe linear map $\Phi\colon V \to W$ such that $\Phi(v_i) = w_i$, for all $i \in I$.
(c) Show that $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} \check{V}$.

1.6.6. Let V and W be vector spaces and $\{v_1, v_2, \dots, v_n\}$ and $\{w_1, w_2, \dots, w_m\}$, bases for V respectively W. Show that there is a bijective map

$$\operatorname{Hom}_{\mathbf{K}}(V, W) \to \mathrm{M}_{m,n}(\mathbf{K}),$$

which is also an isomorphism of vector spaces.

1.7 Bilinear forms

Let V be a finite dimensional vector space over a field $\mathbf{K}$.

Definition 1.7.1. Let V_1, V_2 and W be vector spaces. A *bilinear map* from the Cartesian product (see Example 1.6.5) $V_1 \times V_2$ to W is a map

$$\Phi\colon V_1 \times V_2 \to W,$$

such that, for each scalar a of $\mathbf{K}$, and each pair of vectors x_1, y_1 in V_1 and x_2, y_2 in V_2, we have that:

(i) $\Phi(x_1 + y_1, x_2) = \Phi(x_1, x_2) + \Phi(y_1, x_2)$,
(ii) $\Phi(x_1, x_2 + y_2) = \Phi(x_1, x_2) + \Phi(x_1, y_2)$,
(iii) $\Phi(ax_1, x_2) = \Phi(x_1, ax_2) = a\Phi(x_1, x_2)$.

A *bilinear form* on a vector space is a bilinear map

$$\langle\,,\rangle\colon V \times V \to \mathbf{K}.$$

It is *symmetric* if $\langle x, y\rangle = \langle y, x\rangle$ for all vectors x and y and it is *alternating* if $\langle x, x\rangle = 0$ for all vectors x. Let S be a subset of V. A vector x of V is *orthogonal* to S if $\langle x, y\rangle = 0$ for all vectors y in S. We write

$$S^\perp = \{x \in V \mid \langle x, y\rangle = 0, \quad \text{for all } y \in S\}.$$

Remark 1.7.2. An easier way to phrase that a map $V_1 \times V_2 \to W$ is bilinear, is that, for each vector x_1 in V_1 and x_2 in V_2 we have that the maps $\Phi(*, x_2)\colon V_1 \to W$ and $\Phi(x_1, *)\colon V_2 \to W$, sending y_1 to $\Phi(y_1, x_2)$, respectively y_2 to $\Phi(x_1, y_2)$, are linear. Similarly, one can define *multilinear* maps

$$\Phi\colon V_1 \times \cdots \times V_n \to W,$$

as maps such that $\Phi(x_1, \ldots, *, \ldots, x_n)\colon V_i \to W$, are linear for all x_j in V_j with $j = 1, \ldots, i-1, i+1, \ldots, n$.

For each bilinear form $\langle\, , \rangle : V \times V \to \mathbf{K}$ we have associated a linear map

$$\Psi\colon V \to \check{V},$$

which maps x in V to the map $\Psi(x)\colon V \to \mathbf{K}$ defined by $\Psi(x)(y) = \langle x, y\rangle$. The kernel of Ψ is $V^\perp$.

Definition 1.7.3. We say that the form $\langle\, , \rangle : V \times V \to \mathbf{K}$ is *non-degenerate* if Ψ is injective, that is, if $V^\perp = 0$, or equivalently, if $\langle x, y\rangle = 0$ for all $y \in V$ implies that $x = 0$.

Assume that V is finite dimensional. Since $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} \check{V}$ by Paragraph 1.6.15, we have that Ψ is injective if and only if it is an isomorphism. Assume that the form is non-degenerate. Fix y in V. If we have that $\langle x, y\rangle = 0$ for all x in V we have that $\Psi(x)(y) = 0$ for all x in V. However, since Ψ is surjective, it then follows that $\alpha(y) = 0$ for all linear maps $\alpha\colon V \to \mathbf{K}$. Consequently $y = 0$ (see Exercise 1.7.1). We have proved that for a non-degenerate form $\langle x, y\rangle = 0$ for all x in V implies that $y = 0$. Consequently, the condition to be non-degenerate is symmetric in the two arguments. That is, when the form is non-degenerate the linear map

$$\Phi\colon V \to \check{V},$$

which maps y in V to the map $\Phi(y)\colon V \to \mathbf{K}$, defined by x to $\Psi(y)(x) = \langle x, y\rangle$, is an isomorphism.

Lemma 1.7.4. *Let V be a finite dimensional vector space with a non-degenerate form, and let W be a subspace. Then we have that*

$$\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} W + \dim_{\mathbf{K}} W^\perp.$$

Proof: We have a canonical linear map $\check{V} \to \check{W}$, that maps $\alpha\colon V \to \mathbf{K}$ to $\alpha|_W\colon W \to \mathbf{K}$. This map is surjective, as is easily seen by choosing a basis for W and extending it to a basis of V, see theorem 1.6.9 and Paragraph 1.6.17. Composing the isomorphism $\Psi : V \to \check{V}$, associated to the bilinear form with this surjection, we obtain a surjective map $V \to \check{W}$ with kernel $W^\perp$. Consequently the lemma follows from Lemma 1.6.13. □

Lemma 1.7.5. *Let V be vector space with a non-degenerate form $\langle\, , \rangle$, and let W be a subspace. If $W \cap W^\perp = 0$ then we have that $V = W \oplus W^\perp$ and the form $\langle\, , \rangle$ induces a non-degenerate form on $W^\perp$.*

Proof: If $U = W + W^\perp$, we have that $U = W \oplus W^\perp$ since $W \cap W^\perp = 0$. It follows from Lemma 1.7.4 that $\dim_{\mathbf{K}} V = \dim_{\mathbf{K}} W + \dim_{\mathbf{K}} W^\perp$. Hence U is a subspace of V of dimension $\dim_{\mathbf{K}} V$. Consequently $U = V$ and we have proved the second assertion of the lemma. □

Definition 1.7.6. Let V be a vector space with a non-degenerate bilinear form $\langle\, , \rangle$, and let $\alpha\colon V \to V$ be a linear map. For each y in V we obtain a linear map $V \to \mathbf{K}$ which maps x in V to $\langle \alpha(x), y\rangle$. Since the linear map $\Psi\colon V \to \check{V}$ associated to the form is an isomorphism, there is a unique vector y' in V such that $\langle \alpha(x), y\rangle = \langle x, y'\rangle$, for all x in V. The homomorphism

$$\alpha^*\colon V \to V$$

that maps y to y' is determined by the condition $\langle \alpha(x), y\rangle = \langle x, \alpha^*(y)\rangle$ for all x and y in V and is clearly linear. It is called the *adjoint* of α.

It is clear from the definition that, given two maps α and β of $\mathrm{Hom}_{\mathbf{K}}(V, V)$ and a scalar a of $\mathbf{K}$, we have the formulas

$$(\alpha^*)^* = \alpha, \qquad (\alpha+\beta)^* = \alpha^* + \beta^*, \qquad (a\alpha)^* = a\alpha^*.$$

Definition 1.7.7. Two bilinear forms $\langle\, , \rangle_f$ and $\langle\, , \rangle_g$ on V are *equivalent* if there is an isomorphism $\alpha\colon V \to V$ such that

$$\langle \alpha(x), \alpha(y)\rangle_f = \langle x, y\rangle_g,$$

for all pairs x, y of V.

1.7.8. Let $\langle\, , \rangle$ be a bilinear form on V. Fix a basis $e_1, \dots, e_n$ of V and let $S = (c_{ij})$ be the $n \times n$ matrix with (i,j)'th coordinate $c_{ij} = \langle e_i, e_j\rangle$. Then, for $x = a_1e_1 + \dots + a_ne_n$ and $y = b_1e_1 + \dots + b_ne_n$ we have that

$$\langle x, y\rangle = (a_1, \dots, a_n)\begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \dots & c_{nn}\end{pmatrix}\begin{pmatrix} b_1 \\ \vdots \\ b_n\end{pmatrix} = \sum_{i,j=1}^{n} a_i c_{ij} b_j.$$

If follows, in particular, that the form is non-degenerate if and only if the matrix S is non-singular.

Let $\alpha\colon V \to V$ be a linear map, and let $A = (a_{ij})$ be the corresponding matrix map as in Paragraph 1.6.17. The adjoint map α^* corresponds to the matrix $S^{-1\,t}AS$.

We have that the bilinear form is symmetric if and only if S is symmetric, that is $S = {}^tS$, and it is alternating if and only if $S = -{}^tS$ and $c_{ii} = 0$ for $i = 1, \dots, n$.

Let $f_1, \dots, f_n$ be another basis for V and let T be the matrix associated to the bilinear form, with respect to this basis. Moreover, let $A = (a_{ij})$ be the non-singular matrix defined by $f_i = \sum_{j=1} a_{ji}e_j$. Then $x = \sum_{i=1}^n c_if_i = \sum_{i=1}^n a_ie_i$, and $y = \sum_{i=1}^n d_if_i = \sum_{i=1}^n b_ie_i$ with ${}^t(a_1, \dots, a_n) = A^t(c_1, \dots, c_n)$, respectively ${}^t(b_1, \dots, b_n) = A^t(d_1, \dots, d_n)$. Hence $\langle x, y\rangle = (a_1, \dots, a_n)S^t(b_1, \dots, b_n) = (c_1, \dots, c_n)^tASA^t(d_1, \dots, d_n)$. Consequently we have that

$$T = {}^tASA.$$

Exercises

1.7.1. Let V be a vector space and y a vector of V. Show that if $\alpha(y) = 0$ for all α in $\check{V}$, we have that $y = 0$.

1.8 The orthogonal and symplectic groups

Let V be a vector space over $\mathbf{K}$, with a non-degenerate bilinear form $\langle\, , \rangle$. In the case of a symmetric bilinear form we will always assume that 2 is an invertible element of the field $\mathbf{K}$, i.e., that the characteristic of $\mathbf{K}$ is not equal to 2.

Lemma 1.8.1. *Assume that the form is symmetric. Then there is an element x of V such that $\langle x, x\rangle \neq 0$.*

Proof: Suppose that $\langle x, x\rangle = 0$ for all x in V. Since the form is symmetric we have that $\langle y + z, y + z\rangle = \langle y, y\rangle + 2\langle y, z\rangle + \langle z, z\rangle$ for y, z in V. Since 2 is invertible, we can rearrange this into $\langle y, z\rangle = (\langle y+z, y+z\rangle - \langle y, y\rangle - \langle z, z\rangle)/2$, which is zero by the assumption that $\langle x, x\rangle = 0$ for all x in V. However, this means that $\langle\, , \rangle$ is totally degenerate, which contradicts the assumption made in the beginning of the section that the form should be non-degenerate. Hence there must be an element x in V with $\langle x, x\rangle \neq 0$. □

Proposition 1.8.2. *Assume that the form is symmetric. Then there is a basis for V with respect to which the associated matrix (see Paragraph 1.7.8) is diagonal.*

Moreover, this basis can be chosen so that it includes any given vector x with $\langle x, x\rangle \neq 0$.

Proof: Let x be a vector of V such that $\langle x, x\rangle \neq 0$. It follows from Lemma 1.8.1 that there exists such an element x of V. Let $e_1 = x$ and let $W = \mathbf{K}e_1$. Then W is a subspace of V. Clearly we have that $W \cap W^\perp = 0$ and it follows from Lemma 1.7.5 that $V = W \oplus W^\perp$. Moreover, we have that the restriction of the bilinear form to $W^\perp$ is non-degenerate. We can therefore use induction on $\dim_{\mathbf{K}} V$ to conclude that there is a basis $e_2, \dots, e_n$ of $W^\perp$ such that $\langle e_i, e_j\rangle = 0$ and $\langle e_i, e_i\rangle \neq 0$, for $i, j = 2, \dots n$ and $i \neq j$. By definition, we also have that $\langle e_1, e_i\rangle = 0$ for $i = 2, \dots, n$. Consequently, we have proved the proposition. □

Remark 1.8.3. Another way of phrasing the assertion of the proposition is that there is a basis $e_1, \dots, e_n$ of V such that $\langle e_i, e_i\rangle = c_i$ and $\langle e_i, e_j\rangle = 0$, for $i, j = 1, \dots n$, and $i \neq j$ such that when we write $x = (a_1, \dots, a_n)$ and $y = (b_1, \dots, b_n)$, with respect to this basis, then we have that

$$\langle x, y\rangle = a_1b_1c_1 + \cdots + a_nb_nc_n.$$

Definition 1.8.4. A basis with the properties of Proposition 1.8.2 is called an *orthogonal basis*. When $c_i = 1$, for $i = 1, \dots, n$, the basis is *orthonormal*. A linear map $\alpha\colon V \to V$ such that $\langle \alpha(x), \alpha(y)\rangle = \langle x, y\rangle$, for all pairs x, y of V, is called *orthogonal*. The set of all orthogonal linear maps is denoted by $\mathrm{O}(V, \langle\, , \rangle)$. The subset consisting of linear maps with determinant 1 is denoted by $\mathrm{SO}(V, \langle\, , \rangle)$. As in section 1.1 we see that $\mathrm{O}(V, \langle\, , \rangle)$ is a subgroup of $\mathrm{Gl}(V)$, and that $\mathrm{SO}(V, \langle\, , \rangle)$ is a subgroup of $\mathrm{Sl}(V)$. We call the groups $\mathrm{O}(V, \langle\, , \rangle)$ and $\mathrm{SO}(V, \langle\, , \rangle)$ the *orthogonal group*, respectively the *special orthogonal group* of $\langle\, , \rangle$.

Remark 1.8.5. When the field $\mathbf{K}$ contains square roots of all its elements we can, given an orthogonal basis e_i, replace e_i with $\sqrt{c_i}^{-1}e_i$. We then get an orthonormal basis. In this case, we consequently have that all bilinear forms are equivalent to the form $\langle x, y\rangle = a_1b_1 + \cdots + a_nb_n$, where $x = \sum_{i=1}^n a_ie_i$ and $y = \sum_{i=1}^n b_ie_i$. This explains the choise of terminology in sections 1.1 and 1.4.

Proposition 1.8.6. *Assume that the form is alternating. We then have that $n = 2m$ is even and there is a basis $e_1, \dots, e_n$ for V, with respect to which the associated matrix (see Paragraph* 1.7.8*) is of the form*

$$S = \begin{pmatrix} 0 & J_m \\ -J_m & 0 \end{pmatrix},$$

where J_m be the matrix in $\mathrm{M}_m(\mathbf{K})$ with 1 on the antidiagonal, that is the elements a_{ij} with $i + j = m + 1$ are 1, and the remaining coordinates 0.

Moreover the basis can be chosen so that it contains any given non-zero vector x.

Proof: If $n = 1$ there is no non-degenerate alternating form. So assume that $n > 1$. Let e_1 be an arbitrary non-zero vector. Since the form is non-degenerate there is a vector v such that $\langle e_1, v\rangle \neq 0$. Let $e_n = \frac{1}{\langle e_1, v\rangle}v$. Then $\langle e_1, e_n\rangle = 1$. Let $W = \mathbf{K}e_1 + \mathbf{K}e_n$ be the subspace of V spanned by e_1 and e_n. Clearly we have that $W \cap W^\perp = 0$. It follows from Lemma 1.7.5 that $V = W \oplus W^\perp$ and that the restriction of the bilinear form to $W^\perp$ is non-degenerate. We can now use induction to conclude that $\dim_{\mathbf{K}} W^\perp$ and thus $\dim_{\mathbf{K}} V$ are even, and that there is a basis $e_2, \dots, e_{n-1}$ such that $\langle e_i, e_{n+1-i}\rangle = 1$, for $i = 2, \dots, m$ and all other $\langle e_i, e_j\rangle = 0$. However, $\langle e_1, e_i\rangle = 0 = \langle e_n, e_i\rangle$, for $i = 2, \dots, n-1$. Thus we have a basis $e_1, \dots, e_n$ as asserted in the proposition. □

Remark 1.8.7. The proposition asserts that there is a basis $\{e_1, e_2, \dots, e_n\}$ such that $\langle e_i, e_{n+1-i}\rangle = 1$, for $i = 1, \dots, m$ and all other $\langle e_i, e_j\rangle = 0$. With respect to this basis, we have that

$$\langle x, y\rangle = \sum_{i=1}^m (a_ib_{n+1-i} - a_{n+1-i}b_i).$$

It follows from the proposition that all non-degenerate alternating bilinear forms on a vector space are equivalent.

Definition 1.8.8. A basis with the properties of Proposition 1.8.6 is called a *symplectic basis*. A linear map $\alpha\colon V \to V$ such that $\langle \alpha(x), \alpha(y)\rangle = \langle x, y\rangle$, for all pairs x, y of V, is called *symplectic*. The set of all symplectic linear maps is denoted by $\mathrm{Sp}(V, \langle\,,\rangle)$. As in 1.1 we see that $\mathrm{Sp}(V, \langle\,,\rangle)$ is a subgroup of $\mathrm{Gl}(V)$, We call the group $\mathrm{Sp}(V, \langle\,,\rangle)$ the *symplectic group*, of $\langle\,,\rangle$.

1.9 Generators of the orthogonal and symplectic groups

Let V be a vector space with a fixed non-degenerate bilinear form.

Definition 1.9.1. Assume that $2 = 1 + 1$ is non-zero in $\mathbf{K}$ and that $\langle\, ,\rangle$ is symmetric. A linear map $\alpha\colon V \to V$ that fixes all the vectors in a subspace H of V of *codimension* 1, that is $\dim_{\mathbf{K}} H = \dim_{\mathbf{K}} V - 1$, and is such that $\alpha(x) = -x$ for some non-zero vector x of V, is called a *reflection* of V. Given an element x in V such that $\langle x, x\rangle \neq 0$. The map $s_x\colon V \to V$ defined by

$$s_x(y) = y - 2\frac{\langle y, x\rangle}{\langle x, x\rangle}x,$$

is clearly linear.

Remark 1.9.2. Let $e_1 = x$ and let $\{e_1, e_2, \dots, e_n\}$ be an orthogonal basis with respect to $\langle\, ,\rangle$. Then we have that $s_x(a_1e_1 + a_2e_2 + \cdots + a_ne_n) = -a_1e_1 + a_2e_2 + \cdots + a_ne_n$, and the matrix representing s_x in this basis is given by

$$\begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix}.$$

Thus the determinant of s_x is -1.

The maps of the form s_x are reflections. Indeed, let $W = \mathbf{K}x$. It follows from Lemma 1.7.4 that we have $\dim_{\mathbf{K}} W^\perp = n-1$. For $y \in W^\perp$ we have that $s_x(y) = y$ and we have that $s_x(x) = -x$. In particular s_x^2 is the identity map. Moreover, the maps s_x are orthogonal because

$$\begin{aligned}\langle s_x(y), s_x(z)\rangle &= \left\langle y - 2\frac{\langle y, x\rangle}{\langle x, x\rangle}x, z - 2\frac{\langle z, x\rangle}{\langle x, x\rangle}x\right\rangle \\ &= \langle y, z\rangle - 2\frac{\langle y, x\rangle}{\langle x, x\rangle}\langle x, z\rangle - 2\frac{\langle z, x\rangle}{\langle x, x\rangle}\langle y, x\rangle \qquad (1.9.2.1)\\ &+ 4\frac{\langle y, x\rangle\langle z, x\rangle}{\langle x, x\rangle^2}\langle x, x\rangle = \langle y, z\rangle.\end{aligned}$$

Since $\det s_x = -1$, we have that $s_x \in \mathrm{O}(V) \setminus \mathrm{SO}(V)$.

There are also reflections that are not of the form s_x for any $x \in V$ (see Exercise 1.9.2).

Lemma 1.9.3. *Let x and y be two elements of V such that $\langle x, x\rangle = \langle y, y\rangle \neq 0$. Then there is a linear map $V \to V$, which takes x to y and which is a product of at most 2 reflections of the form s_z.*

Proof: Assume that $\langle x, y\rangle \neq \langle x, x\rangle = \langle y, y\rangle$. Then $\langle x - y, x - y\rangle = 2(\langle x, x - y\rangle) = 2(\langle x, x\rangle - \langle x, y\rangle) \neq 0$. Take $z = x - y$. Then $\langle z, z\rangle \neq 0$ and $s_z(x) = x - 2\frac{\langle x, x-y\rangle}{\langle x-y, x-y\rangle}(x - y) = y$, since $2\frac{\langle x, x-y\rangle}{\langle x-y, x-y\rangle} = 1$.

On the other hand, if $\langle x, y\rangle = \langle x, x\rangle$, we have that $\langle -x, y\rangle \neq \langle x, x\rangle$ since 2 is non-zero in $\mathbf{K}$ and we see from the first part of the proof that we can take s_zs_x, with $z = -x - y$. □

Proposition 1.9.4. *The orthogonal group $\mathrm{O}(V)$ is generated by the reflections of the form s_x with $\langle x, x\rangle \neq 0$, and the subgroup $\mathrm{SO}(V)$ is generated by the products s_xs_y.*

Proof: It follows from Lemma 1.8.1 that there is an element x of V such that $\langle x, x\rangle \neq 0$. Consequently, it follows from Lemma 1.7.5 that, if $W = \mathbf{K}x$, we have that $V = W \oplus W^\perp$, and that the bilinear form induces a non-degenerate bilinear form on $W^\perp$.

Let α be an element of $\mathrm{O}(V)$. Then $\langle \alpha(x), \alpha(x)\rangle = \langle x, x\rangle \neq 0$. It follows from Lemma 1.9.3 that there is a product β of at most 2 reflections of the form s_y such that $\beta(x) = \alpha(x)$. Consequently $\beta^{-1}\alpha$ induces a linear map $\beta^{-1}\alpha|_{W^\perp}$ of $W^\perp$. We now use induction on $\dim_{\mathbf{K}} V$ to write $\beta^{-1}\alpha|_{W^\perp}$ as a product of reflections of the form s_z on $W^\perp$ for z in $W^\perp$. However, the reflection s_z considered as a reflection on $W^\perp$ is the restriction of s_z considered as a reflection on V. Hence $\beta^{-1}\alpha$ and thus α can be written as a product of reflections of the form s_z. We have proved the first part of the proposition. Since $\det s_z = -1$ we have that such a product is in $\mathrm{SO}(V)$ if and only if it contains an even number of factors. Hence the second assertion of the proposition holds. □

Definition 1.9.5. Assume that the bilinear form $\langle\, ,\rangle$ is alternating. Let x be a non-zero vector in V and a an element of $\mathbf{K}$. We define a map $\Psi\colon V \to V$ by $\Psi(y) = y - a\langle y, x\rangle x$. It is clear that Ψ is a linear map. The linear maps of this form are called *transvections*.

Remark 1.9.6. We have that each transvection is in $\mathrm{Sp}(V)$. Indeed, by Proposition 1.8.6 we can choose a symplectic basis $e_1, \dots, e_n$ for the bilinear form with $x = e_1$. Then we have that $\Psi(e_i) = e_i$ for $i \neq n$ and $\Psi(e_n) = e_n + ae_1$. Hence for $y = \sum_{i=1}^n a_i e_i$ we obtain that $\langle \Psi(x), \Psi(y)\rangle = \langle e_1, \sum_{i=1}^n a_i e_i + ae_1\rangle = \langle e_1, \sum_{i=1}^n a_i e_i\rangle = \langle x, y\rangle$. We also see that $\det \Psi = 1$.

Lemma 1.9.7. *Let $\langle\, ,\rangle$ be a non-degenerate alternating form on V. Then for every pair x, y of non-zero vectors of V there is a product of at most 2 transvections that sends x to y.*

Proof: For every pair x, y of elements of V such that $\langle x, y\rangle \neq 0$ the transvection associated to the vector $x - y$ and the element a defined by $a\langle x, y\rangle = 1$ will satisfy $\Psi(x) = y$. Indeed, $\Psi(x) = x + a\langle x - y, x\rangle(x - y) = x - a\langle x, y\rangle x + a\langle x, y\rangle y = y$.

Assume that $x \neq y$. By what we just saw it suffices to find an element z such that $\langle x, z\rangle \neq 0$ and $\langle y, z\rangle \neq 0$. We shall write $\mathbf{K}x = \langle x\rangle$ and $\mathbf{K}y = \langle y\rangle$. First we note that since $\dim_{\mathbf{K}}\langle x\rangle + \dim_{\mathbf{K}}\langle x\rangle^\perp = \dim_{\mathbf{K}} V$ by Lemma 1.7.4 we have that $\langle x\rangle^\perp$ is not equal to V. If $\langle x\rangle^\perp = \langle y\rangle^\perp$ we can take z to be any element outside $\langle x\rangle^\perp$. On the other hand if $\langle x\rangle^\perp \neq \langle y\rangle^\perp$ we take $u \in \langle x\rangle^\perp \setminus \langle y\rangle^\perp$ and $u' \in \langle y\rangle^\perp \setminus \langle x\rangle^\perp$, and let $z = u + u'$. □

Lemma 1.9.8. *Let $\langle\, ,\rangle$ be a non-degenerate alternating form on V and let x, y, x', y' be vectors in V such that $\langle x, y\rangle = 1$ and $\langle x', y'\rangle = 1$. Then there is a product of at most 4 transvections that sends x to x' and y to y'.*

Proof: By Lemma 1.9.7 we can find two transvections, whose product Φ sends x to x'. Let $\Phi(y) = y''$. Then $1 = \langle x', y'\rangle = \langle x, y\rangle = \langle x', y''\rangle$. Consequently it suffices to find two more transvections whose composition maps y'' to y' and that fix x'. If $\langle y', y''\rangle \neq 0$, we let $\Psi(z) = z + a\langle y'' - y', z\rangle(y'' - y')$ with $a = \langle y', y''\rangle$. Then we have that $\Psi(y'') = y'$, and $\Psi(x') = x'$, because $\langle y'' - y', x'\rangle = 1 - 1 = 0$. On the other hand, when $\langle y', y''\rangle = 0$, we have that $1 = \langle x', y'' \rangle = \langle x', x' + y'' \rangle$ and $\langle y'', x' + y'' \rangle \neq 0 \neq \langle y', x' + y'' \rangle$, so we can first find a composition of two transvections that map the pair (x', y'') to $(x', x' + y'')$ and then find a composition of two transvections that map the latter pair to (x', y'). □

Proposition 1.9.9. *The symplectic group* $\mathrm{Sp}(V)$ *is generated by transvections.*

In particular we have that the symplectic group is contained in $\mathrm{Sl}(V)$.

Proof: Choose a basis $e_1, e_1', \dots, e_m, e_m'$ of V such that $\langle e_i, e_i' \rangle = 1$, for $i = 1, \dots, m$, and all other products $\langle e_i, e_j \rangle$ of basis elements are 0. Let Φ be an element in the symplectic group and write $\Phi(e_i) = \bar{e}_i$ and $\Phi(e_i') = \bar{e}_i'$. We have seen above that we can find a product Ψ of transvections that maps the pair (e_1, e_1') to $(\bar{e}_1, \bar{e}_1')$. Then $\Psi^{-1}\Phi$ is the identity on the space $W = \mathbf{K}e_1 + \mathbf{K}e_1'$. Thus $\Psi^{-1}\Phi$ induces a linear map on $W^\perp$, which is generated by $e_2, e_2', \dots, e_m, e_m'$. Hence we can use induction on the dimension of V to conclude that Φ can be written as a product of transvections.

The last part of the proposition follows from Remark 1.9.6. □

Exercises

1.9.1. Write the linear map $V_{\mathbf{C}}^2 \to V_{\mathbf{C}}^2$ corresponding to the matrix $\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$, where $a^2 + b^2 = 1$, as a product of reflections, with respect to the bilinear form corresponding to the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

1.9.2. Let V be a finite dimensional vecor space with a symmetric bilinear form. Show that there are reclections that are not of the form s_x for any x in V.

1.9.3. Show that $\mathrm{Sl}_2(\mathbf{K}) = \mathrm{Sp}_2(\mathbf{K})$, and write all elements in these groups as products of transvections.

1.10 The center of the matrix groups

Definition 1.10.1. Let G be a group. The set $Z(G)$ of elements of G that commute with all elements of G, that is

$$Z(G) = \{a \in G \colon ab = ba, \text{for all } b \in G\}$$

is called the *center* of G.

It is clear that $Z(G)$ is a normal subgroup of G and that isomorphic groups have isomorphic centers.

Proposition 1.10.2. *The center of* $\mathrm{Gl}_n(\mathbf{K})$ *consists of all scalar matrices, that is all matrices of the form* aI_n *for some non-zero element* a *of* $\mathbf{K}$. *The center of* $\mathrm{Sl}_n(\mathbf{K})$ *consists of all matrices of the form* aI_n *with* $a^n = 1$.

Proof: It is clear that the matrices of the form aI_n are in the center of $\mathrm{Gl}_n(\mathbf{K})$. Moreover, we have that the center of $\mathrm{Sl}_n(\mathbf{K})$ is the intersection of the center of $\mathrm{Gl}_n(\mathbf{K})$ with $\mathrm{Sl}_n(\mathbf{K})$. Indeed, every element A of $\mathrm{Gl}_n(\mathbf{K})$ is of the form $(\det A)(\det A^{-1})A$, where $(\det A^{-1})A$ is in $\mathrm{Sl}_n(\mathbf{K})$. In particular, the last assertion of the proposition follows from the first.

Let A in $\mathrm{Gl}_n(\mathbf{K})$ be in the center. Then A must commute with the elementary matrices $E_{ij}(a)$. However, the equality $AE_{ij}(1) = E_{ij}(1)A$ implies that $a_{ij} + a_{jj} = a_{ij} + a_{ii}$ and that $a_{ii} = a_{ii} + a_{ji}$. Consequently we have that $a_{ji} = 0$ and $a_{ii} = a_{jj}$, when $i \neq j$, and we have proved the proposition. □

We shall next determine the center of the orthogonal groups.

Lemma 1.10.3. *Let V be a vector space of dimension at least 3 over a field $\mathbf{K}$ where $2 \neq 0$, and let $\langle\, ,\rangle$ be a symmetric non-degenerate form. If Ψ is an element in $\mathrm{O}(V)$ that commutes with every element of $\mathrm{SO}(V)$. Then Ψ commutes with every element of $\mathrm{O}(V)$.*

In particular we have that $Z(\mathrm{SO}(V)) = Z(\mathrm{O}(V)) \cap \mathrm{SO}(V)$.

Proof: Let x be a vector in V such that $\langle x, x\rangle \neq 0$. It follows from Proposition 1.8.2 that we can find an orthogonal basis $e_1, \dots, e_n$ such that $e_1 = x$.

Let W_1 and W_2 be the spaces generated by e_n, e_1 and e_1, e_2 respectively. Since $n \geq 3$, we have that W_1 and W_2 are different, and we clearly have that $W_1 \cap W_2 = \mathbf{K}e_1 = \mathbf{K}x$. Denote by s_i the reflection s_{e_i} of Definition 1.9.1.

We have that $\Psi(W_i) \subseteq W_i$, for $i = 1, 2$. Indeed, write $s_i = s_{e_i}$ for $i = 1, 2$. We have that $-\Psi(e_1) = \Psi(s_1 s_2(e_1)) = s_1 s_2(\Psi(e_1)) = s_1(\Psi(e_1) - 2\frac{\langle \Psi(e_1), e_2\rangle}{\langle e_2, e_2\rangle} e_2) = \Psi(e_1) - 2\frac{\langle \Psi(e_1), e_1\rangle}{\langle e_1, e_1\rangle} e_1 - 2\frac{\langle \Psi(e_1), e_2\rangle}{\langle e_2, e_2\rangle} e_2$. Consequently, $\Psi(e_1) = \frac{\langle \Psi(e_1), e_1\rangle}{\langle e_1, e_1\rangle} e_1 - \frac{\langle \Psi(e_1), e_2\rangle}{\langle e_2, e_2\rangle} e_2$. Similarly it follows that $\Psi(e_2) \in W_2$. A similar argument, with indices $n, 1$ instead of $1, 2$ gives that $\Psi(W_1) \subseteq W_1$. We obtain that $\Psi(W_1 \cap W_2) \subseteq W_1 \cap W_2$. Consequently we have that $\Psi(x) = ax$, for some $a \in \mathbf{K}$.

Since x was an arbitrary vector with $\langle x, x\rangle \neq 0$, we have that $\Psi(y) = a_y y$, for some element a_y in $\mathbf{K}$ for all y in V such that $\langle y, y\rangle \neq 0$. In particular we have that $\Psi(e_i) = a_i e_i$, for $i = 1, \dots, n$. It is now easy to check that Ψs_x and $s_x \Psi$ take the same value on all the vectors $e_1, \dots, e_n$, and hence $\Psi s_x = s_x \Psi$. It follows from Proposition 1.9.4 that Ψ commutes with all the generators of $\mathrm{O}(V)$, and consequently, with all the elements of $\mathrm{O}(V)$. We have proved the first part of the lemma. The second part follows immediately from the first. □

Proposition 1.10.4. *Let V be a vector space over a field $\mathbf{K}$ with more than 3 elements, where $2 \neq 0$, and let $\langle\, ,\rangle$ be a symmetric non-degenerate form. Then we have that*

(i) $Z(\mathrm{O}(V)) = \{I, -I\}$
(ii) $Z(\mathrm{SO}(V)) = \{I, -I\}$ *if* $\dim_{\mathbf{K}} V > 2$ *and* $\dim_{\mathbf{K}} V$ *is even.*
(iii) $Z(\mathrm{SO}(V)) = \{I\}$ *if* $\dim_{\mathbf{K}} V > 2$ *and* $\dim_{\mathbf{K}} V$ *is odd.*

Proof: Let $n = \dim_{\mathbf{K}} V$ and let Φ be an element in the center of $\mathrm{O}(V)$. It follows from Proposition 1.9.4 that Φ commutes with all reflections of the form s_x, where $\langle x, x\rangle \neq 0$. For all y in V we have that

$$\Phi(y) - 2\frac{\langle y, x\rangle}{\langle x, x\rangle}\Phi(x) = \Phi s_x(y) = s_x \Phi(y) = \Phi(y) - 2\frac{\langle \Phi(y), x\rangle}{\langle x, x\rangle} x.$$

Consequently, we have that $\langle y, x\rangle \Phi(x) = \langle \Phi(y), x\rangle x$. In particular we must have that $\Phi(x) = a_x x$, for some $a_x \in \mathbf{K}$. We get that $a_x^2\langle x, x\rangle = \langle a_x x, a_x x\rangle = \langle \Phi(x), \Phi(x)\rangle = \langle x, x\rangle$. Consequently, we have that $a_x = \pm 1$.

It follows from Proposition 1.8.2 that we have an orthogonal basis $e_1, \dots, e_n$ for $\langle\, ,\rangle$. Then $\Phi(e_i) = a_i e_i$, with $a_i = \pm 1$. We shall show that all the a_i's are equal. To this end we consider $\langle e_i + a e_j, e_i + a e_j\rangle = \langle e_i, e_i\rangle + a^2\langle e_j, e_j\rangle$, for all $a \in \mathbf{K}$. Since $\mathbf{K}$ has more than 3 elements we can find a non-zero element a of $\mathbf{K}$ such that $\langle e_i + a e_j, e_i + a e_j\rangle = \langle e_i, e_i\rangle + a^2\langle e_j, e_j\rangle \neq 0$. We then have that $a_i e_i + a a_j e_j = \Phi(e_i + a e_j) = b(e_i + a e_j)$ for some $b \in \mathbf{K}$. Consequently, we have that $a_i = a_j$, for all i and j, and we have proved the first part of the proposition. The assertions for $\mathrm{SO}(V)$ follow from the first part of the proposition and from Lemma 1.10.3. □

Proposition 1.10.5. *The center of* $\mathrm{Sp}(V)$ *is* $\{I, -I\}$.

Proof: Let Φ be in the center of $\mathrm{Sp}(V)$. It follows from proposition 1.9.9 that Φ commutes with all transvections. Let Ψ be the transvection corresponding to x in V and a in $\mathbf{K}$. Then, for all y in V, we have that $\Phi(y) - a\langle y, x\rangle \Phi(x) = \Phi\Psi(y) = \Psi\Phi(y) = \Phi(y) - \langle \Phi(y), x\rangle x$. Hence $a\langle x, y\rangle \Phi(x) = \langle x, \Phi(y) x\rangle$. Choose y such that $\langle x, y\rangle \neq 0$. Then $\Phi(x) = a_x x$ for some a_x in $\mathbf{K}$. Let z be another vector in V. We obtain, in the same way, that $\Phi(z) = a_z z$ and $\Phi(x+z) = a_{x+z}(x+z)$. Consequently we have that $a_x x + a_z z = \Phi(x+z) = a_{x+z}(x+z)$. Consequently, $a_x = a_z$ and there is an element a in $\mathbf{K}$ such that $\Phi(x) = ax$ for all x in V. Choose y such that $\langle y, x\rangle \neq 0$. We have that $a^2\langle x, y\rangle = \langle ax, ay\rangle = \langle \Phi(x), \Phi(y)\rangle = \langle x, y\rangle$, so that $a = \pm 1$. Hence, we have proved the proposition. □

Example 1.10.6. We have proved the following assertions:

(i) $Z(\mathrm{Gl}_n(\mathbf{C})) \cong \mathbf{C}^* = \mathbf{C} \setminus 0$, for all n.
(ii) $Z(\mathrm{Sl}_n(\mathbf{C})) \cong \mathbf{Z}/n\mathbf{Z}$, for all n (see Example 3.5.2).
(iii) $Z(\mathrm{O}_n(\mathbf{C})) \cong \{\pm 1\}$, for all n.
(iv) $Z(\mathrm{SO}_n(\mathbf{C})) \cong \{\pm 1\}$, when $n \geq 4$ is even.
(v) $Z(\mathrm{SO}_n(\mathbf{C})) \cong \{1\}$, when $n \geq 3$ is odd.
(vi) $Z(\mathrm{Sp}_n(\mathbf{C})) \cong \{\pm 1\}$, for all even n.

Hence $\mathrm{Sl}_n(\mathbf{C})$, for $n > 3$, $\mathrm{SO}_n(\mathbf{C})$, for n odd and $\mathrm{Gl}_n(\mathbf{C})$, are neither isomorphic as groups, nor isomomorphic, as groups to any of the other groups. We can however, not rule out isomorphisms between the remaining groups. The purpose of the next chapter is to give all the groups a geometric structure, and to introduce invariants of this structure that permits us to rule out isomorphisms.

Exercises

1.10.1. Let $\mathbf{K} = \mathbf{F}_3$, i.e., the field with three elements $\{0, 1, 2\}$ where $1+1 = 2$ and $1+1+1 = 0$.

(a) Show that if $\langle\, ,\rangle$ is the form given by $\langle (a_1, b_1), (a_2, b_2)\rangle = a_1 a_2 - b_1 b_2$, we have that $\mathrm{O}(V_{\mathbf{K}}^2, \langle\, ,\rangle)$ consists of 4 elements and is commutative.
(b) Show that if $\langle\, ,\rangle$ is the form given by $\langle (a_1, b_1), (a_2, b_2)\rangle = a_1 a_2 + b_1 b_2$, we have that $\mathrm{O}(V_{\mathbf{K}}^2, \langle\, ,\rangle)$ consists of 8 elements and is non-commutative.

2 The exponential function and the geometry of matrix groups

In Chapter 1 we introduced classical groups over arbitrary fields and we defined algebraic invariants of these matrix groups that made it possible to distinguish many of them. When the groups have coefficients in the real or complex numbers we introduce here a geometric structure on the groups, and define geometric invariants that make it possible to further distinguish the matrix groups.

The geometric structure comes from a norm on all matrices with real or complex coefficients. This norm makes it possible to define and study analytic functions on the matrix groups. In particular we can define curves in the matrix groups, and thus their tangent spaces. The main tool used to introduce a geometric structure on the matrix groups in this chapter is the exponential map. This functions induces an analytic map from the tangent spaces of the *classical groups* introduced in Chapter 1 and the groups themselves. It allows us to describe the tangent spaces as subspaces of all matrices, and we can define and determine the dimension of the matrix groups introduced in Chapter 1 with real or complex coefficients.

2.1 Norms and metrics on matrix groups

Throughout this chapter the field $\mathbf{K}$ will be the real or complex numbers, unless we explicitly state otherwise.

All the matrix groups that we introduced in Chapter 1 were subsets of the $n \times n$ matrices $\mathrm{M}_n(\mathbf{K})$. In this section we shall show how to give $\mathrm{M}_n(\mathbf{K})$ a geometric structure, as a metric space. This structure is inherited by the matrix groups.

Definition 2.1.1. Let $x = (a_1, \dots, a_n)$ be a vector in $V_{\mathbf{K}}^n$. We define the *norm* $\|x\|$ of x by

$$\|x\| = C \max_i |a_i|,$$

where $|a|$ is the usual norm of a in $\mathbf{K}$ and C is some fixed positive real number.

Remark 2.1.2. We have that $V_{\mathbf{K}}^1$ and $\mathbf{K}$ are canonically isomorphic as vector spaces. Under this isomorphism the norm $\| \ \|$ on $V_{\mathbf{K}}^1$ correspond to the norm $| \ |$ on $\mathbf{K}$.

Proposition 2.1.3. *For all vectors x and y of $\mathbf{K}^n$, and elements a of $\mathbf{K}$, the following three properties hold:*

(i) $\|x\| \geq 0$, *and* $\|x\| = 0$ *if and only if* $x = 0$,
(ii) $\|ax\| = |a|\|x\|$,
(iii) $\|x + y\| \leq \|x\| + \|y\|$.

Proof: The properties of the proposition hold for the norm $\| \|$ on $\mathbf{K}$ (see Remark 2.1.2). Consequently, all the properties follow immediately from Definition 2.1.1 of a norm on $V_{\mathbf{K}}^n$. □

Remark 2.1.4. We can consider $\mathrm{M}_n(\mathbf{K})$ as a vector space $V_{\mathbf{K}}^{n^2}$ of dimension n^2, where addition of vectors is the addition of matrices. In the definition of the norm on $\mathrm{M}_n(\mathbf{K})$ we shall choose $C = n$, and in all other cases we choose $C = 1$, unless the opposite is explicitly stated.

Next we shall see how the norm behaves with respect to the product on $\mathrm{M}_n(\mathbf{K})$.

Proposition 2.1.5. *Let X and Y be matrices in $\mathrm{M}_n(\mathbf{K})$. We have that*

$$\|XY\| \leq \|X\|\|Y\|.$$

Proof: Let $X = (a_{ij})$ and $Y = (b_{ij})$. Then we obtain that

$$\begin{aligned}\|XY\| = n \max_{ij}(|\sum_{k=1}^{n} a_{ik}b_{kj}|) &\leq n \max_{ij}(\sum_{k=1}^{n} |a_{ik}||b_{kj}|) \\ &\leq nn \max_{ij}(|a_{ik}||b_{kj}|) \leq n^2 \max_{ij} |a_{ij}| \max_{ij} |b_{ij}| = \|X\|\|Y\|.\end{aligned}$$

□

It is possible to give $V_{\mathbf{K}}^n$ several different, but related, norms (see Exercise 2.1.3). Consequently it is convenient to give a more general definition of a norm, valid for all vector spaces.

Definition 2.1.6. Let V be a vector space. A *norm* on V is a function

$$\|\ \|\colon V \to \mathbf{R},$$

such that for all x and y of V and all a in $\mathbf{K}$ we have that

(i) $\|x\| \geq 0$ and $\|x\| = 0$ if and only if $x = 0$,
(ii) $\|ax\| = |a|\|x\|$,
(iii) $\|x + y\| \leq \|x\| + \|y\|$.

We call the the pair $(V, \|\ \|)$ a *normed space.*

Example 2.1.7. Choose a basis $e = (e_1, \dots, e_n)$ for the vector space V. We obtain a canonical isomorphism

$$\Psi_e\colon V \to V_{\mathbf{K}}^n$$

(see Paragraph 1.6.16). The norm $\|\ \|$ on $V_{\mathbf{K}}^n$ of Definition 2.1.1 induces a norm $\|\ \|_e$ on V by

$$\|x\|_e = \|\Psi_e(x)\|.$$

Choose another basis $f = (f_1, \dots, f_n)$ of V. We get another norm $\|\ \|_f$ of V, which is closely related to $\|\ \|_e$. More precisely, there are two positive constants C_1 and C_2 such that

$$C_2\|x\|_f \leq \|x\|_e \leq C_1\|x\|_f.$$

Indeed, let $f_i = \sum_{j=1}^n a_{ij} e_j$, for $i = 1, \dots, n$. For each vector $x = \sum_{i=1}^n a_i f_i$ of V we obtain that

$$\|x\|_e = \|\sum_{i=1}^n a_i f_i\| = \|\sum_{i=1}^n \sum_{j=1}^n a_i a_{ij} e_j\|_e = \max_j (\|\sum_{i=1}^n a_i a_{ij}\|)$$

$$\le \max_j (\sum_{i=1}^n \|a_i\| \|a_{ij}\|) \le n \max_i (\|a_i\|) \max_{ij} (\|a_{ij}\|) = n\|x\|_f \max_{ij}(\|a_{ij}\|).$$

We can choose $C_1 = n \max(\|a_{ij}\|)$. Similarly, we find C_2.

From a norm on a vector space we can define a distance function on the space.

Definition 2.1.8. Let $(V, \| \|)$ be a normed vector space. Define, for each pair of vectors x and y of V, the *distance* $d(x, y)$ between x and y to be

$$d(x, y) = \|x - y\|.$$

Proposition 2.1.9. *Let $(V, \| \|)$ be a normed vector space. For all vectors x, y and z of V the following three properties hold for the distance function of Definition 2.1.8:*

(i) *$d(x, y) \ge 0$ and $d(x, y) = 0$ if and only if $x = y$,*
(ii) *$d(x, y) = d(y, x)$,*
(iii) *$d(x, z) \le d(x, y) + d(y, z)$.*

Proof: The properties (i) and (ii) follow immediately from properties (i) and (ii) of Definition 2.1.6. For property (iii) we use property (iii) of Definition 2.1.6 to obtain $d(x, z) = \|x - z\| = \|x - y + y - z\| \le \|x - y\| + \|y - z\| = d(x, y) + d(y, z)$. □

Sets with a distance function enjoying the properties of the proposition appear everywhere in mathematics. It is therefore advantageous to axiomatize their properties.

Definition 2.1.10. Let X be a set. A *metric* on X is a function

$$d \colon X \times X \to \mathbf{R}$$

such that, for any triple x, y, z of points of X, we have

(i) $d(x, y) \ge 0$ and $d(x, y) = 0$ if and only if $x = y$,
(ii) $d(x, y) = d(y, x)$,
(iii) $d(x, z) \le d(x, y) + d(y, z)$.

The pair (X, d) is called a *metric space*

Remark 2.1.11. For every subset Y of a metric space (X, d_X), we have a distance function d_Y on Y defined by $d_Y(x, y) = d_X(x, y)$, for all x and y in Y. It is clear that Y, with this distance function, is a metric space. We say that (Y, d_Y) is a *metric subspace* of (X, d_X).

Definition 2.1.12. Let r be a positive real number and x a point in X. A *ball* $B(x, r)$, of radius r with center x, is the set

$$\{y \in X \colon d(x, y) < r\}.$$

We say that a subset U of X is *open* if, for every point x in U, there is a positive real number r such that the ball $B(x,r)$ is contained in U (see Exercise 2.1.4). A subset Y of X is *closed* if its complement $X \setminus Y$ is open. Let x be a point of X, we call an open subset of X that contains x a *neighborhood* of x.

Remark 2.1.13. We have that every ball $B(x,r)$ is open. Indeed, let y be in $B(x,r)$. Put $s = r - d(x,y)$. Then the ball $B(y,s)$ is contained in $B(x,r)$, because, for $z \in B(y,s)$ we have that $d(x,z) \leq d(x,y) + d(y,z) < d(x,y) + s = r$.

The metric on $V_{\mathbf{K}}^n$ is defined by $d((a_1,\dots,a_n),(b_1,\dots,b_n)) = \max_i |a_i - b_i|$. Hence a *ball* $B(x,r)$ with center x and radius r is, in this case, geometrically a cube centered at x and with side length $2r$. We see that, if a subset U of $V_{\mathbf{K}}^n$ is open with respect to the norm given by one constant C, then it is open with respect to the norm defined by all other positive constants. Consequently metrics on a vector space that are associated to the norms on a vector space given by different choices of bases, as in Example 2.1.7, also give the same open sets.

The definition of a continuous map from calculus carries immediately over to metric spaces.

Definition 2.1.14. A map $\Phi\colon X \to Y$, from a metric space (X, d_X) to a metric space (Y, d_Y), is *continuous* if, for each point x of X and any positive real number ε, there is a positive real number δ, such that the image of $B(x,\delta)$ by Φ is contained in $B(\Phi(x),\varepsilon)$. A continuous map between metric spaces that is bijective and whose inverse is also continuous is called a *homeomorphism* of the metric spaces.

We next give a very convenient criterion for a map to be continuous. The criterion easily lends itself to generalizations (see Section 3.3).

Proposition 2.1.15. *Let $\Phi\colon X \to Y$ be a map between metric spaces (X, d_X) and (Y, d_Y). We have that Φ is continuous if and only if, for every open subset V of Y, the inverse image $\Phi^{-1}(V)$ is open in X.*

Proof: Assume first that Φ is continuous. Let V be open in Y. We shall show that $U = \Phi^{-1}(V)$ is open in X. Choose x in U. Since V is open, we can find a positive number ε such that $B(\Phi(x),\varepsilon)$ is in V, and since Φ is continuous, we can find a positive integer δ such that $\Phi(B(x,\delta)) \subseteq B(\Phi(x),\varepsilon)$. That is, the ball $B(x,\delta)$ is contained in U. Consequently, every x in U is contained in a ball in U. Hence U is open.

Conversely, assume that the inverse image by Φ of every open subset of Y is open in X. Let x be in X and let ε be a positive real number. Then $B(\Phi(x),\varepsilon)$ is open in Y. Consequently, the set $U = \Phi^{-1}(B(\Phi(x),\varepsilon)$ is open in X, and U clearly contains x. We can therefore find a positive real number δ such that $B(x,\delta)$ is contained in U. Consequently, we have that $\Phi(B(x,\delta)) \subseteq \Phi(U) = B(\Phi(x),\varepsilon)$. That is, Φ is continuous. □

Remark 2.1.16. Many properties of continuous maps follow directly from Proposition 2.1.15. For example, it is clear that the composite $\Psi\Phi$ of two continuous maps $\Phi\colon X \to Y$ and $\Psi\colon Y \to Z$ of metric spaces is continuous.

Example 2.1.17. The map
$$\det\colon \mathrm{M}_n(\mathbf{K}) \to \mathbf{K}$$
is given by the polynomial
$$\det(x_{ij}) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sign}(\sigma) x_{1\sigma(1)} \cdots x_{n\sigma(n)},$$
of degree n in the variables x_{ij}, where $i, j = 1, \dots, n$. The symbol $\operatorname{sign}\sigma$ is 1 or -1 according to whether σ is an even or odd permutation (see Example 1.2.13). In particular the determinant is a continuous map (see Exercise 2.1.5). For each matrix A in $\mathrm{Gl}_n(\mathbf{K})$ we have that $\det A \neq 0$. Let $\varepsilon = |\det A|$. Then there is a positive real number δ such that the ball $B(A, \delta)$ in $V_{\mathbf{K}}^{n^2}$ maps into the ball $B(\det A, \varepsilon)$ in $\mathbf{K}$. The latter ball does not contain 0. In other words we can find a ball around A that is contained in $\mathrm{Gl}_n(\mathbf{K})$. Hence $\mathrm{Gl}_n(\mathbf{K})$ is an open subset of the space $\mathrm{M}_n(\mathbf{K})$.

Example 2.1.18. We have that the determinant induces a continuous map
$$\det\colon \mathrm{O}_n(\mathbf{K}) \to \{\pm 1\}.$$
The inverse image of 1 by this map is $\mathrm{SO}_n(\mathbf{K})$. Since the point 1 is open in $\{\pm 1\}$ we have that $\mathrm{SO}_n(\mathbf{K})$ is an open subset of $\mathrm{O}_n(\mathbf{K})$.

Exercises

2.1.1. Let X be a set. Define a function
$$d\colon X \times X \to \mathbf{R}$$
by $d(x, y) = 1$ if $x \neq y$ and $d(x, x) = 0$. Show that $(X.d)$ is a metric space, and describe the open sets of X.

2.1.2. Let (X_1, d_1), ..., (X_m, d_m) be metric spaces.

(a) Show that the Cartesian product $X = X_1 \times \cdots \times X_m$ with the function $d\colon X \times X \to \mathbf{R}$ defined by
$$d((x_1, \dots, x_m), (y_1, \dots, y_m)) = d_1(x_1, y_1) + \cdots + d_m(x_m, y_m),$$
is a metric space.

(b) When $X_1 = \cdots = X_m$ and d_i, for $i = 1, \dots, m$ is the metric of Problem 2.1.2, the metric is called the *Hamming metric* on X. Show that $d((x_1, \dots, x_m), (y_1, \dots, y_m))$ is the number of indices i such that $x_i \neq y_i$.

(c) When $X_1 = \cdots = X_m = \mathbf{K}$, where $\mathbf{K}$ is the real or complex numbers, we have that $X = V_{\mathbf{K}}^m$, and we call the metric, the *taxi metric*. Show that the open sets for the taxi metric are the same as the open sets in the metric on $V_{\mathbf{K}}^m$ associated to the norm on $V_{\mathbf{K}}^m$ given in Definition 2.1.8 and Remark 2.1.2.

2.1.3. Let $X = \mathbf{K}^n$ and, for x in X let
$$\|x\| = \sqrt{|x_1|^2 + \cdots + |x_n|^2}.$$
Show that this defines a norm on X.

Hint: Consider the *sesquilinear product*

$$\langle , \rangle \colon \mathbf{C}^n \times \mathbf{C}^n \to \mathbf{C},$$

defined by

$$\langle x, y\rangle = x_1\bar{y}_1 + \cdots + x_n\bar{y}_n,$$

where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ and $\bar{y}_i$ is the complex conjugate of y_i.

For all points x, y and z of $\mathbf{C}^n$ we have that

(a) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$
(b) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
(c) $a\langle x, y\rangle = \langle ax, y\rangle = \langle x, \bar{a}y\rangle$
(d) $\overline{\langle x, y\rangle} = \langle y, x\rangle$.

Then $\|x\| = \sqrt{\langle x, x\rangle}$ and we have *Schwartz inequality*,

$$\|\langle x, y\rangle\| \le \|x\|\|y\|,$$

with equality if and only if $x = ay$ for some $a \in \mathbf{K}$.

In order to prove the Schwartz inequality we square the expression and prove that $\langle x, y\rangle^2 \le \|x\|^2\|y\|^2$. If $\|y\|^2 = 0$, the inequality clearly holds, so you can assume that $\|y\|^2 > 0$. We have that

$$\begin{aligned}
\sum_{j=1}^{n} \left|\|y\|^2 x_j - \langle x, y\rangle y_j\right|^2 &= \sum_{j=1}^{n}(\|y\|^2 x_j - \langle x, y\rangle y_j)(\|y\|^2\bar{x}_j - \overline{\langle x, y\rangle y_j}) \\
&= \|y\|^4\|x\|^2 - \|y\|^2\overline{\langle x, y\rangle}\sum_{j=1}^{n} x_j\bar{y}_j - \|y\|^2\sum_{j=1}^{n}\bar{x}_j y_j + \|\langle x, y\rangle\|^2\|y\|^2 \\
&= \|y\|^4\|x\|^2 - \|y\|^2\|\langle x, y\rangle\|^2 - \|y\|^2\|\langle x, y\rangle\|^2 + \|y\|^2\|\langle x, y\rangle\|^2 \\
&= \|y\|^2(\|y\|^2\|x\|^2 - \|\langle x, y\rangle\|^2).
\end{aligned}$$

Since the first term is nonnegative and $\|y\|^2 > 0$, you obtain the inequality $\|\langle x, y\rangle\|^2 \le \|x\|^2\|y\|^2$. The first term is zero if and only if $x = \frac{\langle x,y\rangle}{\|y\|^2}$, hence equality holds if and only if $x = ay$ for some a.

2.1.4. Let (X, d) be a metric space. Show that the collection $\mathcal{U} = \{U_i\}_{i\in I}$ of open sets satisfies the following three properties:

(a) The empty set and X are in $\mathcal{U}$.
(b) If $\{U_j\}_{j\in J}$ is a collection of sets from $\mathcal{U}$, then the union $\cup_{j\in J} U_j$ is a set in $\mathcal{U}$.
(c) If $\{U_j\}_{j\in K}$ is a finite collection of sets from $\mathcal{U}$, then the intersection $\cap_{j\in K} U_j$ is a set in $\mathcal{U}$.

2.1.5. Show that if f and g are continuous functions $\mathbf{K}^n \to \mathbf{K}$, then cg, $f + g$, and fg are continuous for each c in $\mathbf{K}$. Consequently, all polynomial functions are continuous.

2.1.6. Let f be a non-constant polynomial function $f \colon \mathbf{K}^n \to \mathbf{K}$.

(a) Show that the points where f is zero can not contain a ball $B(a, \varepsilon)$, with $a \in \mathbf{K}^n$ and $\varepsilon > 0$.
(b) Show that if $f(a) \ne 0$, then there is a ball $B(a\varepsilon)$ such that $f(x) \ne 0$ for all x in $B(a, \varepsilon)$.

Hint: Solve the problem for $n = 1$ and use induction on n.

2.1.7. Let $[a, b]$ be a closed interval of the real line, where a or b can be equal to ∞.

(a) Show that the only subsets of $[a,b]$ that are both open and closed are $[a,b]$ and the empty set $\emptyset$.
(b) Show that the only subsets of a ball $B(x,r)$ in $\mathbf{R}^n$ that are both open and closed are $B(x,r)$ and $\emptyset$.

2.2 The exponential map

In Section 2.1 we introduced a norm on the space $\mathrm{M}_n(\mathbf{K})$ which satisfies the inequality of Proposition 2.1.5. To this norm we saw in Definition 2.1.8 that we can associate a metric on $\mathrm{M}_n(\mathbf{K})$. In this section we shall show how this metric allows us to define an exponential and a logarithmic function on $\mathrm{M}_n(\mathbf{K})$. We also show that many of the usual properties of exponential and logarithmic functions hold.

The fundamental notions of calculus carry over to any metric space virtually without any change. Therefore we leave some of the verification to the reader as exercises.

Definition 2.2.1. Let (X,d) be a metric space. A sequence $x_1, x_2, \dots$ of elements in X *converges* to an element x of X if, for every positive real number ε, there is an integer m such that $d(x, x_i) < \varepsilon$, when $i > m$.

A sequence $x_1, x_2, \dots$ is a *Cauchy sequence* if, for every positive real number ε, there is an integer m such that $d(x_i, x_j) < \varepsilon$, when $i, j > m$.

The space X is *complete* if every Cauchy sequence in X converges.

When X is a vector space and the metric comes from a norm, we say that the *series* $x_1 + x_2 + \cdots$ *converges* if the sequence $\{y_n = x_1 + \cdots + x_n\}_{n=1,2,\dots}$ converges.

As in calculus, we have that every convergent sequence in a metric space is a Cauchy sequence (see Exercise 2.2.2).

Proposition 2.2.2. *The space $V_{\mathbf{K}}^n$, with the norm of Definition 2.1.1, is complete.*

Proof: Let $x_i = (a_{i1}, \dots, a_{in})$ be a Cauchy sequence in $V_{\mathbf{K}}^n$. Given ε there is an integer m such that $\|x_i - x_j\| = \max_k |a_{ik} - a_{jk}| < \varepsilon$, when $i, j > m$. Consequently, the sequences $a_{1k}, a_{2k}, \dots$ are Cauchy in $\mathbf{K}$, for $k = 1, \dots, n$. Since the real and complex numbers are complete we have that these sequences converge to elements $a_1, \dots, a_n$. It is clear that $x_1, x_2, \dots$ converges to $x = (a_1, \dots, a_n)$. □

2.2.3. For X in $\mathrm{M}_n(\mathbf{K})$ and $m = 0, 1, \dots$, let $\exp_m(X)$ be the matrix

$$\exp_m(X) = I_n + \frac{1}{1!}X + \frac{1}{2!}X^2 + \cdots + \frac{1}{m!}X^m.$$

The sequence $\{\exp_m(X)\}_{m=0,1,\dots}$ is a Cauchy sequence in $\mathrm{M}_n(\mathbf{K})$ because, for $q > p$, we have that

$$\begin{aligned}\|\exp_q(X) - \exp_p(X)\| &= \|\frac{1}{(p+1)!}X^{p+1} + \cdots + \frac{1}{q!}X^q\| \\ &\le \frac{1}{(p+1)!}\|X^{p+1}\| + \cdots + \frac{1}{q!}\|X^q\| \le \frac{1}{(p+1)!}\|X\|^{p+1} + \cdots + \frac{1}{q!}\|X\|^q,\end{aligned}$$

and the term to the right can be made arbitrary small by choosing p big because the sequence $\{1+\frac{1}{1!}\|X\|+\cdots+\frac{1}{m!}\|X\|^m\}_{m=0,1,\dots}$ converges to $\exp(\|X\|)$, where $\exp(x)$ is the usual exponential function on $\mathbf{K}$.

Definition 2.2.4. For X in $\mathrm{M}_n(\mathbf{K})$ we define $\exp(X)$ to be the limit of the sequence $\exp_0(X), \exp_1(X), \dots$.

Example 2.2.5. Let $X = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Then we have that $X^2 = -I_2$, $X^3 = -X$, $X^4 = I_2$, ... We see that $\exp(yX) = I_2 + \frac{1}{1!}yX + \frac{1}{2!}y^2 I_2 - \frac{1}{3!}y^3 X + \frac{1}{4!}y^4 I_2 + \dots$. Consequently, we have that $\exp(yX) = \left(\begin{smallmatrix} \cos y & \sin y \\ -\sin y & \cos y \end{smallmatrix}\right)$. Let $\Phi\colon \mathbf{C} \to \mathrm{M}_2(\mathbf{R})$ be the map given by $\Phi(x+iy) = \left(\begin{smallmatrix} x & y \\ -y & x \end{smallmatrix}\right)$ (see Example 1.3.12). Then we have that $\Phi(\exp(iy)) = \exp(\Phi(iy))$. In the last formula the exponential function on the left is the usual exponential function for complex numbers and the one to the right the exponential function for matrices.

Remark 2.2.6. The exponential function defines a continuous map $\exp\colon \mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$. Indeed, we have seen that $\|\exp_m(X)\| \leq \exp(\|X\|)$. Let $B(Z,r)$ be a ball in $\mathrm{M}_n(\mathbf{K})$, and choose Y in $\mathrm{M}_n(\mathbf{K})$ such that $\|Z\| + r \leq \|Y\|$. Then, for any X in $B(Z,r)$, we have that $\|X\| \leq \|X - Z\| + \|Z\| \leq r + \|Z\| \leq \|Y\|$. Consequently, we have that $\|\exp_m(X)\| \leq \exp(\|X\|) \leq \exp(\|Y\|)$, for all X in $B(Z,r)$. It follows that the series $\exp_0(X), \exp_1(X), \dots,$ converges uniformly on $B(Z,r)$ (see Exercise 2.2.3 (iii)). The functions $\exp_m\colon \mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$ are given by polynomials, hence they are continuous. Since they converge uniformly their limit exp is therefore continuous on $B(Z,r)$ (see Exercise 2.2.3). Consequently, exp is continuous everywhere. In Section 2.4 we shall show that the exponential function is *analytic*. Hence, in particular, it is differentiable with an analytic derivative.

Example 2.2.7. Let $X = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Then $X^2 = I_2$, $X^3 = -X$, $X^4 = I_2$, ... We obtain, for each t in $\mathbf{K}$ that $\exp tX = I_2 + \frac{1}{1!}tX - \frac{1}{2!}t^2 I_2 - \frac{1}{3!}t^3 X + \cdots$. Consequently, we have that $\exp tX = \left(\begin{smallmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{smallmatrix}\right)$. We see that we get a homomorphism of groups $\mathbf{K} \to \mathrm{SO}_2(\mathbf{K})$, which gives an one to one correspondence between a neighborhood of I_2 in $\mathrm{SO}_2(\mathbf{K})$ and a small neighborhood of 0 in $\mathbf{K}$.

The exponential map in $\mathrm{M}_n(\mathbf{K})$ has the usual properties of the exponential function in $\mathbf{K}$.

Proposition 2.2.8. *For all matrices X and Y of $\mathrm{M}_n(\mathbf{K})$ the following properties hold for the exponential function:*

(i) $\exp(0) = I_n$,
(ii) $\exp(X+Y) = \exp(X)\exp(Y)$, *if* $XY = YX$,
(iii) $\exp(-X)\exp(X) = I_n$. *Consequently* $\exp(X)$ *is in* $\mathrm{Gl}_n(\mathbf{K})$.
(iv) $\exp({}^tX) = {}^t(\exp(X))$,
(v) $\exp(Y^{-1}XY) = Y^{-1}\exp(X)Y$, *for all invertible matrices* Y.
(vi) *If* $X = \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix}$ *is diagonal, then* $\exp(X) = \begin{pmatrix} \exp(a_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp(a_n) \end{pmatrix}$.

Proof: Assertion (i) is obvious. To prove assertion (ii) we observe that, if $XY = YX$, then $(X+Y)^i = \sum_{j=0}^{i} \binom{i}{j} X^j Y^{i-j}$. We obtain that

$$\exp_m(X+Y) = I_n + (X+Y) + \cdots + \sum_{i=0}^{m} \frac{1}{i!(m-i)!} X^i Y^{m-i}.$$

On the other hand, we have that

$$\begin{aligned}\exp_m(X)\exp_m(Y) &= (I_n + \frac{1}{1!}X + \cdots + \frac{1}{m!}X^m)(I_n + \frac{1}{1!}Y + \cdots + \frac{1}{m!}Y^m)\\ &= I_n + \frac{1}{1!}(X+Y) + \cdots + \frac{1}{m!}X^m + \frac{1}{(m-1)!}X^{m-1}Y + \cdots + \frac{1}{m!}Y^m + \frac{1}{m!}g_m(X,Y),\end{aligned}$$

where $g_m(X,Y)$ consists of sums of products of the form $\frac{1}{(p-j)!j!}X^jY^{p-j}$ with $2m \geq p > m$. Consequently $\|g_m(X,Y)\| \leq \|\exp_{2m}(X+Y) - \exp_m(X+Y)\|$. Since the sequence $\{\exp_m(X+Y)\}_{m=0,1,\dots}$ converges to $\exp(X+Y)$, we have that $\{g_m(X,Y)\}_{m=0,1,\dots}$ converges to zero, and we have proved assertion (ii).

Assertion (iii) follows from assertion (i) and assertion (ii) with $Y = -X$. We have that assertion (iv) follows from the formulas ${}^t(X^m) = ({}^tX)^m$, for $m = 1, 2, \dots$, and assertion (v) from the formulas $(Y^{-1}XY)^m = Y^{-1}X^mY$ and $Y^{-1}XY + Y^{-1}ZY = Y^{-1}(X+Z)Y$.

Assertion (vi) follows from the definition of the exponential function. □

Example 2.2.9. Although the exponential function for matrices has features similar to those of the usual exponential function, and, in fact, generalizes the latter, it can look quite different. For example, we have that

$$\exp\begin{pmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & x & y+\frac{1}{2}xz \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix}.$$

For upper triangular matrices with zeroes on the diagonal the exponential function $\exp(X)$ is, indeed, always a polynomial in X (see Exercise 2.2.4).

By direct calculation it is easy to check that, if $X = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ and $Y = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$, then $\exp(X) = \begin{pmatrix} e & e-1 \\ 0 & 1 \end{pmatrix}$, and $\exp(Y) = \begin{pmatrix} e^{-1} & 1-e^{-1} \\ 0 & 1 \end{pmatrix}$. We have that $XY \neq YX$ and $\exp(X+Y) = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, whereas $\exp(X)\exp(Y) = \begin{pmatrix} 1 & 2(e-1) \\ 0 & 1 \end{pmatrix}$.

2.2.10. For A in $\mathrm{M}_n(\mathbf{K})$, and $m = 1, 2, \dots$, let $\log_m(A)$ be the matrix

$$\log_m(A) = (A - I_n) - \frac{1}{2}(A - I_n)^2 + \cdots + (-1)^{m-1}\frac{1}{m}(A - I_n)^m.$$

Assume that $\|A - I\| < 1$. Then the sequence $\log_1(A), \log_2(A), \dots$ is a Cauchy sequence. Indeed, we have that for $q > p$

$$\begin{aligned}\|\log_q(A) - \log_p(A)\| &= \|(-1)^p\frac{1}{p+1}(A - I_n)^{p+1} + \cdots + (-1)^q\frac{1}{q}(A - I_n)^q\|\\ &\leq \frac{1}{p+1}\|A - I_n\|^{p+1} + \cdots + \frac{1}{q}\|(A - I_n)\|^q,\end{aligned}$$

and the term to the right can be made arbitrary small by choosing p big because the sequence $\{\|A-I_n\|+\frac{1}{2}\|A-I_n\|^2+\cdots+\frac{1}{m}\|A-I_n\|^m\}_{m=1,2,\dots}$ converges, when $\|A-I_n\|<1$.

Definition 2.2.11. Let A be a matrix in $\mathrm{M}_n(\mathbf{K})$. We define $\log(A)$ to be the limit of the sequence $\log_1(A), \log_2(A), \dots$, when the sequence converges in $\mathrm{M}_n(\mathbf{K})$.

Proposition 2.2.12. *For all matrices A and B in $\mathrm{M}_n(\mathbf{K})$ the following properties hold for the logarithmic function:*

(i) $\log(I_n)=0$,
(ii) $\log({}^tA)={}^t(\log(A))$,
(iii) *We have that $\log(A)$ is defined if and only if $\log(B^{-1}AB)$ is defined, where B is invertible, and we have that $\log(B^{-1}AB)=B^{-1}\log(A)B$.*
(iv) *If $A=\begin{pmatrix} a_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & a_n\end{pmatrix}$ is diagonal, then $\log(A)=\begin{pmatrix} \log(a_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \log(a_n)\end{pmatrix}$.*
(v) $\log(A)\log(B)=\log(B)\log(A)$, *when $AB=BA$, and log(A), $\log(B)$ and $\log(AB)$ are defined.*

Proof: All the assertions are easily proved by methods similar to those used in the proof of Proposition 2.2.8. For the last assertion we note that when $AB=BA$ the partial sums $\log_m(A)\log_m(B)$ and $\log_m(B)\log_m(A)$ are actually equal. □

Remark 2.2.13. The logarithm defines a continuous map $\log\colon B(I_n,1)\to \mathrm{M}_n(\mathbf{K})$. This follows from the inequality $\|\frac{1}{m}X^m\|\leq\frac{1}{m}\|X\|^m$, since the sequence $\log(1-x)=-(x+\frac{1}{2}x^2+\cdots)$ converges for $|x|<1$ (see Exercise 2.2.3). In Section 2.4 we shall show that the logarithmic function is *analytic*. Hence, in particular, it is differentiable with an analytic derivative.

Exercises

2.2.1. Determine the matrices $\exp\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$, $\exp\left(\begin{smallmatrix}1&1\\4&1\end{smallmatrix}\right)$, and $\exp\left(\begin{smallmatrix}1&1\\-1&3\end{smallmatrix}\right)$.

2.2.2. Show that in a metric space every convergent sequence is a Cauchy sequence.

2.2.3. Let X be a set, S a subset, and (Y,d_Y) a metric space. A sequence $f_0, f_1, \dots$ of functions $f_m\colon S\to Y$ *converges uniformly* to a function $f\colon S\to Y$ if, for every positive real number ε there is an integer m such that $d_Y(f(x),f_p(x))<\varepsilon$, for $p>m$ and all $x\in S$. A sequence $f_0,f_1,\dots$ satisfies the *Cauchy criterion* if, for every positive real number ε, there is an integer m such that $d_Y(f_p(x),f_q(x))<\varepsilon$, for $p,q>m$, and all x in S.

(a) Show that a sequence $f_0,f_1,\dots$ of functions $f_m\colon S\to Y$ that converges to a function $f\colon S\to Y$, satisfy the Cauchy criterion.
(b) Assume that (Y,d_Y) is complete. Show that a sequence $f_0,f_1,\dots$ of functions $f_m\colon S\to Y$ that satisfies the Cauchy criterion converges to a function $f\colon S\to Y$.
(c) Let $f_0,f_1,\dots$ be a sequence of functions $f_m\colon S\to Y$ such that $\|f_m(x)\|\leq a_m$, for $m=0,1,\dots$, where $\sum_{m=0}^{\infty}a_m$ is a convergent sequence. Show that the sequence $\{s_m(x)=f_0(x)+\cdots+f_m(x)\}_{m=0,1,\dots}$ converges uniformly.

(d) Let (X, d_X) be a metric space and let $f_0, f_1, \ldots$ be a sequence of continuous functions $f_m \colon X \to Y$. If the sequence converges uniformly to a function $f \colon X \to Y$, then f is continuous.

2.2.4. Let X be an upper triangular matrix with zeroes on the diagonal. Show that the equality $\exp(X) = I_n + \frac{1}{1!}X + \cdots + \frac{1}{(n-1)!}X^{n-1}$ holds.

2.2.5. Let V be a finite dimensional vector space over $\mathbf{K}$. Moreover let $\langle\, , \rangle$ be a non degenerate bilinear form on V.

(a) Show that the exponential map is well defined from $\mathrm{Hom}_{\mathbf{K}}(V, V)$ to $\mathrm{Gl}(V)$.
(b) Assume that $\langle\, , \rangle$ is symmetric, and let x in V be a vector with $\langle x, x\rangle \neq 0$. Determine $\exp(s_x)$.
(c) Assume that $\langle\, , \rangle$ is alternating. Determine $\exp(\Phi)$, where Φ is the transvection given by $\Phi(y) = y - a\langle y, x\rangle x$.

2.3 Diagonalization of matrices and the exponential and logarithmic functions

Most of the properties of the usual exponential and logarithmic functions hold for the more general functions on matrices. These properties can be proved by similar, but more complicated, methods to those used in calculus courses. The formal manipulations of series can however be quite complicated. Instead of mimicing the techniques from elementary calculus we choose to deduce the properties of the exponential and logarithmic functions for matrices from those of calculus by geometric methods. The idea is that such deductions are immediate for diagonalizable matrices and that, since the functions are continuous, there are *sufficiently many* diagonalizable matrices for the properties to hold for all matrices. From properties (v) and (vi) of Proposition 2.2.8 and properties (iii) and (iv) of Proposition 2.2.12 it follows that the properties of the usual exponential and logarithmic functions are inherited by the exponential and logarithmic functions on the *diagonalizable* matrices. Then we use that the exponential and logarithmic functions are continuous and that there are diagonalizable matrices *sufficiently near* all matrices, to deduce the desired properties for all matrices.

Definition 2.3.1. A subset S of a metric space (X, d) is *dense*, if every ball $B(x, \varepsilon)$ in X contains an element in S. Equivalently, S is dense if every nonempty open set in X contains an element of S.

Lemma 2.3.2. *Let (X, d_X) and (Y, d_Y) be metric spaces, and T a dense subset of X. Moreover, let f and g be continuous functions from X to Y. If $f(x) = g(x)$ for all x in T, then $f(x) = g(x)$, for all x in X.*

Proof: Assume that the lemma does not hold. Then there is a point x in X such that $f(x) \neq g(x)$. Let $\varepsilon = d_Y(f(x), g(x))$. The balls $B_1 = B(f(x), \frac{\varepsilon}{2})$ and $B_2 = B(g(x), \frac{\varepsilon}{2})$ do not intersect, and the sets $U_1 = f^{-1}(B_1)$ and $U_2 = g^{-1}(B_2)$ are open in X and contain x. Since T is dense we have a point y in T contained in $U_1 \cap U_2$. We have that $z = f(y) = g(y)$ and consequently, z is contained in both B_1 and B_2. This is impossible since B_1 and B_2

are disjoint. Consequently there is no point x such that $f(x) \neq g(x)$, and we have proved the lemma. □

Definition 2.3.3. We say that a matrix X in $\mathrm{M}_n(\mathbf{K})$ is *diagonalizable* if there is an invertible matrix B such that $B^{-1}XB$ is diagonal.

Proposition 2.3.4. *Let X be a matrix in $\mathrm{M}_n(\mathbf{C})$. Then there exists a matrix Y, complex numbers d_i and e_i, for $i = 1, \dots, n$, with the e_i all different, and a real positive number ε such that, for all nonzero $t \in \mathbf{C}$ with $|t| < \varepsilon$, we have that $X + tY$ is diagonalizable and with diagonal matrix whose (i,i)'th coordinate is $d_i + te_i$, for $i = 1, \dots, n$.*

Proof: The proposition clearly holds when $n = 1$. We shall proceed by induction on n. Assume that the proposition holds for $n-1$. Choose an eigenvalue d_1 of X and a nonzero eigenvector x_1 for d_1. That is, we have $Xx_1 = d_1x_1$. It follows from Theorem 1.6.9 that we can choose a basis $x_1, \dots, x_n$ of $V^n_{\mathbf{K}}$. With respect to this basis the matrix X takes the form $X = \left(\begin{smallmatrix} d_1 & a \\ 0 & X_1 \end{smallmatrix}\right)$, where $a = (a_{12}, \dots, a_{1n})$ and where X_1 is an $(n-1) \times (n-1)$ matrix. By the induction hypothesis there is a matrix Y_1, elements d_i and e_i of $\mathbf{C}$, for $i = 2, \dots, n$, where the e_i are all different, and an ε_1 such that, for all nonzero $|t| < \varepsilon_1$ there is an invertible matrix $C_1(t)$ such that $X_1 + tY_1 = C_1(t)D_1(t)C_1(t)^{-1}$, where $D_1(t)$ is the $(n-1) \times (n-1)$ diagonal matrix with $(i-1, i-1)$'th entry $d_i + te_i$ for $i = 2, \dots, n$. The equality can also be written

$$(X_1 + tY_1)C_1(t) = C_1(t)D_1(t). \tag{2.3.4.1}$$

Let

$$X = \begin{pmatrix} d_1 & a \\ 0 & X_1 \end{pmatrix}, Y = \begin{pmatrix} e_1 & 0 \\ 0 & Y_1 \end{pmatrix}, C(t) = \begin{pmatrix} 1 & c(t) \\ 0 & C_1(t) \end{pmatrix}, D(t) = \begin{pmatrix} d_1+te_1 & 0 \\ 0 & D_1(t) \end{pmatrix},$$

where $c(t) = (c_{12}(t), \dots, c_{1n}(t))$, for some elements $c_{1i}(t)$ of $\mathbf{C}$. Note that $\det C(t) = \det C_1(t)$ for all $t \neq 0$ such that $|t| < \varepsilon_1$, so that the matrix $C(t)$ is invertible for any choice of the elements $c_{1i}(t)$ of $\mathbf{C}$. Let $X(t) = X + tY$. We shall determine the numbers $c_{1i}(t)$ and e_1 such that the equation

$$X(t)C(t) = C(t)D(t), \tag{2.3.4.2}$$

holds. We have that

$$X(t)C(t) = \begin{pmatrix} d_1+te_1 & a \\ 0 & X_1+tY_1 \end{pmatrix} \begin{pmatrix} 1 & c(t) \\ 0 & C_1(t) \end{pmatrix} = \begin{pmatrix} d_1+te_1 & (d_1+te_1)c(t)+a'(t) \\ 0 & (X_1+tY_1)C_1(t) \end{pmatrix},$$

where $c_{ij}(t)$ for $i = 2, \dots, n$ and $j = 1, \dots, n$ are the coordinates of $C_1(t)$ and where $a'(t) = (\sum_{i=2}^n a_{1i}c_{i2}(t), \dots, \sum_{i=2}^n a_{1i}c_{in}(t))$. On the other hand we have that

$$C(t)D(t) = \begin{pmatrix} 1 & c(t) \\ 0 & C_1(t) \end{pmatrix} \begin{pmatrix} d_1+te_1 & 0 \\ 0 & D_1(t) \end{pmatrix} = \begin{pmatrix} d_1+te_1 & c'(t) \\ 0 & C_1(t)D_1(t) \end{pmatrix},$$

where $c'(t) = ((d_2 + te_2)c_{12}(t), \dots, (d_n + te_n)c_{1n}(t))$. Since the Equation 2.3.4.1 holds the Equality 2.3.4.2 holds exactly when

$$(d_1 + te_1)c_{1i}(t) + a_{12}c_{2i}(t) + \cdots + a_{1n}c_{ni}(t) = (d_i + te_i)c_{1i}(t), \tag{2.3.4.3}$$

for $i = 2, \dots, n$. Choose e_1 different from all the $e_2, \dots, e_n$. Then each equation $d_1 + te_1 = d_i + te_i$ has exactly one solution $t = -(d_i - d_1)/(e_i - e_1)$, and we can choose an $\varepsilon < \varepsilon_1$ such that for a nonzero t with $|t| < \varepsilon$ we have that $(d_i - d_1) + t(e_i - e_1) \neq 0$. Then

$$c_{1i}(t) = \frac{1}{(d_i - d_1) + t(e_i - e_1)}(a_{12}c_{2i}(t) + \cdots + a_{1n}c_{ni}(t)), \quad \text{for } i = 2, \dots n$$

solve the equations 2.3.4.3, and we have proved the proposition. □

Corollary 2.3.5. *The subset of* $\mathrm{M}_n(\mathbf{C})$ *consisting of diagonalizable matrices is dense in* $\mathrm{M}_n(\mathbf{C})$.

Proof: Let X be a matrix of $\mathrm{M}_n(\mathbf{C})$. If follows from the proposition that we can find diagonalizable matrices $X + tY$ for sufficiently small nonzero t. We have that $\|X + tY - X\| = |t|\|Y\|$. Consequently we can find diagonalizable matrices in every ball with center X. □

Theorem 2.3.6. *Let* U *be the ball* $B(I_n, 1)$ *in* $\mathrm{Gl}_n(\mathbf{K})$ *and let* $V = \log(U)$*. The following five properties hold:*

(i) $\log\exp X = X$, *for all* $X \in \mathrm{M}_n(\mathbf{K})$ *such that* $\log\exp X$ *is defined.*
(ii) $\exp\log A = A$, *for all* $A \in \mathrm{Gl}_n(\mathbf{K})$ *such that* $\log A$ *is defined.*
(iii) $\det\exp X = \exp\operatorname{tr} X$, *for all* $X \in \mathrm{M}_n(\mathbf{K})$, *where* $\operatorname{tr}(a_{ij}) = \sum_{i=1}^n a_{ii}$.
(iv) *The exponential map* $\exp\colon \mathrm{M}_n(\mathbf{K}) \to \mathrm{Gl}_n(\mathbf{K})$ *induces a homeomorphism* $V \to U$*. The inverse map is* $\log|_U$.
(v) $\log(AB) = \log A + \log B$, *for all matrices* A *and* B *in* U *such that* $AB \in U$, *and such that* $AB = BA$.

Proof: Since the functions exp and log map real matrices to real matrices it suffices to prove the theorem when $\mathbf{K} = \mathbf{C}$.

To prove assertion (i) we first note that by Remarks 2.2.6 and 2.2.13 we have that $\log\exp X$ and X are continuous maps from V to $\mathrm{M}_n(\mathbf{C})$. It follows from Proposition 2.3.4 that the diagonalizable matrices are dense in V. Consequently, it follows from Lemma 2.3.2 that it suffices to prove the assertion when X is a diagonalizable matrix. From Proposition 2.2.8 (v) and Proposition 2.2.12 (iii) it follows that $Y^{-1}(\log\exp X)Y = \log(Y^{-1}(\exp X)Y) = \log\exp(Y^{-1}XY)$. Consequently it suffices to prove assertion (i) for diagonal matrices. It follows from 2.2.8 (vi) and 2.2.12 (iv) that

$$\log\exp\begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix} = \log\begin{pmatrix} \exp a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp a_n \end{pmatrix} = \begin{pmatrix} \log\exp a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \log\exp a_n \end{pmatrix} = \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix}.$$

Hence we have proved the first assertion.

To prove assertion (ii) we use that $\exp\log A$ and A are continuous functions from U to $\mathrm{M}_n(\mathbf{C})$. Reasoning as in the proof of assertion (i) we see that it suffices to prove assertion (ii) for diagonal matrices. The verification of the assertion for diagonal matrices is similar to the one we used in the proof for diagonal matrices in assertion (i).

To prove assertion (iii) we use that $\det\exp X$ and $\exp\operatorname{tr} X$ are continuous functions from $\mathrm{M}_n(\mathbf{C})$ to $\mathrm{M}_n(\mathbf{C})$. We have that $\det(Y^{-1}XY) = \det X$ and $\operatorname{tr}(Y^{-1}XY) = \operatorname{tr} X$, for all invertible Y (see Exercise 2.3.1). It follows, as in the proofs of assertions (i) and (ii) that it suffices to prove assertion (iii) for diagonal matrices. However,

$$\det\exp\begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix} = \det\begin{pmatrix} \exp a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp a_n \end{pmatrix} = \exp(a_1)\cdots\exp(a_n),$$

and

$$\exp\operatorname{tr}\begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix} = \exp(a_1+\cdots+a_n) = \exp(a_1)\cdots\exp(a_n).$$

Hence we have proved assertion (iii).

Assertion (iv) follows from assertions (i) and (ii) since exp and log are continuous.

Finally, to prove assertion (v), we let A and B be elements in in U, such that AB is in U. It follows from assertion (iv) that we can find X and Y in $\mathrm{M}_n(\mathbf{K})$ such that $A = \exp X$, and $B = \exp Y$. Consequently it follows from assertion (iv) that $X = \log A$ and $Y = \log B$. From Proposition 2.2.12 (v) it follows that $XY = \log A \log B = \log B \log A = YX$. Consequently it follows from Proposition 2.2.8 (ii) that $\exp(X+Y) = \exp(X)\exp(Y)$. Hence it follows from assertion (i) that $\log(AB) = \log(\exp X \exp Y) = \log(\exp(X+Y)) = X + Y = \log A + \log B$, and we have proved assertion (v). □

Part (iv) of Theorem 2.3.6 is a particular case of a much more general result that we shall prove in Chapter 3. We shall next show that a similar assertion to Theorem 2.3.6 (iv) holds for the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, and $\mathrm{G}_S(\mathbf{K})$, when S is invertible. First we shall introduce the relevant subspaces of $\mathrm{M}_n(\mathbf{K})$.

Definition 2.3.7. Let $\mathfrak{gl}_n(\mathbf{K}) = \mathrm{M}_n(\mathbf{K})$. We let

$$\mathfrak{sl}_n(\mathbf{K}) = \{X \in \mathfrak{gl}_n(\mathbf{K}) | \operatorname{tr} X = 0\},$$

where as usual the *trace* $\operatorname{tr} X$ of a matrix $X = (a_{ij})$ is defined by $\operatorname{tr} X = \sum_{i=1}^n a_{ii}$.

Let S be a matrix in $\mathrm{M}_n(\mathbf{K})$. We let

$$\mathfrak{g}_S(\mathbf{K}) = \{X \in \mathfrak{gl}_n(\mathbf{K}) | {}^tXS + SX = 0\}.$$

In the special cases when $S = I_n$, or S is the matrix of Display 1.4.1.1 we denote $\mathfrak{g}_S(\mathbf{K})$ by $\mathfrak{so}_n(\mathbf{K})$ respectively $\mathfrak{sp}_n(\mathbf{K})$.

Remark 2.3.8. All the sets $\mathfrak{sl}_n(\mathbf{K})$ and $\mathfrak{g}_S(\mathbf{K})$ are subspaces of $\mathfrak{gl}_n(\mathbf{K})$. In particular they are metric spaces with the metric induced by the metric on $\mathfrak{gl}_n(\mathbf{K})$. We also note that $\mathfrak{so}_n(\mathbf{K})$ is a subspace of $\mathfrak{sl}_n(\mathbf{K})$ because $\operatorname{tr}{}^tX = \operatorname{tr} X$, so that $2\operatorname{tr} X = 0$, and we always assume that 2 is invertible in $\mathbf{K}$ when we treat the orthogonal groups.

Proposition 2.3.9. *Assume that S is invertible. We have that the exponential map*

$$\exp\colon\ \mathrm{M}_n(\mathbf{K}) \to \mathrm{Gl}_n(\mathbf{K})$$

induces maps

$$\exp\colon \mathfrak{sl}_n(\mathbf{K}) \to \mathrm{Sl}_n(\mathbf{K}),$$

and

$$\exp\colon \mathfrak{g}_S \to \mathrm{G}_S(\mathbf{K}).$$

Let G be any of the groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$ or $\mathrm{G}_S(\mathbf{K})$. Then there is a neighborhood U of I_n in G, on which log *is defined, and such that* exp *induces an homeomorphism* $\log(U) \to U$. *The inverse of* $\exp|_{\log(U)}$ *is given by* $\log|_U$.

In particular we have maps

$$\exp\colon \mathfrak{so}_n \to \mathrm{O}_n(\mathbf{K}),$$

$$\exp\colon \mathfrak{so}_n \to \mathrm{SO}_n(\mathbf{K}),$$

and

$$\exp\colon \mathfrak{sp}_n(\mathbf{K}) \to \mathrm{Sp}_n(\mathbf{K}),$$

and if G *is one of the groups* $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$ *or* $\mathrm{Sp}_n(\mathbf{K})$, *there is an open subset U of* G, *such that these maps induce a homeomorphism* $\log(U) \to U$ *with inverse* $\log|_{\log(U)}$.

Proof: We have already proved the assertions of the proposition for $\mathrm{Gl}_n(\mathbf{K})$. To prove them for $\mathrm{Sl}_n(\mathbf{K})$ we take X in $\mathfrak{sl}_n$. It follows from assertion (iii) Theorem 2.3.6 that $\det\exp X = \exp\operatorname{tr} X = \exp 0 = 1$. Consequently, when $\mathbf{K} = \mathbf{R}$, we have that $\exp X$ is in $\mathrm{Sl}_n(\mathbf{K})$, as asserted.

To prove the second assertion about $\mathrm{Sl}_n(\mathbf{K})$ we take A in $U \cap \mathrm{Sl}_n(\mathbf{K})$. It follows from assertion (iii) of Theorem 2.3.6 that, when $\det A = 1$, we have that $\exp\operatorname{tr}\log A = \det\exp\log A = \det A = 1$. Consequently, when $\mathbf{K} = \mathbf{R}$, we have that $\operatorname{tr}\log A = 0$.

When $\mathbf{K} = \mathbf{C}$ we have to shrink U in order to make sure that $|\operatorname{tr}\log A| < 2\pi$. Then we have that $\operatorname{tr}\log A = 0$. We have shown that exp induces a bijective map $\mathfrak{sl}_n(\mathbf{K}) \cap \log(U) \to \mathrm{Sl}_n(\mathbf{K}) \cap U$. Both this map and its inverse, induced by the logarithm , are induced by continuous maps, and the metric on $\mathrm{Sl}_n(\mathbf{K})$ and $\mathfrak{sl}_n(\mathbf{K})$ are induced by the metrics on $\mathrm{Gl}_n(\mathbf{K})$ respectively $\mathfrak{gl}_n(\mathbf{K})$. Consequently, the induced map and its inverse are continuous and therefore homeomorphisms.

Let X be in $\mathfrak{g}_S(\mathbf{K})$. That is, we have ${}^tXS + SX = 0$. Since S is assumed to be invertible, the latter equation can be written $S^{-1t}XS + X = 0$. Since $S^{-1t}XS = -X$ we have that $S^{-1t}XS$ and X commute. Hence we can use assertions (ii) and (v) of Proposition 2.2.8 to obtain equalities $I_n = \exp(S^{-1t}XS + X) = \exp(S^{-1t}XS)\exp X = S^{-1}(\exp{}^tX)S\exp X = S^{-1}({}^t\!\exp X)S\exp X$. Consequently $({}^t\!\exp X)S\exp X = S$, and $\exp X$ is in $\mathrm{G}_S(\mathbf{K})$, as asserted.

To prove the second assertion about $\mathrm{G}_S(\mathbf{K})$ we take A in $\mathrm{G}_S(\mathbf{K})$. Then ${}^tASA = S$ or equivalently, $S^{-1t}ASA = I_n$. Consequently we have that $\log((S^{-1t}AS)A) = 0$. We have that $S^{-1t}AS$ is the inverse matrix of A, and hence that $S^{-1t}AS$ and A commute. Since the map of $\mathrm{Gl}_n(\mathbf{K})$ that sends A to ${}^tA$ is continuous, as is the map that sends A to $S^{-1}AS$, we can choose U such that log is defined on $S^{-1}({}^tA)S$ for all A in U. It follows from assertion (v) of Theorem 2.3.6 that $\log((S^{-1t}AS)A) = \log(S^{-1t}AS) + \log A = S^{-1}(\log{}^tA)S + \log A = S^{-1}({}^t\!\log A)S + \log A$. We have proved that $S^{-1}({}^t\!\log A)S + \log A = 0$. Multiply to the left with S. We get ${}^t\!\log AS + S\log A = 0$. That is $\log A$ is in $\mathfrak{g}_S(\mathbf{K})$. We have proved that exp induces a bijective map $\mathfrak{g}_S(\mathbf{K}) \cap \exp^{-1}(U) \to \mathrm{G}_S(\mathbf{K}) \cap U$. A similar argument to that used for $\mathrm{Sl}_n(\mathbf{K})$ proves that this bijection is a homeomorphism. Hence we have proved the second assertion of the proposition for $\mathrm{G}_S(\mathbf{K})$.

All that remains is to note that $\mathrm{SO}_n(\mathbf{K})$ is an open subset of $\mathrm{O}_n(\mathbf{K})$ containing I_n. $\square$

Exercises

2.3.1. Show that for all matrices X and Y in $\mathrm{M}_n(\mathbf{K})$, with Y invertible, the formula $\mathrm{tr}(Y^{-1}XY) = \mathrm{tr}\, X$ holds.

2.3.2. Use the fact that diagonalizable matrices are dense in $\mathrm{M}_n(\mathbf{C})$ to prove the *Cayley-Hamilton Theorem* for real and complex matrices.

2.3.3. Let $X = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

(a) For every positive real number ε, find a matrix Y in $\mathrm{M}_n(\mathbf{C})$ such that Y is diagonalizable and $\|Y - X\| < \varepsilon$.

(b) Can you find a matrix Y in $\mathrm{M}_n(\mathbf{R})$ which is diagonalizable, and such that $\|Y - X\| < \varepsilon$, when ε is some small positive real number?

2.4 Analytic functions

We shall, in this section, introduce analytic functions and study their basic properties. The exponential and logarithmic functions are analytic and we shall see how the properties of the matrix groups that were discussed in Section 2.2 can then be reinterpreted as asserting that the matrix groups are analytic manifolds.

2.4.1. Let $\mathcal{I}$ be the set of n-tuples $i = (i_1, \dots, i_n)$ of nonnegative integers i_k and let $\mathcal{R}$ denote the set of n-tuples $r = (r_1, \dots, r_n)$ of positive real numbers r_i. For each $r = (r_1, \dots, r_n)$ and each n-tuple of variables $x = (x_1, \dots, x_n)$, we write $r^i = r_1^{i_1} \cdots r_n^{i_n}$ and $x^i = x_1^{i_1} \cdots x_n^{i_n}$. Moreover, we write $|i| = i_1 + \cdots + i_n$ and for j in $\mathcal{I}$ we write $\binom{j}{i} = \binom{j_1}{i_1} \cdots \binom{j_n}{i_n}$.

For any two n-tuples of real numbers $r = (r_1, \dots, r_n)$ and $s = (s_1, \dots, s_n)$, we write $r < s$ if $r_i < s_i$ for $i = 1, \dots, n$ and for $x = (x_1, x_2, \dots, x_n)$ in $\mathbf{K}^n$, we denote by $|x|$ the n-tuple $(|x_1|, |x_2|, \dots, |x_n|)$. The difference between the notation $|x|$ and $|i|$ will be clear from the context.

We shall, in the following, use *polydiscs* instead of balls. As we shall see, these are equivalent as far as topological and metric properties, like openness and analyticity, is concerned. However, for analytic functions polydiscs are notationally more convenient than balls.

Definition 2.4.2. Let r be in $\mathcal{R}$ and x in $\mathbf{K}^n$. The *open polydisc* around x with radius r is the set

$$P(x, r) = \{y \in \mathbf{K}^n \colon |x - y| < r\}.$$

Remark 2.4.3. Let $P(x, r)$ be a polydisc, and let $\epsilon = \min_i r_i$. Then we have that $B(x, \epsilon) \subseteq P(x, r)$, where $B(x, \epsilon)$ is the ball in $\mathbf{K}^n$ with respect to the norm of Definition 2.1.1, with $C = 1$. Conversely, given a ball $B(x, \epsilon)$ we have that $P(x, r) \subseteq B(x, \epsilon)$, with $r = (\epsilon, \dots, \epsilon)$. It follows that every polydisc is open in $\mathbf{K}^n$, and conversely that every ball can be covered by polydiscs. Hence a set is open if and only if it can be covered by polydiscs.

Definition 2.4.4. We say that a formal power series (see Exercise 1.3.4 and Example 1.3.7)

$$\sum_{i \in \mathcal{I}} c_i x^i,$$

with coefficients in $\mathbf{K}$ *converges* in the polydisc $P(0, r)$ if the sequence $s_1, s_2, \dots$, where

$$s_m = \sum_{|i| \le m} |c_i| r'^i,$$

converges for all $r' < r$. It follows that $s_1(x), s_2(x), \dots$ with $s_n(x) = \sum_{|i| \le n} c_i x^i$ converges uniformly in $P(0, r')$ (see Exercise 2.2.3). In particular the series defines a continuous function

$$f(x) = \sum_{i \in \mathcal{I}} c_i x^i$$

in $P(0, r)$.

2.4.5. We note that the function $f(x)$ is zero for all x where it is defined, if and only if $c_i = 0$, for all $i \in \mathcal{I}$. Indeed, this is clear for $n = 1$ and follows in the general case by induction on n.

Let $r' < r$ and let $C = \sum_{i \in \mathcal{I}} |c_i| r'^i$. Then

$$|c_i r'^i| \le C \quad \text{for all } i \in \mathcal{I}.$$

Conversely, given a formal power series $\sum_{i \in \mathcal{I}} c_i x^i$, such that

$$|c_i| r^i \le C,$$

for some C, then $s_1(x), s_2(x), \dots$ with $s_m(x) = \sum_{|i| \le m} c_i x^i$ converges uniformly in $P(0, r')$ for all $r' < r$. In particular $\sum_{i \in \mathcal{I}} c_i x^i$ converges in $P(0, r)$. Indeed, we have that

$$\sum_{i \in \mathcal{I}} |c_i| r'^i = \sum_{i \in \mathcal{I}} |c_i| r^i \frac{r'^i}{r^i} \le C \sum_{i \in \mathcal{I}} \frac{r'^i}{r^i} = C \prod_{i=1}^{n} (1 - \frac{r'_i}{r_i})^{-1}.$$

Definition 2.4.6. Let U be an open subset of $\mathbf{K}^n$. A function

$$g \colon U \to \mathbf{K}$$

is *analytic* in U if, for each x in U, there is an r in $\mathcal{R}$ and a formal power series $f(x) = \sum_{i \in \mathcal{I}} c_i x^i$ which is convergent in $P(0, r)$, such that

$$g(x + h) = f(h) \quad \text{for all } h \in P(0, r) \text{ such that } x + h \in U.$$

A function

$$g = (g_1, \dots, g_m) \colon U \to \mathbf{K}^m$$

is *analytic*, if all the functions g_i are analytic.

Example 2.4.7. All maps *polynomial maps* $\Phi \colon \mathbf{K}^n \to \mathbf{K}^m$, that is, maps given by $\Phi(x) = (f_1(x), \dots, f_m(x))$, where the f_i are polynomials in n variables, are analytic.

Example 2.4.8. It follows from the estimates of of Paragraphs 2.2.3 and 2.2.10, and the definitions of the exponential and logarithmic functions in Definitions 2.2.4 and 2.2.11, that the exponential and logarithmic functions are analytic.

Proposition 2.4.9. *Let $f(x) = \sum_{i\in\mathcal{I}} c_i x^i$ be a formal power series which is convergent in $P(0,r)$. We have that*

(i) $D^i f = \sum_{j\geq i} c_j \binom{j}{i} x^{j-i}$ *is convergent in* $P(0,r)$.

(ii) *For x in $P(0,r)$ the series $\sum_{i\in\mathcal{I}} D^i f(x)h^i$ converges for h in $P(0, r-|x|)$.*

(iii) *We have that*

$$f(x+h) = \sum_{i\in\mathcal{I}} D^i f(x)h^i \quad \text{for } h \in P(0, r-|x|).$$

In particular we have that f is analytic in $P(0,r)$.

Proof: Let $x \in P(0,r)$. Choose an r' such that $|x| \leq r' < r$ and let $s = r - r'$. We have that

$$(x+h)^j = \sum_{i\leq j} \binom{j}{i} x^{j-i}h^i.$$

Hence, we obtain that

$$f(x+h) = \sum_{j\in\mathcal{I}} c_j \left(\sum_{i\leq j} \binom{j}{i} x^{j-i}h^i \right) \quad \text{for } h \in P(0,s).$$

For $|h| \leq s' < s$ we have that

$$\sum_{j\in\mathcal{I}}\sum_{i\leq j} |c_j \binom{j}{i} x^{j-i}h^i| \leq \sum_{j\in\mathcal{I}}\sum_{i\leq j} |c_j| \binom{j}{i} r'^{j-i} s'^i = \sum_{j\in\mathcal{I}} |c_j|(r'+s')^j < \infty. \tag{2.4.9.1}$$

The last inequality of Formula 2.4.9.1 holds since f converges in $P(0,r)$ and $r'+s' < r$. Assertions (i) and (ii) follow from the inequality 2.4.9.1. Moreover, it follows from inequality 2.4.9.1 that we can rearrange the sum in the above expression for $f(x+h)$. Consequently

$$f(x+h) = \sum_{i\in\mathcal{I}} \left(\sum_{i\leq j} c_j \binom{j}{i} x^{j-i} \right) h^i = \sum_{i\in\mathcal{I}} D^i f(x)h^i.$$

and we have proved the proposition. □

2.4.10. Proposition 2.4.9 and its proof indicate that analytic functions behave in a similar way to the differentiable functions from calculus. We shall next give another example of such behavior.

Let V be an open subset in $\mathbf{K}^p$ and $g\colon V \to \mathbf{K}^n$ an analytic function such that the image $g(V)$ is contained in an open subset U of $\mathbf{K}^n$, and let $f : U \to \mathbf{K}^m$ be an analytic function. Then the composite function $fg\colon V \to \mathbf{K}^m$ of g, is analytic. Indeed, it suffices to consider a neighborhood of 0 in $\mathbf{K}^p$, and we can assume that $g(0) = 0$, that $f(0) = 0$, and that $m = 1$. Let $f(x) = \sum_{i\in\mathcal{I}} c_i x^i$ be a convergent series in $P(0,s)$, for some s in $\mathcal{R}$ and

$g = (g_1, \dots, g_n)$, with $g_k(y) = \sum_{j \in \mathcal{J}} d_{k,j} y^j$, be an n-tuple of series that are convergent in $P(0, r)$ for some r in $\mathcal{R}$, and where $\mathcal{J}$ are p-tuples of non-negative integers. Choose $r' < r$ such that

$$\sum_{i \in \mathcal{J}} |d_{k,i}| r'^i < \frac{s_k}{2} \quad \text{for } k = 1, \dots, n.$$

Then, for $h \in P(0, r')$, we have that

$$\sum_{i \in \mathcal{I}} |c_i| (\sum_{j \in \mathcal{J}} |d_{1,j}||h|^j, \dots, \sum_{j \in \mathcal{J}} |d_{n,j}||h|^j)^i \leq \sum_{i \in \mathcal{I}} |c_i| \left(\frac{s}{2}\right)^i < \infty.$$

Consequently, we have that

$$\sum_{i \in \mathcal{I}} c_i (\sum_{j \in \mathcal{J}} d_{1,j} y^j, \dots, \sum_{j \in \mathcal{J}} d_{n,j} y^j)^i \tag{2.4.10.1}$$

converges in $P(0, r')$, and the series 2.4.10.1 represents $fg(y)$.

Definition 2.4.11. Let U be an open subset of $\mathbf{K}^n$ and let

$$f \colon U \to \mathbf{K}^m$$

be a function. If there exists a linear map $g : \mathbf{K}^n \to \mathbf{K}^m$ such that

$$\lim_{\|h\| \to 0} \frac{\|f(x+h) - f(x) - g(h)\|}{\|h\|} = 0,$$

where $\|h\| = \max_i |h_i|$, we say that f is *differentiable* at x. Clearly, g is unique if it exists, and we write $f'(x) = g$ and $f'(x)h = g(h)$, and call $f'(x)$ the *derivative* of f at x. We say that f is *differentiable in* U if it is differentiable at each point of U.

Remark 2.4.12. Usually the linear map $f'(x)$ is represented by an $m \times n$ matrix with respect to the standard bases of $\mathbf{K}^n$ and $\mathbf{K}^m$ and the distinction between the matrix and the map is often suppressed in the notation. The matrix $f'(x)$ is referred to as the *Jacobian* of the map f.

When $f = (f_1, \dots, f_m)$ we have that f is differentiable, if and only if all the f_i are differentiable, and we have that $f' = (f'_1, \dots, f'_m)$.

Proposition 2.4.13. *Let $f \colon U \to \mathbf{K}$ be an analytic function defined on an open subset U of $\mathbf{K}^n$, and let $f(x) = \sum_{i \in \mathcal{I}} c_i x^i$. Then $f(x)$ is differentiable in U and the derivative $f'(x)$ is an analytic function $f' \colon U \to \mathrm{Hom}_{\mathbf{K}}(\mathbf{K}^n, \mathbf{K}) = \check{\mathbf{K}}^n$ given by*

$$f'(x)h = \sum_{|i|=1} D^i f(x) h^i = \sum_{j \geq i, |i|=1} c_j \binom{j}{i} x^{j-i} h^i, \qquad \text{for all } h \in \mathbf{K}^n,$$

with the notation of Proposition 2.4.9.

Proof: It follows from Proposition 2.4.9 (iii) that $f(x+h) - f(x) - \sum_{|i|=1} D^i f(x) h^i$ is an analytic function of h in $P(0, r)$ whose terms in h of order 0 and 1 vanish. Consequently we have that

$$\lim_{\|h\|\to 0} \frac{\|f(x+h) - f(x) - \sum_{|i|=1} D^i f(x) h^i\|}{\|h\|} = 0,$$

that is $f'(x)h = \sum_{|i|=1} D^i f(x) h^i$. It follows from Proposition 2.4.9 that $f'(x)$ is analytic. □

Remark 2.4.14. Let $m = 1$ in Definition 2.4.11 and let f be analytic. For $i = 1, \dots, n$ we let $\frac{\partial f}{\partial x_i}(x)$ be the $(1, i)$'th coordinate of the $1 \times n$ matrix A. It follows from Proposition 2.4.13 that

$$\frac{\partial f}{\partial x_i}(x) = D^{(0,\dots,1,\dots,0)} f(x),$$

where the 1 in the exponent of D is in the i'th place. Consequently, we have that

$$f'(x) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right).$$

For any m and with $f = (f_1, \dots, f_m)$ we obtain that $f'(x)$ is the $m \times n$ matrix $f'(x) = (\frac{\partial f_i}{\partial x_j}(x))$.

When $g\colon V \to \mathbf{K}^n$ is an analytic function from an open subset V in $\mathbf{K}^p$ such that $g(V) \subseteq U$, Formula 2.4.10.1 shows that for y in V we have that

$$(fg)'(y) = f'(g(y))g'(y) \tag{2.4.14.1}$$

In fact we may clearly assume that $y = 0$ and $g(y) = 0$. Let $e_1, \dots, e_n$ and $f_1, \dots, f_p$ be the standard bases of $\mathbf{K}^n$, respectively $\mathbf{K}^p$. Then, when $m = 1$, we get $f'(0) = (c_{e_1}, \dots, c_{e_n})$ and $g'(0) = (d_{i,f_j})$. Formula 2.4.10.1 gives that $\frac{\partial(fg)}{\partial y_i}(0) = \sum_{j=1}^n c_{e_j} d_{j,f_i}$, and thus $(\frac{\partial g_j}{\partial y_j}(x))(\frac{\partial f}{\partial x_1}(0), \dots, \frac{\partial f}{\partial x_n}(0)) = (\frac{\partial(fg)}{\partial y_1}(0), \dots, \frac{\partial(fg)}{\partial y_p}(0))$. Hence we have that $(fg)'(0) = f'(0)g'(0)$.

Example 2.4.15. Let $f : \mathbf{K} \to \mathbf{K}^m$ be a differentiable function. For t in $\mathbf{K}$ we have a linear map $f'(t) : \mathbf{K} \to \mathbf{K}^m$. We shall identify the linear maps $g : \mathbf{K} \to \mathbf{K}^n$ with $\mathbf{K}^m$ by identifying g with $g(1)$. Then $f'(t)$ is considered as an element in $\mathbf{K}^m$ and we obtain a map $f' : \mathbf{K} \to \mathbf{K}^m$ which maps t to $f'(t)$.

Let X be in $\mathrm{M}_n(\mathbf{K})$ and let $e_X : \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ be the map defined by $e_X(t) = \exp(Xt)$. It follows from Example 2.4.8 that e_X is analytic. Hence we have a map $e'_X : \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$. It follows from Exercise 2.4.1 that $e'_X(t) = \exp'(Xt)X$ and from Proposition 2.4.13 that e'_X is analytic. We have that $e'_X(t) = Xe_X(t)$. In fact, fix an elements t in $\mathbf{K}$ and define $e_t : \mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$ and $e'_t : \mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$ by $e_t(X) = e_X(t)$ and $e'_t(X) = e'_X(t)$. We shall prove that $Xe_t(X) = e'_t(X)$. It follows from Example 2.4.8 and Proposition 2.4.13 that e_t and e'_t are analytic, and in particular continuous, as we noted in Definition 2.4.4. Consequently it follows from Lemma 2.3.2 and Corollary 2.3.5 that it suffices to show that $Xe_t(X)$ and $e'_t(X)$ are equal on diagonalizable matrices X. It follows from assertion (v) of Proposition 2.2.8 that $Y^{-1}XYe_t(Y^{-1}XY) = Y^{-1}XYe_{Y^{-1}XY}(t) = Y^{-1}Xe_X(t)Y$. and from Exercise 2.4.1 that $e'_t(Y^{-1}XY) = e'_{Y^{-1}XY}(t) = (Y^{-1}e_XY)'(t) = Y^{-1}e'_X(t)Y$. Hence, to prove that $e_t = e'_t$ it suffices to prove that they are equal on diagonal matrices. The latter equality follows from the sequence of equalities:

$$\left(\exp\begin{pmatrix} x_1t & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_nt \end{pmatrix}\right)'(t) = \begin{pmatrix} \exp(x_1t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp(x_nt) \end{pmatrix}'(t) = \begin{pmatrix} \exp'(x_1t)'(t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp'(x_nt)'(t) \end{pmatrix}$$
$$= \begin{pmatrix} x_1\exp(x_1t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n\exp(x_nt) \end{pmatrix} = \exp\begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix}\exp\begin{pmatrix} x_1t & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_nt \end{pmatrix}.$$

Remark 2.4.16. The formula $e'_X = Xe_X$ can also easily be proven using Proposition 2.4.13, observing that $\exp_m(Xt)'(t) = X\exp_{m-1}(Xt)$ and that the limit of the sequence $\{\exp_m(Xt)'(t)\}_{m=0,1,\dots}$ is $e'_X(t)$, and the limit of $\{X\exp_n(Xt)\}_{m=0,1,\dots}$ is $Xe_X(x)$.

Remark 2.4.17. Let G be one of the groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$ or $\mathrm{G}_S(\mathbf{K})$, for some invertible S. It follows from Proposition 2.3.9 that for each matrix group G there is an open neighborhood U of the identity, and an open subset V of some vector space, such that the exponential function induces an isomorphism $\exp\colon V \to U$, with inverse $\log|_U$. Let A be an element in G. There is a map $\lambda_A\colon G \to G$, called *left translation*, defined by $\lambda_A(B) = AB$. The left translations are given by polynomials, hence they are analytic. The left translation λ_A induces a homeomorphism $\lambda_A|_U\colon U \to \lambda_A(U)$ of metric spaces onto the open neighborhood $\lambda_A(U)$ of A, with inverse $\lambda_{A^{-1}}$. Consequently, for each A, we have a homeomorphism $\varphi_A\colon V \to U_A$ onto some neighborhood of A. Clearly, if we have another such homeomorphism φ_B, such that $U_A \cap U_B \neq \emptyset$, we have that the map $\varphi_A^{-1}(U_A \cap U_B) \to \varphi_B^{-1}(U_A \cap U_B)$ induced by $\varphi_B^{-1}\varphi_A$, is an analytic map. We summarize these properties by saying that G is an *analytic manifold*.

Exercises

2.4.1. Let X be an $n \times n$-matrix of the form $X(x) = (f_{ij}(x))$, where the functions $f_{ij}\colon V_{\mathbf{K}}^n \to \mathbf{K}$ are analytic, and let Y be in $\mathrm{Gl}_n(\mathbf{K})$. Show that the the derivative $(Y^{-1}XY)'$ of the function $Y^{-1}XY\colon V_{\mathbf{K}}^n \to \mathrm{M}_n(\mathbf{K})$, which takes x to $Y^{-1}X(x)Y$, is equal to $Y^{-1}X'(x)Y$.

2.4.2. Let X be in $\mathrm{M}_n(\mathbf{K})$ and let $f : \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ be defined by $f(x) = Xt$ for t in $\mathbf{K}$. Show that, with the identifications of Example 2.4.15 we have that $f' : \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ is the constant function $f'(t) = X$.

2.4.3. Let $f : \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ be an analytic function and let Y be in $\mathrm{M}_n(\mathbf{K})$. We define $Y^{-1}fY : \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ to be the function that maps t in $\mathbf{K}$ to $Y^{-1}f(x)Y$. Show that $(Y^{-1}fY)' : \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ satisfies $(Y^{-1}fY)'(t) = Y^{-1}f'(t)Y$.

2.4.4. Determine the derivative of the exponential map $\exp : \mathrm{M}_2(\mathbf{C}) \to \mathrm{Gl}_2(\mathbf{C})$.

2.5 Tangent spaces of matrix groups

We shall, in this section, determine the tangent spaces of all the matrix groups that we have encountered so far.

Definition 2.5.1. A *curve in* $V_{\mathbf{K}}^n$ is an analytic map $\gamma\colon B(a,r) \to V_{\mathbf{K}}^n$, from some ball $B(a,r)$ in $\mathbf{K}$. The *tangent of the curve* γ at $\gamma(a)$ is the vector $\gamma'(a)$ in $V_{\mathbf{K}}^n$.

Let $\gamma\colon B(a,r) \to \mathrm{M}_n(\mathbf{K})$ be a curve and let G be one of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, or $\mathrm{G}_S(\mathbf{K})$, for some invertible S. We say that γ is a *curve in* G if $\gamma(B(a,r))$ is in G and if $\gamma(a) = I_n$.

The *tangent space* $T_{I_n}(G)$ of G at I_n is the set of the tangent vectors at a for all curves $\gamma\colon B(a,r) \to \mathrm{M}_n(\mathbf{K})$ in G.

Remark 2.5.2. Since $\mathrm{SO}_n(\mathbf{K})$ is an open subset of $\mathrm{O}_n(\mathbf{K})$ containing I_n, we have that $\mathrm{SO}_n(\mathbf{K})$ and $\mathrm{O}_n(\mathbf{K})$ have the same tangent space.

Example 2.5.3. Let X be a matrix in $\mathrm{M}_n(\mathbf{K})$. The derivative $\exp'(tX)$ of the curve $\gamma(t)\colon \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ that is defined by $\gamma(t) = \exp(tX)$ is equal to $X(\exp tX)$ by Example 2.4.15 (see also Exercise 2.5.3). When X is in $\mathfrak{gl}_n(\mathbf{K})$, $\mathfrak{sl}_n(\mathbf{K})$ or $\mathfrak{g}_S(\mathbf{K})$, for some S, it follows from Proposition 2.3.9 that γ has image contained in $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$ or $\mathrm{G}_S(\mathbf{K})$, respectively.

In particular, the tangent spaces of $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$ and $\mathrm{Sp}_n(\mathbf{K})$ contain the vector spaces $\mathfrak{gl}_n(\mathbf{K})$, $\mathfrak{sl}_n(\mathbf{K})$, $\mathfrak{so}_n(\mathbf{K})$, $\mathfrak{so}_n(\mathbf{K})$, and $\mathfrak{sp}_n(\mathbf{K})$, respectively.

We shall next show that the inclusions of spaces of Example 2.5.3 are equalities.

Proposition 2.5.4. *The tangent spaces at* I_n *of the matrix groups* $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$ *or* $\mathrm{G}_S(\mathbf{K})$, *where* S *is an invertible matrix, are the vector spaces* $\mathfrak{gl}_n(\mathbf{K})$, $\mathfrak{sl}_n(\mathbf{K})$, *and* $\mathfrak{g}_S(\mathbf{K})$ *respectively.*

In particular, the tangent spaces of $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$ *and* $\mathrm{Sp}_n(\mathbf{K})$ *are* $\mathfrak{so}_n(\mathbf{K})$, $\mathfrak{so}_n(\mathbf{K})$, *and* $\mathfrak{sp}_n(\mathbf{K})$ *respectively.*

Proof: Let G be one of the groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, or $\mathrm{G}_S(\mathbf{K})$, and let $\gamma\colon B(a,r) \to \mathrm{M}_n(\mathbf{K})$ be a curve from a ball $B(a,r)$ in $\mathbf{K}$, such that $\gamma(a) = I_n$. It follows from Example 2.4.15, or Exercise 2.5.3, that is suffices to show that, when the image of γ is in G, the derivative $\gamma'(a)$ is in $\mathfrak{gl}_n(\mathbf{K})$, $\mathfrak{so}_n(\mathbf{K})$ or $\mathfrak{g}_S(\mathbf{K})$, respectively.

For $\mathrm{Gl}_n(\mathbf{K})$ this is evident since the tangent space is the whole of $\mathfrak{gl}_n(\mathbf{K}) = \mathrm{M}_n(\mathbf{K})$. If the image of γ is in $\mathrm{Sl}_n(\mathbf{K})$, we have that $\det\gamma(t) = 1$ for all t in $B(a,r)$. We differentiate the last equality and obtain that $0 = (\det\gamma)'(a) = \mathrm{tr}(\gamma')(a)$ (see Exercise 2.5.1), that is, the matrix $\gamma'(a)$ is in $\mathfrak{sl}_n(\mathbf{K})$.

Let γ be in $\mathrm{G}_S(\mathbf{K})$. That is, we have ${}^t\gamma(t)S\gamma(t) = S$, for all t in $B(a,r)$. We differentiate both sides of the latter equation and obtain that ${}^t\gamma'(t)S\gamma(t) + {}^t\gamma(t)S\gamma'(t) = 0$, for all t in $B(a,r)$ (see Exercise 2.5.2). Consequently, we have that ${}^t\gamma'(0)S\gamma(0) + {}^t\gamma(0)S\gamma'(0) = {}^t\gamma'(0)SI_n + {}^tI_nS\gamma'(0) = {}^t\gamma'(0)S + {}^tS\gamma'(0)$, and we have proved the first part of the proposition.

For the last part of the proposition it suffices to note that it follows from Remark 2.5.2 that $\mathrm{SO}_n(\mathbf{K})$ and $\mathrm{O}_n(\mathbf{K})$ have the same tangent space. □

Definition 2.5.5. Let G be one of the groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, or $\mathrm{G}_S(\mathbf{K})$, where S is invertible. The *dimension* $\dim G$ is the dimension of the vector space $T_{I_n}(G)$.

Group	n	Center	Dim.
$\mathrm{Gl}_n(\mathbf{C})$	arb.	$\mathbf{K}^*$	n^2
$\mathrm{Sl}_n(\mathbf{C})$	arb.	$\mathbf{Z}/n\mathbf{Z}$	n^2-1
$\mathrm{O}_n(\mathbf{C})$	arb.	$\{\pm 1\}$	$\frac{n(n-1)}{2}$
$\mathrm{SO}_n(\mathbf{C})$	even	$\{\pm 1\}$	$\frac{n(n-1)}{2}$
$\mathrm{SO}_n(\mathbf{C})$	odd	1	$\frac{n(n-1)}{2}$
$\mathrm{Sp}_n(\mathbf{C})$	arb.	$\{\pm 1\}$	$\frac{n(n+1)}{2}$

TABLE 1. The classical groups over the complex numbers

Proposition 2.5.6. *The dimensions of the matrix groups are:*
$\dim \mathrm{Gl}_n(\mathbf{K}) = n^2$, $\dim \mathrm{Sl}_n(\mathbf{K}) = n^2-1$, $\dim \mathrm{O}_n(\mathbf{K}) = \dim \mathrm{SO}_n(\mathbf{K}) = \frac{n(n-1)}{2}$, *and* $\dim \mathrm{Sp}_n(\mathbf{K}) = \frac{n(n+1)}{2}$.

Proof: We shall use the description of the tangent spaces of Proposition 2.5.4.

The dimension of $\mathfrak{gl}_n(\mathbf{K}) = \mathrm{M}_n(\mathbf{K})$ is clearly n^2. That the dimension of the space $\mathfrak{sl}_n(\mathbf{K})$ of matrices with trace zero is n^2-1 follows from Exercise 2.5.4. The spaces $\mathrm{O}_n(\mathbf{K})$ and $\mathrm{SO}_n(\mathbf{K})$ have the same tangent space $\mathfrak{so}_n(\mathbf{K})$ consisting of skew-symmetric matrices. It follows from Exercise 2.5.5 that this dimension is $\frac{n(n-1)}{2}$.

The space $\mathfrak{sp}_n(\mathbf{K})$ consists of invertible matrices X such that ${}^tXS + SX = 0$, where S is the matrix of the form 1.4.1.1. We have that the linear map $\mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$ that sends a matrix X to SX is an isomorphism (see Exercise 2.5.6). The latter map sends $\mathfrak{sp}_n(\mathbf{K})$ isomorphically onto the space of symmetric matrices. Indeed, we have that ${}^tXS + SX = -{}^tX{}^tS + SX = SX - {}^t(SX)$. However, the space of symmetric matrices has dimension $\frac{n(n+1)}{2}$ (see Exercise 2.5.7). □

We summarize the results of Sections 2.5 and 1.10 in Table 1.

Exercises

2.5.1. Given an $n \times n$ matrix $X(x) = (f_{ij}(x))$, where the coordinates are analytic functions $f_{ij} \colon B(b,s) \to \mathbf{K}$ on a ball $B(b,s)$ in $V^m_{\mathbf{K}}$. We obtain an analytic function $\det_X \colon B(b,s) \to \mathbf{K}$ defined by $\det_X(t) = \det(X(t))$ for t in $B(b,x)$. Show that

$$(\det_X)'(t) = \sum_{i=1}^n \det \begin{pmatrix} f_{11}(t) & \dots & f'_{1i}(t) & \dots & f_{1n}(t) \\ \vdots & & \vdots & & \vdots \\ f_{n1}(t) & \dots & f'_{ni}(t) & \dots & f_{nn}(t) \end{pmatrix}.$$

Assume that $X(0) = I_n$. Show that $(\det_X)'(0) = \sum_{i=1}^n f'_{ii}(0) = \mathrm{tr}(X')(0)$.

2.5.2. Let $X(t) = (f_{ij}(t))$ and $Y(t) = (g_{ij}(t))$ be functions $B(a,r) \to \mathrm{M}_n(\mathbf{K})$ given by analytic functions f_{ij} and g_{ij} on a ball $B(a,r)$ in $\mathbf{K}$. Show that $(XY)'(t) = X'(t)Y(t) + X(t)Y'(t)$.

2.5.3. Let X be a matrix in $\mathrm{M}_n(\mathbf{K})$. Prove the equality $\exp_m(Xt)' = X\exp_m(Xt)$, and use this to show that the tangent of the curve $\gamma \colon \mathbf{K} \to \mathrm{M}_n(\mathbf{K})$ given by $\gamma(t) = \exp(tX)$ at t is $X\exp(tX)$.

2.5.4. Show that the vector space of matrices in $\mathrm{M}_n(\mathbf{K})$ with trace zero, that is with the sum of the diagonal elements equal to zero, has dimension n^2-1.

2.5.5. Show that the vector space of matrices in $\mathrm{M}_n(\mathbf{K})$ consisting of skew-symmetric matrices has dimension $\frac{n(n-1)}{2}$.

2.5.6. Fix a matrix B in $\mathrm{Gl}_n(\mathbf{K})$. Show that the map $\mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$ that sends a matrix X to the matrix BX is an isomorphism of vector spaces.

2.5.7. Show that the subset of $\mathrm{M}_n(\mathbf{K})$ consisting of symmetric matrices is a vector space of dimension $\frac{n(n+1)}{2}$.

2.5.8. Which of the groups $\mathrm{Gl}_n(\mathbf{C})$, $\mathrm{Sl}_n(\mathbf{C})$, $\mathrm{O}_n(\mathbf{C})$, $\mathrm{SO}_n(\mathbf{C})$, and $\mathrm{Sp}_n(\mathbf{C})$, can be distinguished by the table of this section.

2.5.9. Determine the the dimension and a basis of the tangent space $\mathfrak{g}_S(\mathbf{R})$ of the Lorentz group defined by the matrix $S = \begin{pmatrix} I_{n-1} & 0 \\ 0 & -1 \end{pmatrix}$.

2.6 Lie algebras of the matrix groups

Lie algebras will become important in Chapter 4. In this section we indicate how matrix groups are Lie algebras.

Remark 2.6.1. In addition to the usual matrix multiplication on the space of matrices $\mathrm{M}_n(\mathbf{K})$ we have a map

$$[\,,]\colon\ \mathrm{M}_n(\mathbf{K}) \times \mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$$

defined by $[A, B] = AB - BA$. It is easy to check (see Exercise 2.6.1) that $[\,,]$ is an alternating bilinear map which satisfies the *Jacobi Identity*

$$\big[A, [B, C]\big] + \big[C, [A, B]\big] + \big[B, [C, A]\big] = 0, \qquad \text{for all } A, B, C \in \mathrm{M}_n(\mathbf{K})$$

We summarize these properties by saying that $\mathrm{M}_n(\mathbf{K})$ is a *Lie algebra*. When $\mathrm{M}_n(\mathbf{K})$ is considered as a Lie algebra we shall denote it by $\mathfrak{gl}_n(\mathbf{K})$. A subspace V of $\mathfrak{gl}_n(\mathbf{K})$ such that $[A, B] \in V$, for all A and B in V, is called a *Lie subalgebra* of $\mathfrak{gl}_n(\mathbf{K})$. Clearly the Jacobi Identity holds for all elements in V. Hence V is itself a Lie algebra.

Example 2.6.2. The tangent spaces $\mathfrak{gl}_n(\mathbf{K})$, $\mathfrak{sl}_n(\mathbf{K})$, and $\mathfrak{g}_S(\mathbf{K})$, when S is invertible, of $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, and $\mathrm{G}_S(\mathbf{K})$ respectively, are all Lie subalgebras of $\mathfrak{gl}_n(\mathbf{K})$.

In particular, the tangent spaces $\mathfrak{so}_n(\mathbf{K})$, $\mathfrak{so}_n(\mathbf{K})$, and $\mathfrak{sp}_n(\mathbf{K})$ of $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, and $\mathrm{Sp}_n(\mathbf{K})$ respectively, are Lie subalgebras of $\mathfrak{gl}_n(\mathbf{K})$.

It follows from Exercise 2.6.2 that $\mathfrak{sl}_n(\mathbf{K})$ is a Lie subalgebra of $\mathfrak{gl}_n(\mathbf{K})$. That $\mathfrak{g}_S(\mathbf{K})$ is a Lie subalgebra of $\mathfrak{gl}_n(\mathbf{K})$ follows from the calculation $[A, B]S + S^t[B, A] = (AB - BA)S + S^t(AB - BA) = ABS - BAS + S^tB^tA - S^tA^tB = -AS^tB + BS^tA + S^tB^tA - S^tA^tB = S^tA^tB - S^tB^tA + S^tB^tA - S^tA^tB = 0$.

Exercises

2.6.1. Show that $[,]$ defines an alternating bilinear map on $\mathrm{M}_n(\mathbf{K})$ that satisfies the Jacobi Identity.

2.6.2. Show that the subspace of matrices of $\mathfrak{gl}_n(\mathbf{K})$ with trace zero is a Lie algebra.

2.6.3. Let $V = \mathbf{K}^3$ and define $[,]$ as the cross product from linear algebra, that is,

$$[(x_1, y_1, z_1), (x_2, y_2, z_2)] = (y_1z_2 - z_1y_2, z_1x_2 - x_1z_2, x_1y_2 - y_1x_2).$$

Show that V becomes a Lie algebra with this product.

2.6.4. Show that the tangent space $\mathrm{SO}_3(\mathbf{K})$ as a Lie algebra is isomorphic to the Lia algebra of Problem 2.6.3.

2.6.5. Show that the tangent space $\mathfrak{so}_3(\mathbf{K})$ as a Lie algebra is isomorphic to the Lie algebra of Problem PB:tredim.

2.6.6. In quantum mechanics, we have the *Pauli spin matrices*

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(i) Show that the set $\{1, \sigma_x, \sigma_y, \sigma_z\}$ spans the Lie algebra $\mathfrak{gl}_2(\mathbf{C})$.
(ii) Show that the set $\{\sigma_x, \sigma_y, \sigma_z\}$ spans a three dimensional sub Lie algebra of $\mathfrak{gl}_2(\mathbf{C})$ which is identical to $\mathfrak{sl}_2(\mathbf{C})$.
(iii) Show that the Lie algebra of (ii) is isomorphic to the Lie algebras of Problem PB:tredim and Problem 2.6.4.

2.7 One parameter subgroups of matrix groups

One parameter groups play an important part in the theory of Lie groups of Section 4. In this section we determine the one parameter subgroups for matrix groups.

Definition 2.7.1. Let G be one of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$ or $\mathrm{G}_S(\mathbf{K})$, when S is invertible. A *one parameter subgroup* of the matrix groups G is a curve $\gamma\colon K \to G$, which is also a group homomorphism. That is $\gamma(t+u) = \gamma(t)\gamma(u)$, for all t and u in $\mathbf{K}$.

Example 2.7.2. Let X be a matrix in $T_{I_n}(G)$. It follows from Example 2.5.3 that $\gamma\colon \mathbf{K} \to G$ defined by $\gamma(t) = \exp(tX)$ is a curve in G. Since tX and sX commute, it follows from Proposition 2.2.8 (ii), that γ is a one parameter group.

We shall show that all one parameter groups of the matrix groups are of the form of Example 2.7.2.

Proposition 2.7.3. *Let G be one of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$ or $\mathrm{G}_S(\mathbf{K})$, for some invertible matrix S. Then all one parameter groups of G are of the form $\gamma(t) = \exp(tX)$, for some X in $\mathfrak{gl}_n(\mathbf{K})$, $\mathfrak{sl}_n(\mathbf{K})$ or $\mathfrak{g}_S(\mathbf{K})$, respectively.*

Proof: Let $\gamma\colon \mathbf{K} \to G$ be a one parameter group. It follows from Proposition 2.3.9 and Example 2.4.8 that there is a neighborhood U of I_n in G such that the logarithm induces an analytic function $\log\colon U \to \mathrm{M}_n(\mathbf{K})$. We obtain an analytic map $\log\gamma\colon B(a,r) \to \mathrm{M}_n(\mathbf{K})$, on some ball $B(a,r)$ in $\mathbf{K}$, and $\log\gamma(a) = \log(I_n) = 0$. For all t and u in $B(a,r)$, such that $t+u$ is in $B(a,r)$ we have that

$$\begin{aligned}\log\gamma(t+u) - \log\gamma(t) - (\log\gamma)'(t)u &= \log(\gamma(t)\gamma(u)) - \log(\gamma(t)) - (\log\gamma)'(t)u \\ &= \log(\gamma(t)) + \log(\gamma(u)) - \log(\gamma(t)) - (\log\gamma)'(t)u = \log\gamma(u) - (\log\gamma)'(t)u.\end{aligned}$$

Consequently, we have that

$$\lim_{|u|\to 0} \frac{\|\log\gamma(t+u) - \log\gamma(t) - (\log\gamma)'(t)u\|}{\|u\|} = \lim_{|u|\to 0} \frac{\|\log\gamma(u) - (\log\gamma)'(t)u\|}{\|u\|} = 0.$$

That is, we have $(\log\gamma)'(0) = (\log\gamma)'(t)$. Hence $(\log\gamma)'(t)$ is constant equal to $X = (\log\gamma)'(0)$ on some ball $B(a,\varepsilon)$ (see Exercise 2.7.1). We thus have that $\log\gamma(t) = tX$. Using the exponential function on both sides of $\log\gamma(t) = tX$, and Theorem 2.3.6 (i), we obtain that $\gamma(t) = \exp(tX)$, for all t in the ball $B(a,\varepsilon)$. It follows that $\gamma(t) = \exp(tX)$ for all t. Indeed, given an element t of $\mathbf{K}$. Choose an integer n such that $\frac{1}{n}t$ is in $B(a,\varepsilon)$. Then we obtain that $\gamma(t) = \gamma(\frac{n}{n}t) = \gamma(\frac{1}{n}t)^n = \exp(\frac{1}{n}tX)^n = \exp(\frac{n}{n}tX) = \exp(tX)$, which we wanted to prove. □

Exercises

2.7.1. let U be an open subset of $\mathbf{K}$ and let $f : U \to \mathbf{K}^m$ be an analytic function. Show that if the linear map $f' : \mathbf{K} \to \mathbf{K}^m$ is constant, that is $f'(t)$ is the same element x in $\mathbf{K}^m$ for all t in U, then $f(t) = xt$ for all t in U.

3 The geometry of matrix groups

In Chapter 2 we introduced a geometric structure on the *classical groups* of Chapter 1 with coefficients in the real or complex numbers. The main tool was the exponential function. In this chapter we shall consider the geometric structure from a much more general point of view, that of analytic manifolds. The main tool here is the *Implicit function Theorem* for analytic maps.

We prove the *Implicit Function Theorem* and show that the *classical groups* studied in Chapter 2 are algebraic manifolds. Analytic manifolds are introduced after we have defined and studied topological spaces. We define and study tangent spaces of analytic manifolds, and show that the definition of tangent spaces coincides with that given in Chapter 2 for matrix groups. More generally we define tangent spaces of analytic sets and give some criteria for computing the tangent space in the most common geometric situations. Using the so called *epsilon calculus* we use these results to compute the tangent spaces of the *classical groups* of Chapter 2.

In order to further distinguish between the matrix groups we introduce, in the final section, two topological invariants, connectedness and compactness.

Unless explicitly stated otherwise, the field $\mathbf{K}$ will be the real or the complex numbers throughout this chapter.

3.1 The Inverse Function Theorem

We shall in this section prove the Inverse Function Theorem and show how several versions of the Implicit Function Theorem are deduced from the inverse function theorem. Most of these results are probably known for the differentiable case from a course in calculus of several variables. The reason why we give proofs is that the analytic case, although easier, is less standard in calculus books.

Theorem 3.1.1. (Inverse Function Theorem) *Let W be an open subset of $\mathbf{K}^n$ and $\Phi\colon W \to \mathbf{K}^n$ an analytic function. Moreover let y be a point in W such that $\Phi'(y)$ is invertible. Then there exists an open neighborhood U of y in W and an open set V in $\mathbf{K}^n$ such that Φ is injective on U and $\Phi(U) = V$. Moreover the inverse function $\Phi^{-1}\colon V \to \mathbf{K}^n$, defined by*

$$\Phi^{-1}(\Phi(x)) = x, \quad \text{for all } x \in U,$$

is analytic on V.

Proof: We may assume that $y = \Phi(y) = 0$ and, following Φ with the inverse of the analytic linear function $\Phi'(0)\colon \mathbf{K}^n \to \mathbf{K}^n$, we may assume that

$$\Phi_k(x) = x_k - \sum_{|i|>1} c_{ki}x^i = x_k - \varphi_k(x), \qquad \text{for } k = 1, 2, \dots, n.$$

Moreover, we may replace $\Phi(x)$ by $C\Phi(x/C)$ for some positive real number C, and assume that $|c_{ki}| < 1$ for all i in $\mathcal{I}$.

Suppose that the analytic function $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_n)$, given by

$$\Psi_k(y) = \sum_{i \in I} d_{ki} y^i$$

is an inverse function to Φ. Then $y = \Phi_k(\Psi(y)) = \Psi_k(y) - \varphi_k(\Psi(y))$ so that Ψ must satisfy the equation

$$\begin{aligned}\Psi_k(y) &= \sum_{i \in I} d_{ki} y^i = y_k + \varphi_k(\Psi(y)) \\ &= y_k + \sum_{|i|>1} c_{ki} (\sum_{|j|>0} d_{1j} y^j)^{i_1} \cdots (\sum_{|j|>0} d_{nj} y^j)^{i_n}, \qquad \text{for } k = 1, \dots, n.\end{aligned}$$

Comparing coefficients on both sides of the equation we see that $d_{k(0,\dots,1,\dots,0)}$ is 1 when the 1 in $(0, \dots, 1, \dots, 0)$ is in the k'th coordinate, and 0 otherwise. Moreover, we see that d_{kj} is a linear combination with positive integral coefficients of monomials in the d_{mi} and the c_{mi} with $|i| < |j|$. By induction on $|i|$ we obtain that

$$d_{kj} = P_{kj}(c_{mi}), \tag{3.1.1.1}$$

where P_{ki} is a polynomial with positive integral coefficients that depend only on c_{mi}, with $|i| < |j|$. In particular we have that each Ψ_k is uniquely determined if it exist. The problem is to show that the formal power series determined by the solutions of the equation 3.1.1.1 converges. To this end, assume that we can find power series $\bar{\varphi}_k = \sum_{i \in \mathcal{I}} \bar{c}_{ki} x^i$, for $k = 1, \dots, n$, with real positive coefficients $\bar{c}_{ki}$ that converge in some polydisc around 0 in $\mathbf{K}^n$ and which is such that the unique power series $\bar{\Psi}_k = \sum_{i \in \mathcal{I}} \bar{d}_{ki} x^i$ determined by the equation

$$\bar{d}_{ki} = P_{ki}(\bar{c}_{mj}),$$

for $k = 1, \dots, n$, and hence satisfy

$$\bar{\Psi}_k(y) = y_k + \bar{\varphi}_k(\bar{\Psi}(y)), \qquad \text{for } k = 1, \dots, n,$$

converge in some polydisc around 0 and satisfy the conditions

$$|c_{ki}| \le \bar{c}_{ki}, \qquad \text{for all } k \text{ and } i.$$

Then we have that

$$|d_{ki}| = |P_{ki}(c_{mi})| \le P_{ki}(|c_{mj}|) \le P_{ki}(\bar{c}_{mj}) = \bar{d}_{ki},$$

since P_{ki} has positive integral coefficients. Consequently Ψ_k is dominated by $\bar{\Psi}_k$ and thus converges for $k = 1, \dots, n$. It remains to find such series $\bar{\varphi}_k$, for $k = 1, 2, \dots, n$.

Assume that $n = 1$. We have that, for any positive real number p, the series

$$\bar{\varphi}^{(p)}(x) = \sum_{i=2}^{\infty} (px)^i = \frac{(px)^2}{1 - px},$$

will satisfy the conditions. Indeed, it converges and satisfies $|c_i| \leq \bar{c}_i$, since $|c_i| \leq 1$. We must show that the corresponding $\bar{\Psi}^{(p)}$ converges. However, we have the equations

$$\bar{\Psi}^{(p)}(y) = y + \bar{\varphi}^{(p)}(\bar{\Psi}^{(p)}(y)) = y + \frac{(p\bar{\Psi}^{(p)}(y))^2}{1 - p\bar{\Psi}^{(p)}(y))}.$$

Solving the latter equation we obtain that

$$\bar{\Psi}^{(p)}(y) = \frac{1}{2}\frac{(1+yp) - \sqrt{(1+yp)^2 - 4(p^2+p)y}}{p^2+p},$$

which converges in a polydisc around 0.

Let $n > 1$ and put

$$\bar{\varphi}_k(x) = \bar{\varphi}^{(1)}(\sum_{i=1}^n x_i) = \sum_{i=2}^\infty (x_1 + \cdots + x_n)^i = \frac{(x_1 + \cdots + x_n)^2}{1 - (x_1 + \cdots + x_n)}, \qquad \text{for } k = 1, 2, \dots, n.$$

Then $\bar{\varphi}_k$ converges in a neighborhood of 0 and we have that $|c_{ki}| < \bar{c}_{ki}$ for all k and i, since $|c_{ki}| < 1$. Observe that $\bar{\Phi}_j(x) - \bar{\Phi}_k(x) = x_j - x_k$, for $k \neq j$. Hence, if we can find the average of $x_1, \dots, x_n$, from $\bar{\Phi}_1(x), \dots, \bar{\Phi}_n(x)$, we can determine the inverse function $\bar{\Psi}$. In fact we have that

$$\frac{1}{n}\sum_{k=1}^n \bar{\Phi}_k(x) = \frac{1}{n}\sum_{k=1}^n x_k - \bar{\varphi}^{(1)}\left(\sum_{k=1}^n x_k\right) = \frac{1}{n}\sum_{k=1}^n x_k - \bar{\varphi}^{(n)}\left(\frac{1}{n}\sum_{k=1}^n x_k\right) = \bar{\Phi}^{(n)}\left(\frac{1}{n}\sum_{k=1}^n x_k\right). \tag{3.1.1.2}$$

Since $\bar{\Psi}^{(n)}$ and $\bar{\Phi}^{(n)}$ are inverses we obtain, by applying $\bar{\Psi}^{(n)}$ to both sides of Equation 3.1.1.2, we get that $\frac{1}{n}\sum_{k=1}^n x_k = \bar{\Psi}^{(n)}\left(\frac{1}{n}\sum_{k=1}^n \bar{\Phi}_k(x)\right)$, that is,

$$\frac{1}{n}\sum_{k=1}^n \bar{\Psi}_k(y) = \bar{\Psi}^{(n)}\left(\frac{1}{n}\sum_{k=1}^n y_k\right).$$

We can now find $\bar{\Psi}$ by

$$\bar{\Psi}_k(y) = \frac{1}{n}\sum_{j\neq k}\left(\bar{\Psi}_k(y) - \bar{\Psi}_j(y)\right) + \frac{1}{n}\sum_{j=1}^n \bar{\Psi}_j(y) = \frac{1}{n}\sum_{j\neq k}(y_k - y_j) + \bar{\Psi}^{(n)}\left(\frac{1}{n}\sum_{j=1}^n y_j\right),$$

for $k = 1, 2, \dots, n$, where all terms in the equations converge.

We have proved that there is a polydisc $P(0, r)$ around 0, and an analytic function $\Psi\colon P(0, s) \to \mathbf{K}^n$ where $P(0, s) \subseteq \Phi(P(0, r))$ which is an inverse to $\Phi|_{P(0,r)}$. The open sets $V = P(0, s)$ and $U = \Phi^{-1}(V)$ satisfy the conditions of the theorem. □

Theorem 3.1.2. (Implicit Function Theorem — Dual Form) *Let $\Phi\colon V \to \mathbf{K}^{m+n}$ be an analytic map where V is open in $\mathbf{K}^n$. Suppose that x is a point in V where $\Phi'(x)$ has rank n. Then there exist an open set U in $\mathbf{K}^{m+n}$ and an analytic function $\Psi\colon U \to \mathbf{K}^{m+n}$ such that*

(i) *$\Psi(U)$ is an open neighborhood of $\Phi(x)$ in $\mathbf{K}^{m+n}$.*
(ii) *Ψ is injective with an analytic inverse $\Psi^{-1}\colon \Psi(U) \to U$.*

(iii) *There is an* n*-dimensional linear subspace* W *in* $\mathbf{K}^{m+n}$ *such that* Ψ *gives a bijection between the sets* $\Phi(V) \cap \Psi(U)$ *and* $W \cap U$.

Proof: Renumbering the coordinates of $\mathbf{K}^{m+n}$ we may assume that the lower $n \times n$-minor of $\Phi'(a)$ is non-zero. Hence we can write Φ as (Φ_1, Φ_2), where $\Phi_1 \colon V \to \mathbf{K}^m$ and $\Phi_2 \colon V \to \mathbf{K}^n$ with $\operatorname{rank} \Phi_2'(a) = n$. Define the analytic map $\Psi \colon \mathbf{K}^m \times V \to \mathbf{K}^{m+n}$ by $\Psi(x, y) = (x + \Phi_1(y), \Phi_2(y))$, for $x \in \mathbf{K}^n$, $y \in V$. Then we have that

$$\Psi'(0, a) = \begin{pmatrix} I_m & \Phi_1'(a) \\ 0 & \Phi_2'(a) \end{pmatrix}$$

is invertible and it follows from the Inverse Function Theorem 3.1.1 that there is an analytic inverse $U' \to \mathbf{K}^m \times \mathbf{K}^n \cong \mathbf{K}^{m+n}$ defined on some neighborhood U' of $\Psi(0, a)$. Let $U = \Psi^{-1}(U')$. Since we have that $\Phi(y) = \Psi(0, y)$ for all $y \in V$, such that $(0, y) \in U$, we get that $\Phi(V) \cap \Psi(U) = \Psi(W \cap U)$, where W is the n-dimensional subspace $\{(x, y) \in \mathbf{K}^m \times \mathbf{K}^n \mid x = 0\}$ of $\mathbf{K}^{m+n}$. □

Theorem 3.1.3. (Implicit Function Theorem) *Let* U *be an open subset of* $\mathbf{K}^{m+n}$ *and let* $\Phi \colon U \to \mathbf{K}^m$ *be an analytic function. Suppose that* $x \in U$ *is a point where* $\Phi(x) = 0$ *and that* $\Phi'(x)$ *has rank* m*. Then there exist a neighborhood* V *of* x *in* U *and an analytic function* $\Psi \colon V \to \mathbf{K}^{m+n}$*, such that*

(i) $\Psi(V)$ *is open set in* $\mathbf{K}^{m+n}$.
(ii) Ψ *is injective with an analytic inverse* $\Psi^{-1} : \Psi(V) \to V$.
(iii) *There is an* n*-dimensional linear subspace* W *of* $\mathbf{K}^{m+n}$ *such that* Ψ *gives a bijection between the sets* $V \cap \Phi^{-1}(0)$ *and* $W \cap \Psi(V)$.

Proof: Renumbering the coordinates of $\mathbf{K}^{m+n}$ we may assume that the leftmost $m \times m$-minor of $\Phi'(x)$ is non-zero. Let π_1 and π_2 be the projections of $\mathbf{K}^{m+n} = \mathbf{K}^m \times \mathbf{K}^n$ to its two factors and let W be the kernel of π_1. Define $\Psi : U \to \mathbf{K}^m \times \mathbf{K}^n$ by $\Psi(y) = (\Phi(y), \pi_2(y))$, for all $y \in U$. Then Ψ is analytic and

$$\Psi'(x) = \begin{pmatrix} \Phi_1'(x) & \Phi_2'(x) \\ 0 & I_n \end{pmatrix},$$

where $\Phi_1'(x)$ and $\Phi_2'(x)$ are the $m \times m$- respectively $m \times n$-submatrices of $\Phi'(x)$ consisting of the m first, respectively n last columns of $\Phi'(x)$. In particular $\Psi'(x)$ is invertible and we can use the Inverse Function Theorem 3.1.1 to find an open subset V of U containing x and an analytic inverse function $\Psi^{-1} \colon \Psi(V) \to V$. It is clear that for y in V we have that $\Psi(y)$ is in W if and only if $\Phi(y) = 0$. □

Remark 3.1.4. We have stated the Implicit Function Theorem 3.1.3 in a slightly different way from what is usually done in that we consider the embedding of the set of zeroes of Φ into $\mathbf{K}^{m+n}$ through the analytic map Ψ. Usually, only the restriction of Ψ to the subspace W is mentioned in the Implicit Function Theorem. Of course the usual form follows from the one given above.

3.2 Matrix groups in affine space

We shall show how the Implicit Function Theorem can be used to induce a structure as manifolds on groups that are defined as the zeroes of analytic functions.

We have seen in Example 2.1.17 that $\mathrm{Gl}_n(\mathbf{K})$ is open in $\mathrm{M}_n(\mathbf{K})$. However, the subgroups $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{G}_S(\mathbf{K})$, for a matrix S are not open sets in $\mathrm{Gl}_n(\mathbf{K})$. Quite to the contrary they are zeroes of polynomials in the variables x_{ij} which are the matrix entries (see Exercise 2.1.6). We have that $\mathrm{Sl}_n(\mathbf{K})$ is the subset of $\mathrm{Gl}_n(\mathbf{K})$ which consists of the zeroes of the polynomial

$$\det(x_{ij}) = 1. \tag{3.2.0.1}$$

The set $\mathrm{G}_S(\mathbf{K})$ is the zeroes of the n^2 quadratic equations in the variables x_{ij} obtained by equating the n^2 coordinates on both sides of

$$(x_{ij})S^t(x_{ij}) = S.$$

Finally, $\mathrm{SG}_S(\mathbf{K})$ is the subset of $\mathrm{Gl}_n(\mathbf{K})$ which is the intersection of $\mathrm{G}_S(\mathbf{K})$ with the matrices satisfying Equation 3.2.0.1.

On the other hand we have that $\mathrm{Gl}_n(\mathbf{K})$ itself can be considered as the zeroes of polynomials in the space $\mathrm{M}_{n+1}(\mathbf{K})$. Indeed we have seen in Example 1.2.11 that we have an injection $\Phi\colon \mathrm{Gl}_n(\mathbf{K}) \to \mathrm{Sl}_{n+1}(\mathbf{K})$. As we just saw $\mathrm{Sl}_{n+1}(\mathbf{K})$ is the zeroes of a polynomial of degree $n+1$ in the variables x_{ij}, for $i, j = 1, \dots, n+1$, and clearly $\operatorname{im}\Phi$ is given, in $\mathrm{Sl}_{n+1}(\mathbf{K})$ by the relations $x_{1i} = x_{i1} = 0$ for $i = 2, \dots, n+1$.

3.2.1 Zeroes of analytic functions in affine space

We will now study the more general problem of a subset $Z \in \mathbf{K}^n$ which is given as the common zeroes of some collection of analytic functions defined on $\mathbf{K}^n$. The main result is that in such a set we can always find points around which Z locally looks exactly like an open set in $\mathbf{K}^m$, for some m.

Definition 3.2.1. Let U be an open subset of $\mathbf{K}^n$ and denote by $\mathcal{O}(U)$ the ring of analytic functions $f : U \to \mathbf{K}$. A subset Z of U is an *analytic set* if there is a set of analytic functions $\{f_i\}_{i\in I}$ in $\mathcal{O}(U)$ such that $Z = \{x \in U \mid f_i(x) = 0, \quad \text{for all } i \in I\}$. For an analytic set Z we define the *ideal in $\mathcal{O}(U)$ of analytic functions* vanishing on Z by

$$\mathcal{I}(Z) = \{f\colon U \to \mathbf{K} \mid f \text{ is analytic and } f(x) = 0, \quad \text{for all } x \in Z\}.$$

Furthermore, at each point $x \in Z$, we define the *normal space* by

$$N_x(Z) = \left\{ \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) \,\middle|\, f \in I(Z) \right\}.$$

Remark 3.2.2. It is clear that the *normal space* $N_x(Z)$ is a linear subspace of $\mathbf{K}^n$ whose dimension may vary over Z. Let Z_r be the set of points $x \in Z$ where $\dim_{\mathbf{K}}(N_x(Z)) \le r$.

Then Z_r is given by the points $x \in Z$ where all the determinants

$$\begin{vmatrix} \frac{\partial f_{i_1}}{\partial x_{j_1}} & \cdots & \frac{\partial f_{i_1}}{\partial x_{j_{r+1}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{i_r}}{\partial x_{j_1}} & \cdots & \frac{\partial f_{i_r}}{\partial x_{j_{r+1}}} \end{vmatrix}$$

are zero, for $i_1, i_2, \dots, i_{r+1} \in \mathcal{I}$ and $j_1, j_2, \dots, j_{r+1} \in \{1, 2, \dots, n\}$. These determinants are analytic functions and hence the set Z_r is an analytic set. In particular Z_r is closed for all integers $r = 0, 1, \dots, n$, which implies that $\dim_{\mathbf{K}} N_x(Z)$ takes its maximal value on an open subset of Z.

Theorem 3.2.3. *Let Z be an analytic subset subset of an open subset of $\mathbf{K}^n$ and let $x \in Z$ be a point where $\dim_{\mathbf{K}} N_x(Z)$ attains its maximal value m. Then there exists a neighborhood U of x in $\mathbf{K}^n$ and an analytic bijection $\Phi : V \to U$ where V is open in $\mathbf{K}^n$ such that*

(i) *$\Phi^{-1} \colon U \to V$ is analytic.*
(ii) *$Z \cap U = \Phi(V \cap W)$, where W is a linear subspace of $\mathbf{K}^n$ of dimension $n - m$.*
(iii) *If y is another point where $\dim_{\mathbf{K}} N_y(Z) = m$, and $\Psi : V' \to U'$ is the corresponding analytic function, then the function*

$$\Psi^{-1}\Phi : \Phi^{-1}(U \cap U') \to \Psi^{-1}(U \cap U'),$$

is analytic as well as its restriction to $W \cap \Phi^{-1}(U \cap U')$.

Proof: We first prove the theorem for the special case where $m = 0$. Then we have that $N_x(Z)$ is zero-dimensional for all points $x \in Z$ and it follows that for any analytic function f in $\mathcal{I}(Z)$, we have that $\partial f / \partial x_i(x) = 0$, for all $i = 1, 2, \dots, n$ and all $x \in Z$. This means that the analytic functions $\partial f / \partial x_i$, for $i = 1, 2, \dots, n$, are in $I(Z)$. By induction on $|i|$ we get that all partial derivatives $D^i f$ of f for all i in $\mathcal{I}$ are in $I(Z)$. In a neighborhood of each point $x \in Z$, we can write f as the convergent power series $f(x + h) = \sum_{i \in \mathcal{I}} D^i f(x) h^i$, which is now identically zero. Hence there is a neighborhood of x in $\mathbf{K}^n$ contained in Z which shows that Z is open in $\mathbf{K}^n$. Thus we can take $Z = U = V$, and $W = \mathbf{K}^n$, and we have proved the first two assertions of the theorem when $m = 0$.

Assume that $m > 0$. We can pick a subset $\{f_1, f_2, \dots, f_m\}$ of $I(Z)$ such that the vectors $(\partial f_i / \partial x_1(x), \partial f_i / \partial x_2(x), \dots, \partial f_i / \partial x_n(x))$, for $i = 1, 2, \dots, m$, for a basis for $N_x(Z)$.

Let Z' be the common zeroes of $f_1, f_2, \dots, f_m$. Then we have by the Implicit Function Theorem 3.1.3 applied to the analytic map $\Psi : U \to \mathbf{K}^m$ defined by $\Psi(y) = (f_(y), \dots, f_m(y))$ that there is a neighborhood U of x in $\mathbf{K}^n$ and a bijective analytic map $\Phi : V \to U$ with analytic inverse such that $Z' \cap U = \Phi(V \cap W)$, where V is open in $\mathbf{K}^n$ and $W \subseteq \mathbf{K}^n$ is a vector space of dimension $n - m$.

We have that $Z \subseteq Z'$. To prove the first two assertions of the theorem it suffices to prove that there is a neighborhood U' of x contained in U such that $Z \cap U' = Z' \cap U'$, or equivalently that the common zeroes $\Phi^{-1}(Z \cap U')$ of the functions $f\Phi|V' \cap W : V' \cap W \to \mathbf{K}$ for f in $\mathcal{I}(Z)$, with $V' = \Phi^{-1}(U')$, is the same as the common zeroes $\Phi^{-1}(Z' \cap U') = W \cap V'$ of the functions $f_i\Phi|(W \cap V') : W \cap V' \to \mathbf{K}$ for $i = 1, \dots, m$.

We have that the vector $(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x))$ is in the vector space generated by the vectors $(\frac{\partial f_i}{\partial x_1}(x), \dots, \frac{\partial f_i}{\partial x_n}(x))$ for $i = 1, \dots, m$. However, the functions $f_i\Phi|V \cap W : V \cap W \to \mathbf{K}$ are identically zero. Hence all the partial derivatives $\frac{\partial f_i \Phi}{\partial x_j}$ are zero on $V \cap W$. It follows from the case $m = 0$ that $\Phi^{-1}(Z \cap U)$ is open in W. Since $\Phi^{-1}(Z \cap U) \subseteq W \cap V$ we can find an open subset V' of V such that $W \cap V' = \Phi^{-1}(Z \cap U)$. Let $U' = \Phi(V')$. Then $Z \cap U' = Z' \cap U'$, and we have proved the two first assertions of the Theorem.

For the third assertion, we note that the composition of analytic functions, as well as the restriction to linear subspaces of analytic functions are analytic. □

Corollary 3.2.4. *Let G be one of the groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$ or $\mathrm{SG}_S(\mathbf{K})$, for a matrix S. Then, for each A in G there exists an open neighborhood U of A in $\mathrm{M}_n(\mathbf{K})$, an open set V in some affine space $\mathbf{K}^m$, depending only on the group, and an injective analytic map $\Phi\colon V \to U$, whose image is $U \cap G$.*

Moreover, if $\Psi : V' \to U'$ is another such map, then $\Psi^{-1}\Phi\colon \Phi^{-1}(U \cap U') \to \Psi^{-1}(U \cap U')$ is analytic.

Proof: By the theorem, we can find a point B of G, an open neighborhood U_B of B in $\mathrm{M}_n(\mathbf{K})$, an open set V in $\mathbf{K}^m$, and an analytic map $\Phi_B : V \to U_B$ with the properties of the corollary. For each A in G we saw in Remark 2.4.17 that we have an analytic bijection $\lambda_{AB^{-1}} : U_B \to \lambda_{AB^{-1}}(U)$ which maps B to A. Hence we can take $U = \lambda_{AB^{-1}}(U)$ and $\Phi = \lambda_{AB^{-1}}\Phi_B$ □

Exercises

3.2.1. Determine quadratic polynomials in x_{ij} that with common zeroes $\mathrm{O}_n(\mathbf{K})$ in $\mathrm{Gl}_n(\mathbf{K})$ and $\mathrm{Sp}_4(\mathbf{K})$ in $\mathrm{Gl}_4(\mathbf{K})$.

3.2.2. Use Exercise 1.4.6 to find directly maps $\mathbf{R}^1 \to \mathrm{SO}_2(\mathbf{R})$ that give bijections from open sets of $\mathbf{R}^1$ to some $U \cap \mathrm{SO}_2(\mathbf{R})$, where U is open in $\mathrm{M}_2(\mathbf{R})$.

3.3 Topological spaces

In Proposition 2.3.9 we saw that the groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$, for any invertible matrix S, and thus $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, $\mathrm{Sp}_n(\mathbf{K})$, in a natural way, can be covered by subsets that are homeomorphic to open subsets in $\mathbf{K}^m$, for some m. Like we used the algebraic properties of these groups to motivate the abstract structure of groups, rings and fields, we shall use the geometric properties to motivate the geometric structures, topology, manifold, and algebraic variety.

Definition 3.3.1. A *topological space* is a set X together with a collection of subsets $\mathcal{U} = \{U_i\}_{i \in I}$ of X satisfying the following three properties:

(i) The empty set and X are in $\mathcal{U}$.
(ii) If $\{U_i\}_{i \in J}$ is a collection of sets from $\mathcal{U}$, then the union $\bigcup_{i \in J} U_i$ is a set in $\mathcal{U}$.
(iii) If $\{U_i\}_{i \in K}$ is a finite collection of sets from $\mathcal{U}$, then the intersection $\bigcap_{i \in K} U_i$ is a set in $\mathcal{U}$.

The sets of the form U_i will be called *open* and their complement $X \setminus U_i$ will be called *closed*.

Let x be a point of X, we call an open subset of X that contains x a *neighborhood* of x.

Example 3.3.2. In Section 3.2 we have already seen one of the most important topologies on the space $X = \mathbf{K}^n$. Indeed, the subsets of $\mathbf{K}^n$ that are unions of balls, form the open sets of a topology (see Exercise 2.1.4). We call this topology on $\mathbf{K}^n$, the *metric topology* (compare Definition 2.1.10).

Example 3.3.3. Let X and Y be topological spaces given by open subsets $\{U_i\}_{i\in I}$ respectively $\{V_j\}_{j\in J}$. On the Cartesian product the collection of sets consisting of all unions of the sets in $\{U_i \times V_j\}_{(i,j)\in I\times J}$ defines a topology, called the *product topology* (see Exercise 3.3.4).

Example 3.3.4. The metric topology on the set $\mathbf{K}^n$ is the product, n times, of the metric topology on $\mathbf{K}$.

Definition 3.3.5. Let X and Y be topological spaces. A map $\Phi\colon X \to Y$ is *continuous* if, for every open subset V of Y, we have that $\Phi^{-1}(V)$ is open in X. We say that a continuous map is a *homeomorphism* if it is bijective, and the inverse is also continuous.

Example 3.3.6. We saw in Proposition 2.1.15 that, when $\mathbf{K}$ is the real or the complex numbers, the definition coincides with the usual definition of continuous maps from analysis.

Example 3.3.7. An analytic map $\Phi\colon \mathbf{K}^n \to \mathbf{K}^m$ is continuous in the metric topology. Indeed, it suffices to show that the inverse image of a polydisc $P(a,r)$ in $\mathbf{K}^m$ is open in $\mathbf{K}^n$, that is, there is a polydisc around every point b in the inverse image that is contained in the inverse image. Let $\Phi = (\Phi_1, \dots, \Phi_m)$. Then $\Phi^{-1}(P(a,r)) = \bigcap_{i=1}^m \Phi_i^{-1}(P(a_i,r_i))$. Consequently, it suffices to prove that the map is continuous when $m = 1$. With $m = 1$ and $\Phi = \Phi_1$, let b in $\mathbf{K}^n$ a point such that $\Phi(b) = a$. It follows from Definition 2.4.11 that we have $\Phi(a+x) = \Phi(a) + \Phi'(a)(x) + r(x)$, for x in some polydisc $P(0,s)$ in $\mathbf{K}^n$, where $r(x)$ is analytic in the polydisc, and where $\lim_{x\to 0} \frac{\|r(x)\|}{\|x\|} = 0$. Hence, by choosing $\|x\|$ small we can make $\|\Phi(a+x) - \Phi(a)\|$ as small as we like, and Φ is continuous.

Example 3.3.8. Let $\Phi\colon [0,2\pi) \to \{z \in \mathbf{C}\colon |z| = 1\}$ be the map defined by $\Phi(x) = e^{ix}$. Then Φ is continuous and bijective. However, it is not a homeomorphism because the image of the open subset $[0,\pi)$ of $[0,2\pi)$ is the upper half circle plus the point $(1,0)$, which is not open in the circle. Hence, the inverse map is not continuous.

Definition 3.3.9. Let Y be a subset of a topological space X and $\{U_i\}_{i\in I}$ the open subsets of X. Then the sets $\{Y \cap U_i\}_{i\in I}$ are the open sets of a topology of Y which we call the *induced topology.*

Example 3.3.10. We saw in Corollary 3.2.4 that the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$, for all invertible matrices S considered as subsets of $\mathrm{M}_n(\mathbf{K})$, are covered by sets that are in bijective correspondence, via analytic maps, with balls in affine spaces. We also saw that these sets can be taken to be the intersection of the group with a ball in $\mathrm{M}_n(\mathbf{K})$. Consequently, these subsets are open sets in the the topology induced by the metric topology on $\mathrm{M}_n(\mathbf{K})$, and given a point x in one of the groups G and an open set U of G in the induced topology, then there is an open subset V of U, obtained as in Corollary 3.2.4, such that $x \in V \subseteq U$.

Exercises

3.3.1. A topological space X is *Hausdorff*, if, given two points x and y of X, there are open neighborhoods of x and y that do not intersect. Show that every metric topology is Hausdorff.

3.3.2. Let $X = \mathbf{K}^n$. Show that the two metrics, associated by Definition 2.1.10 to the norms of Definition 2.1.6 and Example 2.1.7, define the same topology on X.

3.3.3. Let X be a set. Show that the family of all finite subsets of X, together with X itself and the empty set, are the closed sets of a topology. We call this topology the *finite topology*. Show that the finite topology is not Hausdorff.

3.3.4. Let X and Y be topological spaces given by open subsets $\{U_i\}_{i\in I}$ respectively $\{V_j\}_{j\in J}$. Show that the collection of sets consisting of all unions of the sets in $\{U_i \times V_j\}_{(i,j)\in I\times J}$ defines a topology on the Cartesian product.

3.3.5. Let $X = \mathbf{Z}$ and for $a \in \mathbf{Z} \setminus \{0\}$ and $b \in \mathbf{Z}$ define $X_{a,b} = \{ax + b \mid x \in \mathbf{Z}\}$. Let $\mathcal{U}$ consist of all unions of sets of the form $X_{a,b}$.

(a) Show that $\mathcal{U}$ is a topology on X.
(b) Show that all the sets $X_{a,b}$, for $a, b \in \mathbf{Z}$, are both open and closed.
(c) Let $P \subseteq \mathbf{Z}$ be the set of prime numbers. Show that $\bigcup_{p\in P} X_{p,0} = X \setminus \{1, -1\}$.
(d) Show that $\{-1, 1\}$ is not open, and that this implies that P is infinite.

3.3.6. Let R be a commutative ring. An ideal P in R is a *prime ideal* if $ab \in P$ implies that $a \in P$ or $b \in P$. Let

$$\operatorname{Spec}(R) = \{P \subseteq R \mid P \text{ is a prime ideal}\}.$$

For any ideal I in R, let $\mathcal{V}(I) = \{P \in \operatorname{Spec}(R) \mid I \subseteq P\}$.

(a) Show that the sets $\operatorname{Spec}(R) \setminus \mathcal{V}(I)$ form the open sets of a topology on $\operatorname{Spec}(R)$.
(b) Show that a one point set $\{P\}$ in $\operatorname{Spec}(R)$ is closed if and only if P is a *maximal ideal*, that is an ideal that is not contained in any other ideal different from R.

3.3.7. For every ideal I in the polynomial ring $\mathbf{K}[x_1, \dots, x_n]$ in n variables $x_1, \dots, x_n$ with coefficients in an arbitrary field $\mathbf{K}$ we write $\mathcal{V}(I) = \{(a_1, \dots, a_n) \in \mathbf{K}^n \mid f(a_1, \dots, a_n) = 0 \text{ for all } f \text{ in } I\}$.

(a) Show that the sets $\mathbf{K}^n \setminus \mathcal{V}(I)$ for all ideals I form the open sets of a topology on $\mathbf{K}^n$.
(b) Show that when $\mathbf{K}$ has a finite number of elements all subsets of $\mathbf{K}$ are open and closed.
(c) Show that when $\mathbf{K}$ is infinite any two open sets intersect. In particular, the space $\mathbf{K}^n$ is not Hausdorff.

3.4 Manifolds

In Remark 2.4.17 we summarized certain properties of the matrix groups under the term manifold. The same properties that we used were also stated in Corollary 3.2.4. In this section we introduce manifolds and show how Remark 2.4.17 and Corollary 3.2.4 are reinterpreted in the language of manifolds.

Definition 3.4.1. Let X be a topological space. A *chart* of X of *dimension* n consists of an open set U of X, an open subset V in $\mathbf{K}^n$ with the metric topology, and a homeomorphism $\Phi\colon V \to U$. A family of charts $\{(\Phi_i, V_i, U_i)\}_{i\in I}$ is called an *atlas* if the open sets $\{U_i\}_{i\in I}$ cover X and if the map $\Phi_j^{-1}\Phi_i\colon \Phi_i^{-1}(U_i \cap U_j) \to \Phi_j^{-1}(U_i \cap U_j)$ is analytic for all i and j in I, when $U_i \cap U_j$ is non-empty. Here, and in the following, we write, for simplicity, $\Phi_j^{-1}\Phi_i$ for the map

$$\Phi_j^{-1}|_{(U_i\cap U_j)}\Phi_i|_{\Phi_i^{-1}(U_i\cap U_j)}.$$

The set where $\Phi_j^{-1}\Phi_i$ is defined will be clear from the context.

A topological space M together with an atlas of equal-dimensional charts is called an *analytic manifold*. It is often convenient to include in the atlas all the homeomorphisms $\Phi\colon V \to U$, from an open subset in $\mathbf{K}^n$ to an open subset in X, such that, for all $x \in U$ and some U_i in the chart that contains x, we have that $\Phi_i^{-1}\Phi$ is analytic on $\Phi^{-1}(U \cap U_i)$ and $\Phi^{-1}\Phi_i$ is analytic on $\Phi_i^{-1}(U \cap U_i)$. Clearly the maps $\Phi_j^{-1}\Phi$ and $\Phi^{-1}\Phi_j$ are then analytic on $\Phi^{-1}(U \cap U_j)$ respectively on $\Phi_j^{-1}(U \cap U_j)$ for all j in I such that $U \cap U_j$ is non-empty. Such a maximal chart is called an *analytic structure*.

For each open subset U of M the charts $\Phi_i\colon \Phi_i^{-1}(U \cap U_i) \to U \cap U_i$ define a structure as manifold on U, called the *induced structure*.

3.4.2. The number n that appear in the definition of a manifold is uniquely determined by the analytic structure, in the sense that if $\Phi\colon U \to M$ is a homeomorphism from an open set in $\mathbf{K}^m$ to an open subset of M, such that for all x in U and some member U_i of a chart that contains x, we have that $\Phi_i^{-1}\Phi$ and $\Phi^{-1}\Phi_i$ are analytic on $\Phi^{-1}(U \cap U_i)$, respectively on $\Phi_i^{-1}(U \cap U_i)$ then $m = n$. Indeed, it follows from Equation 2.4.14.1 that $I_m = ((\Phi^{-1}\Phi_i)(\Phi_i^{-1}\Phi))' = (\Phi^{-1}\Phi_i)'(\Phi_i^{-1}(x))(\Phi_i^{-1}\Phi)'(\Phi^{-1}(x))$. Hence we have that $m \leq n$ (see Exercise 3.4.1). A similar reasoning shows that $m \geq n$. Hence the linear maps $(\Phi\Phi_i)'(\Phi_i^{-1}(x))$ and $(\Phi_i^{-1}\Phi)'(\Phi^{-1}(x))$ are both invertible and we have that $m = n$.

Definition 3.4.3. The number n that appear in Definition 3.4.1 is called the *dimension* of M and denoted by $\dim M$.

Example 3.4.4. The space $\mathbf{K}^n$ is a manifold. A chart is $\mathbf{K}^n$ itself with the identity map.

Example 3.4.5. Clearly Corollary 3.2.4 states that the topological spaces $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$ are analytic manifolds. In particular, we have that $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, $\mathrm{Sp}_n(\mathbf{K})$ are analytic manifolds.

Example 3.4.6. The homeomorphism $\mathbf{R}^2 \to \mathbf{C}$ mapping (a, b) to $a+bi$ defines a structure on $\mathbf{C}$ as a real manifold of dimension 2. Similarly the homeomorphism $\mathbf{R}^4 \to \mathbf{H}$ mapping (a, b, c, d) to $a + ib + jc + kd$ (see Example 1.3.13) defines a structure of a real manifold of dimension 4 on the quaternions $\mathbf{H}$.

Example 3.4.7. Let M and N be manifolds defined by charts $\{(\Phi_i, U_i)\}_{i\in I}$ respectively $\{(\Psi_j, V_j)\}_{j\in J}$. We give the *Cartesian product* $M \times N$ product topology (see Example 3.3.3). The maps $\Phi_i \times \Psi_j\colon U_i \times V_j \to M \times N$ clearly are homeomorphisms of topological spaces. Moreover, these maps define a chart on $M \times N$ because if $\Phi \times \Psi\colon U \times V \to M \times N$ is another one of these homeomorphisms, then the map $(\Phi_i \times \Psi_j)(\Phi \times \Psi)^{-1}$ on $(U_i \times V_j) \cap (\Phi_i \times \Psi_j)^{-1}((\Phi \times \Psi)(U \times V) = (U \cap \Phi_i^{-1}\Phi(U) \times (V_j \cap \Psi_j^{-1}\Psi(V))$ is given by the analytic map $\Phi_i\Phi^{-1} \times \Psi_j\Psi^{-1}$. In this way $M \times N$ becomes a manifold which we call the *product manifold*.

Definition 3.4.8. Let M be an analytic manifold and U an open subset. A function $f\colon U \to \mathbf{K}$ is *analytic* if for every x in U and some chart $\Phi_i\colon V_i \to U_i$, where x is contained in U_i, we have that the map $f\Phi_i$ is analytic on $\Phi_i^{-1}(U \cap U_i)$. The condition then holds for all charts $\Phi_j : V_j \to U_j$. We denote by $\mathcal{O}_M(U)$ the set of all analytic functions on U.

Remark 3.4.9. The set $\mathcal{O}_M(U)$ is clearly a ring, and for an open subset V of M contained in U there is a natural ring homomorphism $\rho_{U,V}\colon \mathcal{O}_M(U) \to \mathcal{O}_M(V)$ sending a function f to its restriction $f|_V$. The following two fundamental properties hold:

(i) If $f \in \mathcal{O}_M(U)$ and there is an open cover $\{U_i\}_{i\in I}$ of U such that $\rho_{U,U_i}(f) = 0$, for all $i \in I$, we have that $f = 0$.

(ii) If $\{U_i\}_{i\in I}$ is an open covering of U and $\{f_i\}_{i\in I}$ is a collection of functions $f_i \in \mathcal{O}_M(U_i)$ such that $\rho_{U_i,U_i\cap U_j}(f_i) = \rho_{U_j,U_i\cap U_j}(f_j)$, for all i and j, there is a function $f \in \mathcal{O}_M(U)$ such that $\rho_{U,U_i}(f) = f_i$, for all $i \in I$.

We summarize these properties by saying that $\mathcal{O}_M$ is a *sheaf* on M.

Definition 3.4.10. Let N and M be analytic manifolds and $\Phi\colon N \to M$ a continuous map. We say that Φ is *analytic* if, for every open subset U of M and every analytic function$f\colon U \to \mathbf{K}$ on U, we have that $f\Phi$ is analytic on $\Phi^{-1}(U)$. When Φ has an analytic inverse, we say that Φ is an *isomorphism* of manifolds.

Remark 3.4.11. It follows immediately from the definition that if $\Psi\colon P \to N$ is another analytic map of manifolds, then the composite $\Psi\Phi\colon P \to M$ is also analytic.

Let X be a topological space and U an open subset. We denote by $\mathcal{C}_X(U)$ the ring of all continuous functions $U \to \mathbf{K}$. A continuous map $\Phi\colon N \to M$ of topological spaces induces, for all open subsets U of M, a ring homomorphism $\mathcal{C}_M(U) \to \mathcal{C}_N(f^{-1}(U))$, which sends a function $f\colon U \to \mathbf{K}$ to the composite $f\Phi\colon \Phi^{-1}(U) \to \mathbf{K}$. When M and N are analytic manifolds, the map $f\Phi$ is analytic, by definition, if and only if it induces a map $\Phi^*(U)\colon \mathcal{O}_M(U) \to \mathcal{O}_N(\Phi^{-1}(U))$ on the subrings of analytic functions. Clearly $\Phi^*(U)$ is a ring homomorphism and, when V is an open subset of U we have that the diagram

$$\begin{array}{ccc} \mathcal{O}_M(U) & \xrightarrow{\Phi^*(U)} & \mathcal{O}_N(\Phi^{-1}(U)) \\ {\scriptstyle \rho_{U,V}}\downarrow & & \downarrow{\scriptstyle \rho_{\Phi^{-1}(U),\Phi^{-1}(V)}} \\ \mathcal{O}_M(V) & \xrightarrow{\Phi^*(V)} & \mathcal{O}_N(\Phi^{-1}(V)) \end{array} \tag{3.4.11.1}$$

is commutative.

Remark 3.4.12. When M and N are open subsets of $\mathbf{K}^m$ respectively $\mathbf{K}^n$, with the induced manifold structures, we have that a map $\Phi\colon N \to M$ is an analytic map of manifolds if and only if it is an analytic map from an open subset of $\mathbf{K}^m$ to $\mathbf{K}^n$, in the sense of Definition 2.4.6. Indeed, the two notions clearly coincide when $M = \mathbf{K}$, and since composition of analytic functions in the sense of Definition 2.4.6 is again analytic in the same sense, we have that if a function is analytic as in Definition 2.4.6, it is an analytic map of manifolds. Conversely, write $\Phi = (\Phi_1, \dots, \Phi_m)$ where $\Phi_i : N \to \mathbf{K}^m$ are analytic maps. The *coordinate functions* $x_i \colon M \to \mathbf{K}$ defined by $x_i(a_1, \dots, a_m) = a_i$ are clearly analytic according to both definitions. Hence $x_i\Phi = \Phi_i$ is analytic, for $i = 1, \dots, m$. Consequently Φ is analytic in the sense of Definition 2.4.6.

Example 3.4.13. It follows from Corollary 3.2.4 that the inclusion map into $\mathrm{M}_n(\mathbf{K})$ of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$. In particular $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, $\mathrm{Sp}_n(\mathbf{K})$ are analytic manifolds.

Example 3.4.14. The group homomorphisms of Examples 1.2.10, 1.2.11, and 1.2.12 are all analytic.

Example 3.4.15. Let M and N be manifolds. Then the maps $\pi_1 \colon M \times N \to M$ and $\pi_2 \colon M \times N \to N$, from the product manifold to the *factors* are analytic maps. We call π_1 and π_2 the *projection* to the first, respectively second, factor.

Exercises

3.4.1. Show that if A is an $m \times n$-matrix and B is an $n \times m$-matrix such that $AB = I_m$, then $m \leq n$.

3.5 Equivalence relations and applications

Equivalence relations are fundamental in all parts of mathematics. Here we shall define equivalence relations and give some important examples. The reason that we introduce the material at this point is that it is used to define the ring of germs of analytic functions at a point, which is very convenient for the treatment of tangent spaces of analytic sets.

Definition 3.5.1. A *partition* of a set S is a family of *disjoint* subsets $\{S_i\}_{i \in I}$ of S that *cover* S. That is

$$S_i \cap S_j = \emptyset, \quad \text{if } i \neq j$$

and

$$S = \bigcup_{i \in I} S_i.$$

A *relation* on the set S is a subset T of $S \times S$. If (x, y) is in T we write $x \equiv y$ and say that x and y are related. We say that the relation $\equiv$ is an *equivalence relation* if the following three properties hold, for all x, y and z in S:

(i) (reflexivity) $x \equiv x$,
(ii) (symmetry) if $x \equiv y$, then $y \equiv x$,
(iii) (transitivity) if $x \equiv y$ and $y \equiv z$, then $x \equiv z$.

Given a partition $\{S_i\}_{i \in I}$ of a set S, we obtain an equivalence relation on S by defining x to be related to y if x and y lie in the same subset S_i for some i. Conversely, given an equivalence relation $\equiv$ on a set S we obtain a partition $\{S_i\}_{i \in I}$ of S as follows:

For each x in S let $S_x = \{y \in S\colon y \equiv x\}$ be the set of all elements in S related to x. Then we have that $x \in S_x$, and $S_x = S_y$ if and only if $x \equiv y$. Let $I = S/\equiv$ be the set whose elements are the different sets S_x. For x in S we write $[x]$ for the element of $S/\equiv$ corresponding to the set S_x. Then $[x] = [y]$ if and only if $x \equiv y$, and each i in $S/\equiv$ is of the form $[x]$ for some x in S. For i in $S/\equiv$ we let S_i be the set S_x for any x such that $i = [x]$.

Given a *multiplication* on S, that is, a map

$$S \times S \to S,$$

and denote by xy the image of (x, y) by this map. If, for all elements x, y and z of S such that $x \equiv y$, we have that $xz \equiv yz$ and $zx \equiv zy$, we obtain a multiplication

$$(S/\equiv) \times (S/\equiv) \to S/\equiv,$$

defined by $[x][y] = [xy]$. Indeed, if $[x] = [x']$ and $[y] = [y']$, we have that $xy \equiv x'y \equiv x'y'$, and consequently that $[xy] = [x'y']$.

Example 3.5.2. Let G be a group and H a subgroup. Define a relation on G by $a \equiv b$ if $ab^{-1} \in H$. This is an equivalence relation. Indeed, it is reflexive because $aa^{-1} = e \in H$, symmetric because, if $ab^{-1} \in H$, then $ba^{-1} = (ab^{-1})^{-1} \in H$, and transitive because if $ab^{-1} \in H$ and $bc^{-1} \in H$, then $ac^{-1} = ab^{-1}(bc^{-1}) \in H$. We write $G/H = G/\equiv$. If H is a normal subgroup of G, we have that G/H has a multiplication. Indeed, if $a \equiv b$, then $ca \equiv cb$ and $ac \equiv bc$, because $ca(cb)^{-1} = cab^{-1}c^{-1} \in H$ and $ac(bc)^{-1} = ab^{-1} \in H$. It is easily checked that, with this multiplication, G/H is a group with unit $[e]$. Moreover, the canonical map

$$G \to G/H,$$

that sends a to $[a]$ is a group homomorphism with kernel H. We call the group G/H the *residue group* of G with respect to H (see Exercise 3.5.4).

Let R be a commutative ring and $I \subseteq R$ an ideal (see Definition 1.3.1). Let R/I be the residue group. The multiplication on R induces a multiplication

$$R/I \times R/I \to R/I$$

on R/I, which sends $([a], [b])$ to $[ab]$. With this multiplication R/I becomes a ring, and the map

$$R \to R/I$$

is a ring homomorphism with kernel I. We call R/I the *residue ring* of R with respect to I (see Exercise 3.5.5).

The best known case of a residue ring is the residue ring $\mathbf{Z}/n\mathbf{Z}$ of $\mathbf{Z}$ with respect to the ideal $n\mathbf{Z} = \{m \in \mathbf{Z}\colon n|m\}$ (see Exercises 3.5.1 and 3.5.2).

Example 3.5.3. Let $S = \mathbf{K}^{n+1} \setminus (0)$. Defining $(a_0, \dots, a_n)$ and $(b_0, \dots, b_n)$ to be related, if there is a non-zero element a of $\mathbf{K}$ such that $a_i = ab_i$, for $i = 0, \dots, n$, we obtain a relation on S. This relation clearly is an equivalence relation. The set $(\mathbf{K}^{n+1} \setminus (0))/\equiv$ is denoted $\mathbf{P}^n(\mathbf{K})$, and is called the *projective space* of dimension n over $\mathbf{K}$. We have a canonical map

$$\Phi\colon \mathbf{K}^{n+1} \setminus \{0\} \to \mathbf{P}^n(\mathbf{K}).$$

The sets U in $\mathbf{P}^n(\mathbf{K})$ such that $\Phi^{-1}(U)$ is open in the metric topology on $\mathbf{K}^{n+1} \setminus \{0\}$, are the open sets in a topology on $\mathbf{P}^n(\mathbf{K})$. By definition, the map Φ is continuous with respect to this topology and the metric topology on $\mathbf{K}^n \setminus \{0\}$.

For $i = 0, \dots, n$ we denote by H_i the subset of $\mathbf{P}^n(\mathbf{K})$ consisting of points of the form $[(a_0, \dots, a_{i-1}, 0, a_{i+1}, \dots, a_n)]$. Then H_i is closed in the topology of $\mathbf{P}^n(\mathbf{K})$. Let $U_i = \mathbf{P}^n(\mathbf{K}) \setminus H_i$. Then the sets U_i, for $i = 0, \dots n$, form an open covering of $\mathbf{P}^n(\mathbf{K})$. Let

$$\Phi_i \colon \mathbf{K}^n \to \mathbf{P}^n(\mathbf{K})$$

be the map defined by $\Phi_i(a_1, \dots, a_n) = [(a_1, \dots, a_{i-1}, 1, a_i, \dots, a_n)]$. Then Φ_i is a homeomorphism of $\mathbf{K}^n$ onto the open subset U_i of $\mathbf{P}^n(\mathbf{K})$. We have that the map $\Phi_j^{-1}\Phi_i$ is defined on the set $\Phi_i^{-1}(U_i \cap U_j)$ and is given by $\Phi_j^{-1}\Phi_i(a_1, \dots, a_n) = (\frac{a_1}{a_j}, \dots, \frac{a_{j-1}}{a_j}, \frac{a_{j+1}}{a_j}, \dots, \frac{a_n}{a_j})$, where $a_j \neq 0$ because $\Phi_i(a_1, \dots, a_n)$ is in $U_i \cap U_j$. We see that (U_i, Φ_i), for $i = 0, \dots, n$ define a chart on $\mathbf{P}^n(\mathbf{K})$, which makes $\mathbf{P}^n(\mathbf{K})$ into a manifold over $\mathbf{K}$ of dimension n.

Example 3.5.4. Let M be a manifold and x a point of M. Let S be the set consisting of pairs (U, f), where U is an open neighborhood of x and f an analytic function on U. We give a relation on S by defining (U, f) to be related to (V, g) if there is an open neighborhood W of x, contained in $U \cap V$ such that $f|_W = g|_W$. Clearly this relation is an equivalence relation. The residual set $S/\equiv$ is denoted by $\mathcal{O}_{M,x}$. The elements of $\mathcal{O}_{M,x}$ can be added and multiplied by the rules $[(U, f)] + [(V, g)] = [(U \cap V, (f + g)|_{U\cap V})]$ and $[(U, f)][(V, g)] = [(U \cap V, (fg)|_{U\cap V})]$. Clearly $O_{M,x}$ becomes a ring with this addition and multiplication, zero being the element $[(M, 0)]$ and the unity the element $[(M, 1)]$.

For every open neighborhood U of x we obtain a ring homomorphism

$$\mathcal{O}_M(U) \to \mathcal{O}_{M,x},$$

sending f to $[(U, f)]$. The ring $\mathcal{O}_{M,x}$ is called the *ring of germs* of analytic functions at x. When g is a function which is analytic in a neighborhood U of x we have that $[(U, g)] = [(V, g)]$ for any neighborhood V of x contained in U. We often write g for the elements $[(U, g)]$ when there can be no confusion as to whether g is considered in $\mathcal{O}_M(U)$ or in $\mathcal{O}_{M,x}$. There is also a ring homomorphism

$$\mathcal{O}_{M,x} \to \mathbf{K},$$

mapping f to $f(x)$. This map is called the *augmentation map* at x.

Given an analytic map $\Phi \colon N \to M$ of analytic manifolds, we have a natural ring homomorphism

$$\Phi_x^* \colon \mathcal{O}_{M,\Phi(x)} \to \mathcal{O}_{N,x}$$

defined by $\Phi_x^*[(U, f)] = [(\Phi^{-1}(U), f\Phi)]$.

Exercises

3.5.1. Show that the ring $\mathbf{Z}/n\mathbf{Z}$ has n elements.

3.5.2. Show that $\mathbf{Z}/n\mathbf{Z}$ is a field if and only if n is a prime number.

3.5.3. Show that R/I is a field, if and only if I is not contained in any other ideal.

3.5.4. Let H be an invariant subgroup of a group G. Show that the product $[a][b] = [ab]$ is well defined for all a and b in G/H and that G/H with this product is a group. Also, show that the map $G \to G/H$ that sends a to $[a]$ is a groups homomorphism.

3.5.5. Let R be a commutative ring and $I \subseteq R$ an ideal. Show that the multiplication on R induces a multiplication

$$R/I \times R/I \to R/I$$

on R/I, which sends $([a], [b])$ to $[ab]$. Moreover, show that with this multiplication R/I becomes a ring, and the map

$$R \to R/I$$

is a ring homomorphism with kernel I.

3.5.6. Let $X = \mathbf{R}$, and for each open set $U \in X$, let $\mathcal{O}_X(U)$ be the set of all functions $f : U \to \mathbf{R}$, such that for all $x \in U$, there exist two polynomials g and h and a neighborhood V of x in U, with the property that for all $y \in V$

$$f(y) = \begin{cases} g(y), & \text{if } y \leq x, \\ h(y) & \text{if } y \geq x. \end{cases}$$

Hence $\mathcal{O}_X(X)$ consists of all piecewise polynomial functions on X.

(a) Show that $\mathcal{O}_X$ is a sheaf of rings on X.
(b) Determine the ring of germs $\mathcal{O}_{X,x}$ for $x \in X$.
(c) Determine the tangent space $T_x(X)$, that is, the vector space of derivations $D : \mathcal{O}_{X,x} \to \mathbf{R}$, with respect to the augmentation map $\varphi : \mathcal{O}_{X,x} \to \mathbf{R}$, sending f to $f(x)$.
(d) Determine the set of vector fields Z on X, that is, the set of derivations $Z_U : \mathcal{O}_X(U) \to \mathcal{O}_X(U)$, with respect to the identity map $\mathcal{O}_X(U) \to \mathcal{O}_X(U)$, that commute with the restriction maps $\rho_{U,V} : \mathcal{O}_X(U) \to \mathcal{O}_X(V)$.
(e) Determine the set of left-invariant vector fields on X, if the group operation on X is addition.
(f) Determine the set of left-invariant vector fields on $X \setminus \{0\}$, if the group operation on $X \setminus \{0\}$ is multiplication.

3.5.7. Let X be a topological space, and let $\mathbf{K}$ be a field which is a topological space, for example $\mathbf{C}$ with the metric topology. For every open subset of X we define

$$\mathcal{O}_X(U) = \{f : U \to \mathbf{K} \mid f \text{ is continuous}\}.$$

(a) Show that, with the obvious restriction maps, we have that $\mathcal{O}_X$ is a sheaf.
(b) Show that if $\mathbf{K}$ has the discrete topology, then $\mathcal{O}_X$ is a sheaf of rings.
(c) Compute the stalks $\mathcal{O}_{X,x}$ in the case when $\mathbf{K}$ has the discrete topology.

3.5.8. Let $S = \mathbf{K} \times \mathbf{K} \setminus \{(0, 0)\}$. We write $(x, y) \equiv (x', y')$ if $xy' = x'y$.

(a) Show that $\equiv$ is an equivalence relation on S.

(b) Define the *projective line* $\mathbf{P}^1_{\mathbf{K}}$ to be $S/\equiv$. Let Φ and Ψ be maps $\mathbf{K} \to \mathbf{P}^1_{\mathbf{K}}$ be defined by $\Phi(x) = (x, 1)$ respectively $\Psi(x) = (1, x)$, for all x in $\mathbf{K}$ and let $U = \operatorname{im}\Phi$ and $V = \operatorname{im}\Psi$. Moreover let $\mathcal{U}$ be the subsets W of $\mathbf{P}^1_{\mathbf{K}}$ such that $\Phi^{-1}(W)$ and $\Psi^{-1}(W)$ are open. Show that $\{(U, \mathbf{K}, \Phi), (V, \mathbf{K}, \Psi)\}$ is an atlas on $\mathbf{P}^1_{\mathbf{K}}$, which defines $\mathbf{P}^1_{\mathbf{K}}$ as an analytic manifold.

(c) Show that the ring $\mathcal{O}_{\mathbf{P}^{-1}_{\mathbf{R}}}$ of analytic functions on $\mathbf{P}^1_{\mathbf{R}}$ is isomorphic to the ring of analytic functions $f : \mathbf{R} \to \mathbf{R}$ such that $\lim_{x \to +\infty} f(x) = \lim_{x \to -\infty} f(x)$.

(d)) Show that $\mathbf{P}^1_{\mathbf{R}}$ is isomorphic to $\mathrm{SO}_2((R)$ as an analytic manifold. [(e)]Show that the ring $\mathcal{O}_{\mathbf{P}^1_{\mathbf{C}}}(\mathbf{P}^1_{\mathbf{C}})$ of analytic functions on $\mathbf{P}^1_{\mathbf{C}}$ consists entirely of constant functions.
Hint: Use Liouville's Theorem.

3.6 Tangent spaces

In this section we shall introduce the tangent space of an analytic manifold. We start by studying the tangent vectors to curves in $\mathbf{K}^n$ in order to motivate the definitions.

3.6.1. A *curve* in $\mathbf{K}^n$ is an analytic map $\gamma : U \to \mathbf{K}^n$ on a ball U of $\mathbf{K}$ (see Definition 2.5.1). Let $c \in U$. Then the curve passes through $y = \gamma(c)$. The *tangent* to the curve at y is the derivative $\gamma'(c)$ of γ at c (see Definition 2.4.11). Each vector v of $V^n_{\mathbf{K}}$ is the derivative of the curve $\gamma\colon \mathbf{K} \to \mathbf{K}^n$ through y, defined by $\gamma(t) = y + tv$.

Let $\gamma : U \to \mathbf{K}^n$ be a curve with tangent $v = \gamma'(c)$ at c. We obtain a linear homomorphism

$$D_v\colon \mathcal{O}_{\mathbf{K}^n, y} \to \mathbf{K}$$

which maps an element $[(V, f)]$ to the derivative $(f\gamma)'(c)$ at c of the composite map $f\gamma\colon U \cap \gamma^{-1}(V) \to \mathbf{K}$. We write $D_v(f) = (f\gamma)'(c)$. It follows from the Formula 2.4.14.1 that

$$D_v(f) = f'(y)\gamma'(c) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(y)\gamma_i'(c).$$

In particular, the map D_v depends only on the tangent vector $v = \gamma'(c)$. Let $[(W, g)]$ be another element of $\mathcal{O}_{\mathbf{K}^n, y}$. From the derivation rules for analytic functions in one variable, applied to $f\gamma|_{V \cap W}$ and $g\gamma|_{V \cap W}$, we obtain that the function D_v is a $\mathbf{K}$-linear map and that

$$D_v(fg) = f(y)D_v g + g(y)D_v f.$$

Definition 3.6.2. Let $\mathbf{K}$ be any field, and let R and S be K-algebras. Moreover let $\varphi\colon S \to R$ be a ring homomorphism which is the identity on $\mathbf{K}$. Such a map is called a $\mathbf{K}$-*algebra homomorphism*. A linear map

$$D\colon S \to R$$

such that

$$D(ab) = \varphi(a)Db + \varphi(b)Da,$$

for all elements a and b of S, is called a *derivation* with respect to φ.

3.6.3. With this terminology D_v is a derivation on $\mathcal{O}_{\mathbf{K}^n, y}$, with respect to the augmentation map.

Conversely, given a **K**-linear map

$$D\colon \mathcal{O}_{\mathbf{K}^n,y} \to \mathbf{K},$$

which is a derivation for the augmentation map. There is a unique vector v such that $D = D_v$. Indeed, let x_i be the coordinate function in $\mathcal{O}_{\mathbf{K}^n,y}$ defined by $[(\mathbf{K}^n, x_i)]$, where $x_i(a_1, \dots, a_n) = a_i - y_i$. Let $[(U, f)]$ be in $\mathcal{O}_{\mathbf{K}^n,y}$. It follows from Remark 2.4.14 that

$$f(x) = f(y) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(y)x_i(x) + \sum_{i=1}^{n}\sum_{j=1}^{n} x_i(x)x_j(x)g_{ij}(x), \qquad \text{for all } x \text{ in } U,$$

where the g_{ij} are analytic functions on U. Since D is a derivation with respect to the augmentation map, we obtain that $D(1) = D(1 \cdot 1) = 1D(1) + 1D(1)$, which implies that $D(1) = 0$. Moreover, $D(x_ix_jg_{ij}) = x_j(y)g_{ij}(y)D(x_i)+x_i(y)g_{ij}(y)D(x_j)+x_i(y)x_j(y)D(g_{ij}) = 0$. Thus we get that

$$Df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(y)D(x_i) = D_v f,$$

where $v = (D(x_1), D(x_2), \dots, D(x_n))$ is the tangent vector of the curve $\gamma\colon \mathbf{K} \to \mathbf{K}^n$, defined by $\gamma(t) = y + tv$.

From the above considerations it is natural to make the following definition:

Definition 3.6.4. Let M be a manifold, and x a point of M. The *tangent space* $T_x(M)$ of M at x is the space of derivations $\mathcal{O}_{M,x} \to \mathbf{K}$, with respect to the augmentation map.

Example 3.6.5. Let y be a point of $\mathbf{K}^n$. Then it follows from Paragraph 3.6.3 that $T_y(\mathbf{K}^n)$ is a vector space of dimension n and a basis is given by the derivations $D_1, D_2, \dots, D_n$ defined by

$$D_i(x_j) = \delta_{ij}, \qquad \text{for } 1 \le i, j \le n,$$

where $x_1, x_2, \dots, x_n$ are the coordinate functions in $\mathcal{O}_{\mathbf{K}^n,y}$ with respect to the standard basis of $\mathbf{K}^n$. We sometimes write $D_i = \partial/\partial x_i$. It will cause no confusion that we write $\frac{\partial f}{\partial x_i}(x)$ for the partial derivative of f at x, when we consider $\frac{\partial}{\partial x_i}$ as a partial derivative on $\mathbf{K}^n$, and that we write $\frac{\partial f}{\partial x_i}$ when $\frac{\partial}{\partial x_i}$ is considered as a tangent $\mathcal{O}_{\mathbf{K}^n,x} \to \mathbf{K}$ of $\mathbf{K}^n$ at x.

Example 3.6.6. Let N be a manifold and U an open subset with the induced topology. Then clearly U is a manifold and $\mathcal{O}_{U,x} = \mathcal{O}_{M,x}$, for all x in U. Hence we have that $T_x(U) = T_x(N)$.

3.6.7. The advantage of Definition 3.6.4 to that of Section 2.5 is that it is independent of chaise of charts. On the other hand, the advantage of the considerations of Section 2.5 is that they give an explicit description of the tangent space as vectors in the space $\mathbf{K}^n$. In particular, it follows from the above description that $T_x(M)$ is a vector space of dimension equal to $\dim M$. To be more precise, let $\Phi\colon V \to M$ be a chart with V open in $\mathbf{K}^n$ and let $U = \Phi(V)$. Then, for $y = \Phi^{-1}(x) \in V$, we have an isomorphism of rings

$$\Phi_x^*\colon \mathcal{O}_{M,x} \to \mathcal{O}_{V,\Phi^{-1}(x)},$$

(see Example 3.5.4), and consequently an isomorphism

$$T_{\Phi^{-1}(x)}(V) \to T_x(M)$$

of tangent spaces. We have a basis of $T_x(M)$ consisting of the derivations D_i, which are the images of the derivations $\partial/\partial x_i \colon \mathcal{O}_{\mathbf{K}^n,y} \to \mathbf{K}$. Hence, for $[(W,f)]$ in $\mathcal{O}_{M,x}$ we get that $D_i(f) = \partial f\Phi/\partial x_i(y)$. Note that the basis $D_1, D_2, \dots, D_n$ depends on the chart (V, Φ). On the other hand, when we have chosen one chart, it will give a natural basis for the tangent space in all points of this chart. We often write $D_i = \partial/\partial x_i$, as mentioned in Example 3.6.5, when we have specified a chart.

3.6.8. Let $\Phi\colon N \to M$ be an analytic map of manifolds. For each x in N we have a ring homomorphism

$$\Phi_x^* \colon \mathcal{O}_{M,\Phi(x)} \to \mathcal{O}_{N,x},$$

(see Example 3.5.4). Hence we obtain a map

$$T_x\Phi \colon T_x(N) \to T_{\Phi(x)}(M),$$

that sends a derivation D of $\mathcal{O}_{N,x}$, with respect to the augmentation map on $\mathcal{O}_{N,x}$, to the derivation $D\Phi_x^*$ of $\mathcal{O}_{M,\Phi(x)}$, with respect to the augmentation on $\mathcal{O}_{M,\Phi(x)}$. Clearly, the map $T_x\Phi$ is a $\mathbf{K}$-linear map. Moreover, if $\Psi\colon P \to N$ is an analytic map and x is a point of P, we have that $T_{\Psi(x)}\Phi T_x\Psi = T_x(\Phi\Psi)$.

Definition 3.6.9. A *curve* in a manifold M is an analytic map $\gamma\colon B(a,r) \to M$, for a ball in $\mathbf{K}$. The *tangent* $\gamma'(a)$ of the curve in $\gamma(a)$ is the image $T_a\gamma(d/dt)$ of the standard basis d/dt of $T_a(\mathbf{K}) = V_{\mathbf{K}}^1$ by the map $T_a\gamma\colon T_a(\mathbf{K}) \to T_{\gamma(a)}(M)$.

Remark 3.6.10. It follows from Paragraph 3.6.8 that, given a chart $\Phi\colon U \to M$ such that $\gamma(a) \in \Phi(U)$, and $\Phi^{-1}\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$ for t in a neighborhood of a, then

$$T_{\Phi^{-1}(\gamma(a))}\Phi(\gamma_1'(a), \dots, \gamma_n'(a)) = \gamma'(a).$$

Consequently, the definition of the tangent to a curve of a manifold corresponds, via a chart, to the tangent to the corresponding curve of $\mathbf{K}^n$, as given in Paragraph 3.6.1 and Definition 2.5.1.

Definition 3.6.11. Let M and N be manifolds, where N is a subset of M. We say that N is a *submanifold* of M if the inclusion map of N in M is analytic and if the resulting map $T_xN \to T_xM$ of Paragraph 3.6.8 is injective, for all x in N.

Example 3.6.12. It follows from Example 3.4.13 and Proposition 2.5.4 that the groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$, and thus $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, $\mathrm{Sp}_n(\mathbf{K})$ are submanifolds of $\mathrm{M}_n(\mathbf{K})$.

Example 3.6.13. Let M and N be manifolds. Then $M \times N$ with the product topology is a manifold (see Example 3.4.7). For each point y in N we have a closed subset $M \times \{y\}$ of $M \times N$, and we have an isomorphism $\Phi_y\colon M \to M \times \{y\}$ of metric spaces that maps a point x of M to (x,y). This map defines a structure of manifold on $M \times \{y\}$, and it is clear that, with this structure, we have that the inclusion of $M \times \{y\}$ in $M \times N$ is analytic.

The inclusion Φ_y induces a map $T_x\Phi_y\colon T_xM \to T_{(x,y)}(M \times N)$. Moreover, the composite map of Φ_y with the projection $\pi_1\colon M \times N \to M$ onto the second factor is the identity on M. The map $T_{(x,y)}\pi_1$ is therefore the inverse map to $T_x\Phi_y$. Let $\Psi_x\colon N \to M \times N$ be the map defined by $\Psi_x(y) = (x, y)$ for all y in N. We obtain a map:

$$T_x\Phi_y \times T_y\Psi_x\colon T_xM \oplus T_yN \to T_{(x,y)}(M \times N),$$

from the direct sum of the spaces T_xM and T_yN (see Example 1.6.4) and a reverse map:

$$T_{(x,y)}\pi_1 \times T_{(x,y)}\pi_2\colon T_{(x,y)}(M \times N) \to T_xM \oplus T_yN,$$

which sends (D, D') to $T_{(x,y)}\pi_1(D)+T_{(x,y)}\pi_2(D')$. It is clear that the two maps are inverses to each other. Consequently, there is a canonical isomorphism

$$T_{(x,y)}(M \times N) \xrightarrow{\sim} T_xM \oplus T_yN$$

of vector spaces.

Example 3.6.14. Let N be the subset $\{(a^2, a^3)\colon a \in \mathbf{K}\}$ of $\mathbf{K}^2$, and let N have the topology induced by the metric topology on $\mathbf{K}^2$. The map $f\colon \mathbf{K} \to N$ defined by $f(a) = (a^2, a^3)$ defines a *chart*, and *atlas* on N. Hence N is a manifold. The inclusion map i is clearly analytic. However, N is not a submanifold of $\mathbf{K}^2$, because the map on tangent spaces $T_afi\colon T_a(\mathbf{K}) \to T_{(a^2,a^3)}(\mathbf{K}^2)$, maps the basis vector D to the vector $(2aD_1, 3a^2D_2)$, where $\{D\}$, $\{D_1, D_2\}$ are the bases on the tangent spaces corresponding to the standard bases on $\mathbf{K}$ and $\mathbf{K}^2$. However, this map is zero at $a = 0$.

Lemma 3.6.15. *Let M and N be manifolds of dimensions m and n. Suppose that $N \subseteq M$ and that the inclusion map is analytic. Then N is a submanifold of M if and only if around each point x of N, there is a chart $\Phi\colon U \to M$ such that $\Phi^{-1}(N)$ is the intersection of U by a linear subspace $W \in \mathbf{K}^m$ of dimension n and $\Phi|_W\colon W \cap U \to N$ is a chart of N.*

Proof: It is clear that, if the condition of the lemma holds, then the map of tangent spaces is injective, indeed, up to isomorphisms, the tangent map is equal to the inclusion map of W into $\mathbf{K}^m$.

Conversely, assume that N is a submanifold of M. Fix x in N and choose charts $\psi\colon V \to N$ and $\varphi\colon U \to M$, such that $V \subseteq U$. It follows from Paragraph 3.6.7 that we have isomorphisms $T_{\psi^{-1}(x)}V \to T_xN$ and $T_{\phi^{-1}(x)}U \to T_xM$. Consequently the map $\varphi^{-1}\psi\colon V \to U$ gives an injective map $T_{\psi^{-1}(x)}V \to T_{\varphi(x)}U$. The latter map is the same as $(\varphi^{-1}\psi)'(x)$. It follows from Theorem 3.1.2 that the condition of the lemma holds. □

Proposition 3.6.16. *Let N be a submanifold of M, then N is locally closed in M, that is, for each x in N there is a neighborhood U of x in M such that $N \cap U$ is closed in M.*

Proof: Since a linear subspace W of $\mathbf{K}^m$ is closed in $\mathbf{K}^m$, it follows from Lemma 3.6.15 that N is locally closed in M. □

Proposition 3.6.17. *Let N be a submanifold of M, then the map $\mathcal{O}_{M,x} \to \mathcal{O}_{N,x}$ is surjective, for all points x in N.*

Proof: Let x be a point of N. Since N is a submanifold of M, it follows from Lemma 3.6.15 that we can find a chart $\Phi\colon U \to M$ around x such that $\Phi^{-1}(N) \cap U$ is the intersection of a linear space W with U and such that $\Phi|_W$ is a chart for N around x. Thus it suffices to show that any analytic function f defined on $W \cap U$ can be extended to an analytic function on all of U. This can be done by composing f with some linear projection of $\mathbf{K}^m$ onto W that maps U to $W \cap U$. It suffices to find a projection mapping V to $V \cap W$ for some neighborhood V of x contained in U. Hence we may assume that $\Phi^{-1}(x) = 0$ in $\mathbf{K}^m$, and that U is a ball with center 0. Then it is clear how to find the linear projections. Since all linear maps are analytic, this composition will be analytic on U. □

3.7 The tangent spaces of zeroes of analytic functions

We shall, in this section, give an easy method to compute the tangent spaces of subsets of $\mathbf{K}^n$ defined as the zeroes of analytic functions, and use the method to compute the tangent spaces of the matrix groups.

Let Z be the set contained in an open subset U of $\mathbf{K}^n$ which is the common zeroes of analytic functions $\{f_i\}_{i \in I}$ on U. When Z is a submanifold of U we know that in each point x of Z, there is an injective map $T_xZ \to T_xU$. We want to describe the linear space T_xZ as a subspace of T_xU.

Let $D\colon \mathcal{O}_{\mathbf{K}^n,x} \to \mathbf{K}$ be an element of T_xU which is also an element of T_xZ. For any element f in $\mathcal{I}(Z)$ we have that f is mapped to zero by the map $\mathcal{O}_{U,z} \to \mathcal{O}_{Z,x}$ and we must consequently have that $Df = 0$. Thus we get

$$T_xZ \subseteq \{D \in T_xU \mid Df = 0, \quad \text{for all } f \in \mathcal{I}(Z)\}. \tag{3.7.0.1}$$

We know from Example 3.6.5 that T_xU is the set of derivations $\sum_{i=1}^n a_i \partial/\partial x_i$, where $a_1, a_2, \dots, a_n \in \mathbf{K}$. Thus the set $\{D \in T_xU \mid Df = 0, \quad \text{for all } f \in \mathcal{I}(Z)\}$ can be written as

$$\left\{ \sum_{i=1}^n a_i \frac{\partial}{\partial x_i} \;\middle|\; \sum_{i=1}^n a_i \frac{\partial f}{\partial x_i} = 0, \quad \text{for all } f \in \mathcal{I}(Z) \right\},$$

which we can also describe as $N_x(Z)^\perp$. On the other hand, we know from Theorem 3.2.3 that the dimension of Z is $n - \dim_{\mathbf{K}} N_x(Z)$. Thus it follows from Lemma 1.7.4 that $\dim_{\mathbf{K}} T_x(Z) = \dim_{\mathbf{K}} N_x(Z)^\perp$, which proves that

$$T_xZ = \{D \in T_xU \mid Df = 0 \text{ for all } f \in \mathcal{I}(Z)\}. \tag{3.7.0.2}$$

We note that if V is a neighborhood of x contained in U then $N_x(Z) \subseteq N_x(Z \cap U)$, and it follows from the equality $T_x(Z) = N_x(Z)^\perp = N_x(Z \cap U)^\perp$ that $N_x(Z) = N_x(Z \cap U)$.

3.7.1 The epsilon calculus

The observation that the tangent space of an analytic set N, defined as the zeroes of analytic functions, depends on the the linear terms of the analytic functions only, can be conveniently expressed by the, so called, *epsilon calculus*. This calculus disregards, in a natural way, all terms of degree higher than 1. To explain the calculus we notice that the ring of dual numbers $\mathbf{K}[\varepsilon]$ of $\mathbf{K}$ (see Example 1.3.15) is isomorphic, as a vector space, to $\mathbf{K}^2$ and thus has a norm, which makes it possible for us to talk about analytic functions

$f\colon U \to \mathbf{K}[\varepsilon]$ defined on open subsets U of $\mathbf{K}[\varepsilon]^n$. Let $f\colon U \to \mathbf{K}$ be an analytic function defined in a neighborhood U of a point $x \in \mathbf{K}^n$. Then we can extend f to an analytic function $\bar{f}\colon V \to \mathbf{K}[\varepsilon]$, where V is open in $\mathbf{K}[\varepsilon]^n$, by using the same power series. In fact, assume that f is given by $f(x+h) = \sum_{i\in\mathcal{I}} c_i h^i$, for small h. Then we define $\bar{f}$ by $\bar{f}(x+h_1+h_2\varepsilon) = \sum_{i\in\mathcal{I}} c_i(h_1+h_2\varepsilon)^i$. Since we have that $(h_1+h_2\varepsilon)^i = h_1^i + \sum_{|j|=1} h_1^{i-j}h_2^j\varepsilon$ is a sum with $n+1$ terms for each i, we can change the order of summation to get

$$\begin{aligned}\bar{f}(x+h_1+h_2\varepsilon) &= \sum_{i\in\mathcal{I}} c_i h_1^i + \varepsilon\sum_{|j|=1}\sum_{i\in\mathcal{I}}\binom{i}{j} h_1^{i-j}h_2^j \\ &= f(x+h_1) + \varepsilon\sum_{|j|=1} D^j f(x+h_1)h_2^j.\end{aligned}$$

Since $\bar{f}(x+h_1) = f(x+h_1)$ Equality 3.7.0.1 can now be expressed as

$$T_x(Z) = \{v \in \mathbf{K}^n \mid \bar{f}(x+\varepsilon v) - \bar{f}(x) = 0, \quad \text{for all } f \in \mathcal{I}(Z)\}.$$

3.7.2 Computation of the tangent spaces

The disadvantage of the description 3.7.0.1 of the tangent spaces is that it requires knowledge of the ideal $\mathcal{I}(Z)$. In most cases we only know a subset of $\mathcal{I}(Z)$ such that the zeroes of the functions in the subset is Z. Often we can find analytic functions $f_i : U \to \mathbf{K}$ for $i = 1, \dots, m$ such that Z is the common zeroes of $f_1, \dots, f_m$ and such that the subspace $N_x(f_1, \dots, f_m)$ of T_xU is generated by the vectors $(\frac{\partial f_i}{\partial x_1}, \dots \frac{\partial f_i}{\partial x_n})$ for $i = 1, \dots, m$ has dimension m, or equivalently that the subspace $T_x(f_1, \dots, f_m) = N_x(f_1, \dots, f_m)^\perp$ of T_xU has dimension $n-m$. We then have that

$$T_xZ = N_x(f_1, \dots, f_m) = \{D \in T_xU \mid Df_i = 0 \text{ for } i = 1, \dots, m\}. \tag{3.7.0.3}$$

or, with the epsilon calculus as

$$T_xZ = \{v \in \mathbf{K}^n \mid \bar{f}_i(x+\varepsilon v) - \bar{f}_i(x) = 0 \text{ for } i = 1, \dots, m\}. \tag{3.7.0.4}$$

Since $N_x(Z)$ contains $N_x(f_1, \dots, f_m)$ and consequently $N_x(f_1, \dots, f_m)^\perp$ is contained in T_xZ, the Equality 3.7.0.3 is equivalent to the equality $\dim_{\mathbf{K}} T_xZ = n-m$. To prove the latter equality we consider the analytic map $\Phi : U \to \mathbf{K}^m$ defined by $\Phi(y) = (f_1, (y), \dots, f_m(y))$. By assumption we have that the rank of $\Phi'(x)$ is m. Hence it follows from Theorem 3.1.3 that there is an analytic bijection $\Psi : V \to \Psi(V)$ with an analytic inverse from a neighborhood V of x contained in U to an open subset $\Psi(V)$ of $\mathbf{K}^m$, that maps $Z \cap U$ bijectively onto $W \cap \Psi(V)$, where W is a linear subspace of $\mathbf{K}^m$ of dimension $n-m$. We consequently have an isomorphism $\Psi'(x) : T_xV \to T_{\Psi(x)}(\Psi(V))$ that induces an isomorphism between $T_x(Z \cap V)$ and $T_{\Psi(x)}(W \cap \Psi(V))$. Since W is a linear subspace of dimension $n-m$ we clearly have that $\dim_{\mathbf{K}} T_{\Psi(x)}(W \cap \Psi(V)) = n-m$, and we have proved the desired equality $\dim_{\mathbf{K}} T_xZ = \dim_{\mathbf{K}} T_x(Z \cap V) = n-m$.

We are now ready to compute the tangent spaces of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$ and $\mathrm{Sp}_n(\mathbf{K})$.

Example 3.7.1. The group $\mathrm{Sl}_n(\mathbf{K})$ is the subset of $\mathrm{M}_n(\mathbf{K}) \cong \mathbf{K}^{n^2}$, defined by the polynomial equation $f(x_{ij}) = \det(x_{ij}) - 1 = 0$. Consider the space of derivations D in $T_{I_n}\,\mathrm{M}_n(\mathbf{K})$ such that $Df = 0$. By the epsilon calculus this space equals

$$\{A \in \mathrm{M}_n(\mathbf{K}) \bigm| \det(I_n + \varepsilon A) - \det I_n = 0\}.$$

A short calculation shows that $\det(I_n + \varepsilon A) = 1 + \varepsilon \operatorname{tr} A$, where $\operatorname{tr} A$ is the *trace* of A, that is, the sum of the diagonal elements of A (see Exercise 3.7.2). Since the trace is a non-zero linear equation in the entries of A, the subspace of matrices of trace zero has dimension $n^2 - 1$. Thus we have that

$$T_{I_n}(\mathrm{Sl}_n(\mathbf{K})) = \{(a_{i,j}) \in \mathrm{M}_n(\mathbf{K}) \bigm| \operatorname{tr} A = 0\}.$$

That is, $T_{I_n}(\mathrm{Sl}_n(\mathbf{K}))$ consists of all matrices whose trace is equal to zero. In particular we have that the tangent space, and hence $\mathrm{Sl}_m(\mathbf{K})$ both have dimension $n^2 - 1$ (see Exercise 2.5.4).

Example 3.7.2. The group $\mathrm{O}_n(\mathbf{K})$ is the subset of $\mathrm{M}_n(\mathbf{K}) \cong \mathbf{K}^{n^2}$ defined by the n^2 polynomials, in n^2 variables, that are the coefficients in the matrix ${}^tXX - I_n$ However, these polynomials are not independent, since ${}^tXX - I_n$ is a symmetric matrix. Thus there are only $n(n+1)/2$ different entries, $f_{ij}(X)$, for $1 \leq i \leq j \leq n$. The space of derivations D in $T_{I_n}\,\mathrm{O}_n(\mathbf{K})$ such that $Df_{ij} = 0$, for all $1 \leq i \leq j \leq n$ can by epsilon calculus be written as

$$\{A \in \mathrm{M}_n(\mathbf{K}) \bigm| {}^t(I_n + A\varepsilon)(I_n + A\varepsilon) - I_n = 0\}.$$

We have that ${}^t(I_n + A\varepsilon)(I_n + A\varepsilon) - I_n = ({}^tI_n + {}^tA\varepsilon)(I_n + A\varepsilon) - I_n = I_n + {}^tA\varepsilon + A\varepsilon - I_n = ({}^tA + A)\varepsilon$. Consequently, the space we consider is

$$T_{I_n}(\mathrm{O}_n(\mathbf{K})) = \{A \in \mathrm{M}_n(\mathbf{K}) \bigm| {}^tA + A = 0\}.$$

That is, the set of all *skew-symmetric* matrices. This space has dimension $n(n-1)/2$ (see Exercise 2.5.5). In particular, we have that $n(n-1)/2 + n(n+1)/2 = n^2$, and $T_{I_n}\,\mathrm{O}_n(\mathbf{K})$ is equal to the set of skew-symmetric matrices.

The subspace $\mathrm{SO}_n(\mathbf{K})$ is defined in $\mathrm{M}_n(\mathbf{K})$ by the same equations as $\mathrm{O}_n(\mathbf{K})$ plus the equation $\det(x_{i,j}) - 1 = 0$. Since at any point of $\mathrm{O}_n(\mathbf{K})$, the determinant is either 1 or -1 all points in a neighborhood of x in $\mathrm{O}_n(\mathbf{K})$ are in $\mathrm{SO}_n(\mathbf{K})$, and the new equation will not contribute to $N_x(\mathrm{SO}_n(\mathbf{K}))$. Thus the tangent space of $\mathrm{SO}_n(\mathbf{K})$ is equal to the tangent space of $\mathrm{O}_n(\mathbf{K})$ at any point of $\mathrm{SO}_n(\mathbf{K})$.

Example 3.7.3. The symplectic group $\mathrm{Sp}_n(\mathbf{K})$ is the subset of $\mathrm{M}_n(\mathbf{K})$ of common zeroes of the n^2 polynomials in n^2 variables that are the coefficients in the matrix $XS{}^tX - S$. These polynomials are not independent, since $XS{}^tX - S$ is skew-symmetric and we have, in fact, only $n(n-1)/2$ different equations $f_{ij}(X)$, for $1 \leq i < j \leq n$. We consider the space of derivations D in $T_{I_n}\mathbf{K}^{n^2}$ such that $Df_{ij} = 0$, for all $1 \leq i < j \leq n$, and obtain by the epsilon calculus that the space can be written as

$$\{A \in \mathrm{M}_n(\mathbf{K}) \bigm| {}^t(I_n + A\varepsilon)S(I_n + A\varepsilon) = S\}.$$

We have that ${}^t(I_n + A\varepsilon)S(I_n + A\varepsilon) - S = S + {}^tAS\varepsilon + SA\varepsilon - S$. Consequently, we have that this space is equal to

$$\{A \in \mathrm{M}_n(\mathbf{K})) \mid {}^tAS + SA = 0\}.$$

However ${}^tAS + SA = SA - {}^tA{}^tS = SA - {}^t(SA)$. Consequently, the isomorphism of vector spaces $\mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$, which maps a matrix A to SA (see Exercise 2.5.6), maps this space isomorphically onto the subspace of $\mathrm{M}_n(\mathbf{K})$ consisting of symmetric matrices. In particular, this space has dimension $n(n+1)/2 = n^2 - n(n-1)/2$, which shows that the tangent space $T_{I_n} \mathrm{Sp}_n(\mathbf{C})$ has dimension $n(n+1)/2$ (see Exercise 2.5.7).

We shall now indicate another way of computing the tangent spaces in the previous examples that is closer to the methods we shall use in the algebraic case in Chapter **??**. The treatment will not be self-contained, since we need results from the theory of several complex variables which it would take too much space to explain and prove here. We refer to Griffiths–Harris [4] for these results.

First we need some concepts from algebra, which also will be utterly important in the study of algebraic varieties in Chapter **??**.

Definition 3.7.4. A ring R where no non-zero element is a zero-divisor is called an *integral domain* or sometimes just *domain*. In an integral domain, we say that an element f is *irreducible* if in any factorization $f = gh$, either g or h is invertible. An integral domain R is a *unique factorization domain* if every non-zero element f can be uniquely – up to invertible elements – written as a product of irreducible elements.

Example 3.7.5. The integers $\mathbf{Z}$ is the standard model of a domain. The irreducible elements are ± 1 and $\pm p$, where p is prime number. It is also a unique factorization domain, since we have unique prime factorization of all positive integers. An example of a domain which does not have unique factorization has to be somewhat more complicated; $R = \{a + bi\sqrt{5} \mid a, b \in \mathbf{Z}\}$ is an example, since 6 have two different factorizations, $2 \cdot 3$ and $(1 + i\sqrt{5})(1 - i\sqrt{5})$, into irreducible elements. (It is a domain, since it is a subring of the complex numbers.)

We will now quote two results from the theory of several complex variables without proof.

Theorem 3.7.6. *The local ring $\mathcal{O}_{\mathbf{C}^n,x}$ of analytic functions defined in neighborhood of a point x is a unique factorization domain.*

Theorem 3.7.7. (Weak Nullstellensatz) *If $h \in \mathcal{O}_{\mathbf{C}^n,x}$ vanishes on the zeroes of an irreducible function $f \in \mathcal{O}_{\mathbf{C}^n,x}$, then $h = fg$ for some $g \in \mathcal{O}_{\mathbf{C}^n,x}$.*

Remark 3.7.8. Observe that this is not true for real analytic functions, since for example $x^2 + y^2$ is an irreducible analytic function defined around the origin in $\mathbf{R}^2$, while neither x nor y is divisible by $x^2 + y^2$, though they both vanish at the origin, which is the zero set of $x^2 + y^2$.

From these two theorems, we can prove the following fundamental theorem for the zeroes of analytic functions which basically states that the set of zeroes of an analytic function on $\mathbf{C}^n$ cannot have dimension less than $n-1$. This is not true for real analytic functions, as the remark above shows.

Theorem 3.7.9. (Dimension Theorem) *Let $f : \mathbf{C}^n \to \mathbf{C}$ be an analytic function and let Z be the set of zeroes of f. Then we have that*

$$\dim_{\mathbf{C}} N_x Z \leq 1, \qquad \text{for all points } x \text{ in } Z.$$

Proof: By Theorem 3.7.6 we can write f as a product of irreducible functions $f = f_1 f_2 \cdots f_m$. Since the zero set of a power of an analytic function equals the zero set of the function itself, we may assume that the functions $f_1, f_2, \dots, f_m$ are distinct and do not divide each other. Let h be any function in $I(Z)$. Since h vanishes on the zero set of $f_1 f_2 \cdots f_m$, it must vanish on the zero set of each f_i. Thus Theorem 3.7.7 says that f_i divides h for $i = 1, 2, \dots, m$, but since the f_i's does not divide each other, we conclude that $h = f_1 f_2 \cdots f_n g = fg$, for some analytic function g. Now if $D \in T_x\mathbf{C}^n$ is a derivation, where $x \in Z$, we have that

$$Dh = f(x)Dg + g(x)Df = g(x)Df.$$

Thus the vector $(\partial f/\partial x_1, \partial f/\partial x_2, \dots, \partial f/\partial x_n)$ spans N_xZ, whose dimension thereby is at most 1. □

Corollary 3.7.10. *Let $f_1, f_2, \dots, f_m$ be analytic functions on $\mathbf{C}^n$ and let Z be the subset of $\mathbf{C}^n$ where they vanish. Then we have that $\dim_{\mathbf{C}} N_x Z \leq m$, for any point x in Z.*

Proof: Let Z_1 be the zero set of f_1. By the theorem we have that $\dim_{\mathbf{C}} N_x Z_1 \leq 1$, for $x \in Z$. By Theorem 3.2.3, we can parameterize the Z_1 around any point $x \in Z$ where $\dim_{\mathbf{C}} N_x Z$ is maximal, by an open subset of $\mathbf{C}^n$ or $\mathbf{C}^{n-1}$. The analytic functions $f_2, \dots, f_m$ define analytic functions on this set and the corollary follows by induction. □

If the maximum in the corollary is attained, we say that Z is a *complete intersection*. In particular, if we have that

$$\{D \in T_x\mathbf{C}^n \mid Df_i = 0, \quad \text{for } i = 1, 2, \dots, m\} \tag{3.7.10.1}$$

has dimension $n-m$, we get that $\dim_{\mathbf{C}} T_x Z \leq n-m$, while by the corollary, $\dim_{\mathbf{C}} N_x Z \leq m$. These two inequalities together imply that we have equality, and Z is a complete intersection. Thus 3.7.10.1 gives an expression for the tangent space $T_x Z$.

We shall now see that the ordinary complex matrix groups are complete intersections in the affine space of matrices.

Exercises

3.7.1. Prove that for any integer d, the set $\mathbf{Z}[\sqrt{d}] = \{a + b\sqrt{d} \mid a, b \in \mathbf{Z}\}$ is an integral domain.

3.7.2. Show that $\det(I_n + \varepsilon A) = 1 + \sum_{i=1}^{n} A_{ii}\varepsilon$.

3.8 Connectedness

As we observed in Sections 1.10 and 3.6 we can not yet distinguish $\mathrm{O}_n(\mathbf{K})$ from $\mathrm{SO}_n(\mathbf{K})$. There is however, an important topological invariant, connectedness, that distinguishes $\mathrm{O}_n(\mathbf{K})$ from the other matrix groups.

Definition 3.8.1. Let X be a topological space. An *arch* in X is a continuous map $\gamma\colon [0,1] \to X$ from the closed unit interval, with the metric topology, to X. We call $\gamma(0)$ and $\gamma(1)$ the *beginning*, respectively *end*, of the arch.

Remark 3.8.2. If we have two arches, given by, $\gamma\colon [0,1] \to X$ and $\delta[0,1] \to X$ such that $\gamma(1) = \delta(0)$, then the map $\varepsilon\colon [0,1] \to X$ defined by $\varepsilon(a) = \gamma(2a)$, when $a \in [0, \frac{1}{2}]$, and $\varepsilon(a) = \delta(2a-1)$, when $a \in [\frac{1}{2}, 1]$, gives an arch which begins in $\gamma(0)$ and ends in $\delta(1)$. Thus the property that x and y can be connected by an arch yields an equivalence relation on X.

Definition 3.8.3. A topological space X is *archwise connected* if, for every pair of points x, y of X, there is an arch which begins in x and ends in y.

The space X is *connected* if it can not be written as the union of two disjoint non-empty open sets. That is, there does not exist open sets U and V of X such that $X = U \cup V$ and $U \cap V = \emptyset$.

A subset Y of X which is connected in the induced topology, and not contained in any other connected subset is called a *connected component*.

Remark 3.8.4. The assertion that X is connected can be expressed in many different ways, like X is not the union of two disjoint non-empty closed sets, the complement of an non-empty open set can not be open, or, the complement of a non-empty closed set can not be closed. A connected component Y of X is closed because when Y is connected then the *closure* $\bar{Y}$ of Y, that is, the smallest closed set in X containing Y, is connected (see Exercise 3.8.6).

Example 3.8.5. The unit interval $[0,1]$ is connected (see Exercise 3.8.1).

Example 3.8.6. The space $\mathbf{K}^n$ is archwise connected in the metric topology. Indeed, any two points can be joined by a straight line.

Example 3.8.7. When the field $\mathbf{K}$ is infinite, the space $\mathbf{K}^n$ is connected in the Zariski topology (see Exercise 3.3.7). On the other hand, when $\mathbf{K}$ is finite, all subsets are open, and $\mathbf{K}^n$ is not connected.

Lemma 3.8.8. *Let X be a topological space. Then X can be written uniquely as a union $X = \{X_i\}_{i \in I}$, where the X_i are the connected components. We have that $X_i \cap X_j = \emptyset$, when $i \neq j$.*

Proof: Let $\{Y_j\}_{j\in J}$ be an ordered set of connected subsets in X, that is $Y_j \subseteq Y_{j'}$ or $Y_{j'} \subseteq Y_j$ for all j, j' in J. Then we have that $\bigcup_{j\in J} Y_j$ is connected because, if $\bigcup_{j\in J} Y_j$ is the union of two disjoint open sets, the same must be true for at least one of the Y_j. Given a point x in X, and let $\{Y_j\}_{j\in J}$ be the family of all connected subsets of X that contain x. Since the union of an ordered family of connected sets is connected there must be a maximal set in the family $\{Y_j\}_{j\in J}$. Consequently every point is contained in a connected component of X. Hence we have that X is the union of connected components. Two components can not intersect, because then the union would be connected. Similarly, we see that a composition into connected components is unique. □

Lemma 3.8.9. *An archwise connected topological space is connected.*

Proof: Assume that X is archwise connected. If $X = U \cup V$, where U and V are open, non-empty, disjoint sets such that $X = U \cup V$ we choose points x and y in U respectively V. There is an arch given by $\gamma\colon [0,1] \to X$, beginning in x and ending in y. Then $[0,1]$ is the union of the two non-empty open sets $\gamma^{-1}(U)$ and $\gamma^{-1}(V)$, and $\gamma^{-1}(U) \cap \gamma^{-1}(V) = \gamma^{-1}(U \cap V) = \emptyset$. However, this is impossible, since $[0,1]$ is connected (see Example 3.8.5). Hence X is connected. □

Lemma 3.8.10. *A connected manifold is archwise connected.*

Proof: Let M be a connected manifold. For each point x of M, denote by U_x the set of points that are the ends of arches in M that begin at x. We have that U_x is open, because, if y is in U_x, then there is a chart $f\colon B(0,r) \to M$, from a ball in $\mathbf{K}^m$, such that $f(0) = y$. Clearly the ball is archwise connected. Consequently we have that $f(B(0,r))$ is archwise connected, and hence is contained in U_x. Hence U_x is open. Fix x in M. If U_x is not all of M, then, for every point y in M outside of U_x, the set U_y is disjoint from U_x by Remark 3.8.2. Consequently the complement of U_x is open, which contradicts the connectivity of M. We thus have that $M = U_x$, and hence is archwise connected. □

Lemma 3.8.11. *Let $f\colon X \to Y$ be a continuous map of topological spaces. If X is connected, then $f(X)$ is connected.*

Proof: Assume that Y can be written $Y = U \cup V$ where U and V are non-empty open sets such that $f(X) \cap f^{-1}(U)$ and $f(X) \cap f^{-1}(V)$ are disjoint. Then $X = f^{-1}(U) \cup f^{-1}(V)$ expresses X as a union of disjoint open sets. Since X is connected we must have that $f(X) \subseteq U$ or $f(X) \subseteq V$. Consequently $f(X)$ is connected, and we have proved the lemma. □

Proposition 3.8.12. *The groups* $\mathrm{Gl}_n(\mathbf{C})$, $\mathrm{Sl}_n(\mathbf{C})$ *and* $\mathrm{Sl}_n(\mathbf{R})$ *are connected in the metric topologies, whereas* $\mathrm{Gl}_n(\mathbf{R})$ *consists of two connected components.*

Proof: It follows from Proposition 1.5.2 that every element A of $\mathrm{Gl}_n(\mathbf{K})$ can be written as a product of matrices $E_{ij}(a)$ and $E(b)$ with a and b in $\mathbf{K}$, and where $E(b)$ is the matrix 1.5.2.1. Moreover if A is in $\mathrm{Sl}_n(\mathbf{K})$ then $b = 1$.

Let $\mathbf{K} = \mathbf{R}$. When we replace the matrices $E_{ij}(a)$ in the product with $E_{ij}(at)$ we obtain an arch $\gamma : [0,1] \to \mathrm{Gl}_n(\mathbf{K})$ from $E(c)$ to A for some c in $\mathbf{K}$. When A is in $\mathrm{Sl}_n(\mathbf{K})$ we have that $c = 1$. Hence $E(c) = I_n$, and $\mathrm{Sl}_n(\mathbf{K})$ is connected.

If $\mathbf{K} = \mathbf{C}$, we can find an arch $\gamma : [0,1] \to \mathrm{Gl}_n(\mathbf{C})$ from $E(a)$ to I_n, since $\mathbf{C} \setminus \{0\}$ is connected. Thus $\mathrm{Gl}_n(\mathbf{C})$ is connected. For $\mathrm{Gl}_n(\mathbf{R})$, we can connect $E(a)$ by an arch to $E(-1)$ or $I_n = E(1)$, depending on the sign of $\det A$. On the other hand $\det^{-1}(1)$ and $\det^{-1}(-1)$ are disjoint open sets of $\mathrm{Gl}_n(\mathbf{R})$ whose union is $\mathrm{Gl}_n(\mathbf{R})$. Thus $\mathrm{Gl}_n(\mathbf{R})$ consists of two connected components. □

Proposition 3.8.13. *The group* $\mathrm{SO}_n(\mathbf{K})$ *is connected in the metric topology, and* $\mathrm{O}_n(\mathbf{K})$ *consists of two connected components.*

Proof: We have that $\det^{-1}(1) \cup \det^{-1}(-1)$ gives a partition of $\mathrm{O}_n(\mathbf{K})$ into two disjoint open sets $\mathrm{SO}_n(\mathbf{K})$ and $E(b)\,\mathrm{SO}_n(\mathbf{K})$ where $E(b)$ is the matrix of 1.5.2.1. Hence, when we have proved that $\mathrm{SO}_n(\mathbf{K})$ is connected, it will follow that $\mathrm{O}_n(\mathbf{K})$ consists of two connected components.

It follows from Proposition 1.9.4 that $\mathrm{SO}_n(\mathbf{K})$ is generated by products of two reflections of the form s_x, where $\langle x, x\rangle \neq 0$. Let $A = \prod s_{x_i} s_{y_i}$ be an element of $\mathrm{SO}_n(\mathbf{K})$. If we can show that the set $\{x \in V^n(\mathbf{K}) \mid \langle x, x\rangle \neq 0\}$ is connected, we can find arches $\gamma_i\colon [0,1] \to V^n(\mathbf{K})$ from x_i to y_i, for all i. Thus we can define an arch $\gamma\colon [0,1] \to \mathrm{SO}_n(\mathbf{K})$ by $\gamma(t) = \prod s_{\gamma_i(t)} s_{y_i}$, which goes from A to I_n and $\mathrm{SO}_n(\mathbf{K})$ is connected.

It remains to prove that $X = \{x \in V^n(\mathbf{K}) \mid \langle x, x\rangle \neq 0\}$ is connected. For $\mathbf{K} = \mathbf{R}$, we have that $X = V^n(\mathbf{R}) \setminus \{0\}$, which is connected for $n > 1$. The case $n = 1$ is trivial, since $\mathrm{SO}_1(\mathbf{K}) = \{1\}$. For $\mathbf{K} = \mathbf{C}$, we can take a complex line $tx + (1-t)y$ through any two points $x, y \in X$. On this line, there are at most the two points corresponding to the zeroes of the equation $\langle tx + (1-t)y, tx + (1-t)y\rangle = t\langle x, x\rangle + 2t(1-t)\langle x, y\rangle + (1-t)^2\langle y, y\rangle = 0$ in t that are not in X. However the complex line minus two points is still connected. Hence we can find an arch between x and y in X. □

Proposition 3.8.14. *The group* $\mathrm{Sp}_n(\mathbf{K})$ *is connected.*

Proof: It follows from Proposition 1.9.9 that every element of $\mathrm{Sp}_n(\mathbf{K})$ can be written as a product of transvections $\tau(x, a)$, where $\tau(x, a)(y) = y - a\langle y, x\rangle x$. From Remark 3.8.2 it follows that it suffices to find an arch $\gamma\colon [0,1] \to \mathrm{Sp}_n(\mathbf{K})$ such that $\gamma(0) = I_n$ and $\gamma(1) = \tau(x, a)$. However, we can define such an arch by $\gamma(t) = \tau(x, ta)$, for $t \in [0,1]$. □

Example 3.8.15. We collect the information we have about the matrix groups in the Table 2.

As we observed in Section 1.10 the size of the center alone suffices to distinguish $\mathrm{Gl}_n(\mathbf{C})$ from the remaining groups. Moreover, for $n > 2$, the same is true for $\mathrm{Sl}_n(\mathbf{C})$, and when n is odd, for $\mathrm{SO}_n(\mathbf{C})$. Hence none of these sets are isomorphic as groups when n is odd and $n > 2$. The group $\mathrm{O}_n(\mathbf{C})$ is the only one that is not connected and can not be homeomorphic, as topological space, to any of the other groups. Finally, for m even, $\mathrm{SO}_m(\mathbf{C})$ can not be equal to $\mathrm{Sp}_n(\mathbf{C})$ as a manifold for n even, since then they must have

Group	n	Center	Dimension	Connected
$\mathrm{Gl}_n(\mathbf{C})$	arb.	$\mathbf{K}^*$	n^2	yes
$\mathrm{Sl}_n(\mathbf{C})$	arb.	$\mathbf{Z}/n\mathbf{Z}$	n^2-1	yes
$\mathrm{O}_n(\mathbf{C})$	arb.	$\{\pm 1\}$	$\frac{n(n-1)}{2}$	no
$\mathrm{SO}_n(\mathbf{C})$	even	$\{\pm 1\}$	$\frac{n(n-1)}{2}$	yes
$\mathrm{SO}_n(\mathbf{C})$	odd	1	$\frac{n(n-1)}{2}$	yes
$\mathrm{Sp}_n(\mathbf{C})$	arb.	$\{\pm 1\}$	$\frac{n(n+1)}{2}$	yes

TABLE 2. The classical groups over the complex numbers

the same dimension and then we must have that $n(2n+1) = m(2m-1)$, for some positive integers m and n. This implies that $2(m-n) = -1$, which is impossible. Consequently we can distinguish the matrix groups over $\mathbf{C}$. We see that we have used notions from group theory, topology, and from the theory of manifolds to separate the groups. It is therefore natural to introduce structures on the classical groups that take both the algebraic and geometric properties into account. We shall do this in Chapter 4.

In the case when the field $\mathbf{K}$ is the real numbers, the center of $\mathrm{Sl}_n(\mathbf{R})$ is also $\pm I_n$. In this case we can, as above, distinguish all groups except $\mathrm{Sl}_n(\mathbf{R})$ and $\mathrm{Sp}_{2m}(\mathbf{R})$, when $n^2-1 = \frac{2m(2m+1)}{2}$, and $\mathrm{Sl}_n(\mathbf{R})$ and $\mathrm{SO}_{2m}(\mathbf{R})$, when $n^2-1 = \frac{2m(2m-1)}{2}$ (see Exercise 3.8.3). The possibility that $\mathrm{Sl}_n(\mathbf{R})$ can be isomorphic to $\mathrm{SO}_n(\mathbf{R})$ can be ruled out by introducing compactness, which is a topological invariant. We shall see how this is done in the Section 3.9.

Exercises

3.8.1. Show that the unit interval $[0,1]$ is connected.

3.8.2. Let $\mathrm{SO}_2(\mathbf{R}, S)$ be the special orthogonal group with respect to the form $S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Show that $\mathrm{SO}_2(\mathbf{R}, S) = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbf{R}, a^2 - b^2 = 1 \right\}$, and that $\mathrm{SO}_2(\mathbf{R}, S)$ is not connected.

3.8.3. Determine all positive integers m and n such that

$$n^2 - 1 = \frac{2m(2m+1)}{2}.$$

3.8.4. Prove that X is connected if and only if X is not the union of two disjoint non-empty closed sets, or, if and only if the complement of an non-empty open set can not be open, or, if and only if the complement of a non-empty closed set can not be closed.

3.8.5. For x and y in a topological space X we write $x \equiv y$ if there is an arch in X that begins in x and ends in y. Show that $\equiv$ is an equivalence relation on X.

3.8.6. Let Y be a subset of a topological space X.

(a) Show that the intersection of all closed subsets of X that contain Y is the smallest closed subset $\bar{Y}$ of X that contains Y.
(b) Show that the connected components of X are closed.

3.9 Compact topological spaces

To distinguish the groups $\mathrm{Sl}_n(\mathbf{R})$ and $\mathrm{SO}_n(\mathbf{R})$ we introduce another topological invariant.

Definition 3.9.1. Let X be a topological space. A subset S of X is *compact* if, for every family of open sets $\{U_i\}_{i\in I}$ that *cover* S, that is, such that $S = \bigcup_{i\in I} U_i$, there is a finite subset $U_{i_1}, \dots, U_{i_n}$, for some n, that cover S.

Proposition 3.9.2. *A subset of $\mathbf{R}^n$ is compact if and only if it is closed and bounded.*

Proof: Assume that S is compact. First we show that S is bounded. Every point x in S is contained in a ball $B(x,1)$ of radius 1. We have that the family $\bigcup_{x\in S} B(x,1)$ covers S. Hence, there is a finite subcover $B(x_1,1), \dots, B(x_n,1)$. The union of this finite family of bounded sets is clearly bounded. Hence S is bounded. We next show that S is closed. Let y be a point of $\mathbf{R}^n$ not in S. For every $x \in S$ there are balls $B(y,\varepsilon_x)$ and $B(x,\varepsilon_x)$ such that $B(y,\varepsilon_x) \cap B(x,\varepsilon_x) = \emptyset$. The sets $\{B(x,\varepsilon_x)\}_{x\in S}$ cover S. Hence there is a finite subcover $B(x_1,\varepsilon_{x_1}), \dots, B(x_m,\varepsilon_{x_m})$. We have that $U = \bigcap_{i=1}^m B(y,\epsilon_{x_i})$ is an open subset containing y such that $U \cap B(x_i,\epsilon_{x_i}) = \emptyset$, for $i = 1, \dots, m$. Consequently $U \cap S = \emptyset$. Since every point of $\mathbf{R}^n$ that is not in S has a neighborhood that does not intersect S, we have that S is closed.

Conversely assume that S is a closed and bounded subset of $\mathbf{R}^n$. Let $\{U_i\}_{i\in I}$ be an open covering of S. Assume that S can not be covered by a finite subfamily of this covering. Since S is bounded we have that S is contained in a box of the form $B_0 = \{x \in \mathbf{R}^n \colon |x_i| \leq \frac{a}{2}\}$ of side-length a, for some a. Divide B_0 into 2^n boxes, B_{11}, B_{12^n} of side-length $\frac{a}{2}$. Since S can not be covered by a finite number of the U_i the same is true for at least one of the sets, say $B_1 = B_{1j} \cap S$. We subdivide B_1 into 2^n boxes $B_{21}, \dots, B_{22^n}$ of side-length $\frac{a}{2^2}$. Since $B_1 \cap S$ can not be covered by a finite number of the open sets U_i the same is true for at least one of the, say $B_2 = B_{2j_2}$. We continue this reasoning and obtain a sequence of boxes $B_0 \supset B_1 \supset B_2 \supset \cdots$, where B_i has side-length $\frac{a}{2^i}$, and such that $B_i \cap S$ can not be covered by a finite number of the U_i.

Let, for $j = 1, \dots, m$, the j'th side of B_i be $[a_{ij}, b_{ij}]$. Then $a_{1j} \leq a_{2j} \leq \cdots \leq b_{2j} \leq b_{1j}$, and $b_{ij} - a_{ij} = \frac{a}{2^i}$. Let b_j be the *greatest lower bound* for the set $b_{1j} \geq b_{2j} \geq \cdots$. Then $a_{ij} \leq b_j \leq b_{ij}$, for all i. Consequently, the point $(b_1, \dots, b_n)$ is in $\cap_{i=1}^\infty B_i$. We have that $b \in U_l$, for some l in I. Since the side of B_i is $\frac{a}{2^i}$, we can find a j such that $B_j \subseteq U_l$. In particular B_j can be covered by a finite number, in fact one, of the sets U_i for i in I. This contradicts the assumption that B_j can not be covered by a finite number of the U_i, and we have finished the proof. □

Group	n	Center	Dimension	Connected	Compact
$\mathrm{Gl}_n(\mathbf{R})$	arb.	$\mathbf{K}^*$	n^2	no	no
$\mathrm{Sl}_n(\mathbf{R})$	arb.	$\{\pm 1\}$	n^2-1	yes	no
$\mathrm{O}_n(\mathbf{R})$	arb.	$\{\pm 1\}$	$\frac{n(n-1)}{2}$	no	yes
$\mathrm{SO}_n(\mathbf{R})$	even	$\{\pm 1\}$	$\frac{n(n-1)}{2}$	yes	yes
$\mathrm{SO}_n(\mathbf{R})$	odd	$\{1\}$	$\frac{n(n-1)}{2}$	yes	yes
$\mathrm{Sp}_n(\mathbf{R})$	arb.	$\{\pm 1\}$	$\frac{n(n+1)}{2}$	yes	no

TABLE 3. The classical groups over the real numbers

Example 3.9.3. The groups $\mathrm{Gl}_n(\mathbf{R})$ and $\mathrm{Sl}_n(\mathbf{R})$ are not compact, for $n > 1$. Indeed, they contain the matrices $E_{i,j}(a)$ for all $i \neq j$, and consequently are not bounded.

Example 3.9.4. Both of the groups $\mathrm{O}_n(\mathbf{R})$ and $\mathrm{SO}_n(\mathbf{R})$ are compact. Indeed, they are defined as the zeroes of the n^2 polynomials that are the coefficients of the matrix identity $X^tX = 1$, and $\mathrm{SO}_n(\mathbf{R})$ is the zero also of the polynomial $\det X - 1$. Hence the groups are closed. However, the relations $x_{i1}^2 + \cdots + x_{in}^2 = 1$, for $i = 1, \dots, n$, which are obtained by considering the diagonal entries of the matrix relation, show that the points of $\mathrm{O}_n(\mathbf{R})$, and thus those of $\mathrm{SO}_n(\mathbf{R})$, are contained in the unit cube in $\mathbf{R}^n$.

Example 3.9.5. The group $\mathrm{Sp}_n(\mathbf{R})$ is not compact. Indeed, it contains the element $E_{i,n+1-j}(a)$, for all i, and hence is not bounded.

Example 3.9.6. We can now return to the case of matrix groups over the real numbers as we mentioned in Example 3.8.15. Over the real numbers we get a Table 3.

In this case we can, as above, distinguish all groups except $\mathrm{Sl}_n(\mathbf{R})$ and $\mathrm{Sp}_{2m}(\mathbf{R})$, when $n^2 - 1 = \frac{2m(2m+1)}{2}$ (see Exercise 3.3.7).

Exercises

3.9.1. Let $\mathrm{O}_2(\mathbf{R}, \langle, \rangle)$ be the orthogonal group over the real numbers with respect to the form defined by the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Show that $\mathrm{O}_2(\mathbf{R}, \langle, \rangle)$ contains the matrices $\begin{pmatrix} \frac{1}{2}(t+\frac{1}{t}) & \frac{1}{2}(t-\frac{1}{t}) \\ \frac{1}{2}(t-\frac{1}{t}) & \frac{1}{2}(t+\frac{1}{t}) \end{pmatrix}$, and that $\mathrm{O}_2(\mathbf{R}, \langle, \rangle)$ is not compact.

3.9.2. Let H be the set of $(n+2) \times (n+2)$-matrices of the form

$$\begin{pmatrix} 1 & {}^tx & c \\ 0 & I_n & y \\ 0 & 0 & 1 \end{pmatrix} \quad \text{for } x, y \in \mathbf{C}^n \text{ and } c \in \mathbf{C}.$$

(a) Show that H is a subgroup of $\mathrm{Gl}_n(\mathbf{C})$.
(b) Determine the tangent space $\mathfrak{h}$ of H.
(c) Show that the map $\exp : \mathfrak{h} \to H$ is a polynomial and globally invertible.
(d) Determine the center of H.
(e) Determine whether H is compact.
(f) Determine whether H is connected.
(g) Find a natural set of generators of H.

3.9.3. Let $B_n \subseteq \mathrm{Gl}_n = \mathrm{Gl}_n(\mathbf{C})$ be the subset of upper triangular matrices

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

(a) Show that B_n is a subgroup of Gl_n.
(b) Determine the center of B_n.
(c) Determine the tangent space $\mathfrak{b}_n$ of B_n.
(d) Show that the map $\exp : \mathfrak{b}_n \to B_n$ is surjective.
(e) Show that B_n acts on Gl_n by left multiplication.
(f) Show that all orbits are homeomorphic to B_n.
(g) Show that B_n is not a normal subgroup of Gl_n.
(h) Give $Fl_n = \mathrm{Gl}_n / B_n$ the structure of an analytic manifold and determine the dimension of its tangent space.
(i) Determine whether Fl_n is compact.
(j) Determine whether Fl_n is connected.

4 Lie groups

We shall in this Chapter join the algebraic point of view of Chapter 1 with the geometric point of view of Chapter 3, and consider manifolds that have a natural group structure. In this way we obtain a powerful tool for studying the matrix groups.

4.1 Lie groups

In Chapter 3 we gave a natural geometric framework for the classical groups introduced in Chapter 1, when the groups have coefficients in the real or complex numbers. We shall in this chapter join the algebraic point of view of Chapter 1 with the geometric point of view of Chapter 3. In this way we obtain a natural and powerful tool for studying matrix groups.

Throughout this chapter the field $\mathbf{K}$ will be the real or complex numbers unless we explicitly state otherwise.

Definition 4.1.1. Let G be a manifold which is also a group and let $G \times G$ be the product manifold (see Example 3.4.7). We say that G is a *Lie group* when the product map

$$G \times G \to G,$$

which sends (a, b) to ab, and the inverse map

$$G \to G,$$

which sends a to a^{-1}, are analytic.

Remark 4.1.2. We note that the inverse map is an analytic isomorphism. In fact, it is its own inverse.

Example 4.1.3. The manifolds $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SL}_S(\mathbf{K})$, and hence, in particular, $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, $\mathrm{Sp}_n(\mathbf{K})$ are all Lie groups (see Example 3.4.5). Indeed the multiplication map is given by polynomials, and the inverse is given by a rational function with denominator the determinant $\det(X_{ij})$ of a $n \times n$ matrix with variables as coefficients.

Example 4.1.4. The manifold $\mathbf{K}$ with addition as group operation is a Lie group.

Example 4.1.5. Let H be a Lie group, and G a submanifold of H, which is also a subgroup of H. Then G is also a Lie group. Indeed, the multiplication map $G \times G \to G$ and the inverse map $G \to G$ of G are the composite of the inclusions $G \times G \subseteq H \times H$ and $G \subseteq H$, with the multiplication, respectively inverse, on H. Since the inclusion maps are analytic, by the definition of submanifolds, the multiplication and inverse on G are analytic.

Definition 4.1.6. Let G and H be Lie groups. We say that G is a *Lie subgroup* of H if it is a submanifold, and the inclusion map is also a group homomorphism.

The most remarkable feature of a Lie group is that the structure is the same in the neighborhood of each of its points. To make this precise we introduce the left translations.

Definition 4.1.7. Let G be a group and a an element of G. The map

$$\lambda_a \colon G \to G$$

defined by $\lambda_a(b) = ab$ is called a *left translation* by a.

Remark 4.1.8. When G is a Lie group the left translations are analytic. Indeed λ_a is the composite of the inclusion $a \times G \to G \times G$ with the multiplication $G \times G \to G$, and both the latter maps are analytic. The map λ_a is also an isomorphism of the manifold G, because it has the analytic inverse $\lambda_{a^{-1}}$.

let a and b be points of a Lie group G. Then the map $\lambda_{ba^{-1}}$ is an isomorphism of the manifold G, which sends a to b. We obtain, for each open set U of G, an isomorphism of rings

$$(\lambda^*_{ba_{-1}})_U \colon \mathcal{O}_G(\lambda_{ba^{-1}}(U)) \to \mathcal{O}_G(U)$$

which sends an analytic function $f\colon \lambda_{ba^{-1}}(U) \to \mathbf{K}$ to the function $f\lambda_{ba^{-1}} \colon U \to \mathbf{K}$, which sends c to $f(ba^{-1}c)$. In particular, we obtain an isomorphism

$$(\lambda_{ba^{-1}})_b \colon \mathcal{O}_{G,b} \to \mathcal{O}_{G,a}$$

of rings. Consequently, we have an isomorphism of vector spaces

$$T_a\lambda_{ba^{-1}} \colon T_aG \to T_bG,$$

sending a derivation D in T_aG to the derivation in T_bG which maps a function f of $\mathcal{O}_{G,b}$ to $D(f\lambda_{ba^{-1}})$.

Definition 4.1.9. Let G and H be Lie groups. A *homomorphism* of Lie groups is a map $\Phi\colon G \to H$ which is an analytic map of manifolds and a homomorphism of groups. We say that a homomorphism of Lie groups is an *isomorphism* if it has an inverse map, which is a homomorphism of Lie groups.

Example 4.1.10. The maps of Examples 1.2.10, 1.2.11, 1.2.12, and the inclusion in $\mathrm{Gl}_n(\mathbf{K})$, of all the groups $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$ are all homomorphisms of Lie groups. In particular, $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, $\mathrm{Sp}_n(\mathbf{K})$ are all Lie groups.

Remark 4.1.11. Let a and b be two points of a Lie group G. Then we have that $\lambda_{ab} = \lambda_a\lambda_b$. Moreover, given a map $\Phi\colon G \to H$ of Lie groups. For each point a of G we have that $\Phi\lambda_a = \lambda_{\Phi(a)}\Phi$, i.e., the diagram

$$\begin{array}{ccc} G & \xrightarrow{\lambda_a} & G \\ {\scriptstyle\Phi}\downarrow & & \downarrow{\scriptstyle\Phi} \\ H & \xrightarrow[\lambda_{\Phi(a)}]{} & H \end{array} \tag{4.1.11.1}$$

is commutative.

4.2 Lie algebras

We noticed in Example 2.6.2 that the tangent spaces of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$, and in particular $\mathrm{O}_n(\mathbf{K})$, $\mathrm{SO}_n(\mathbf{K})$, $\mathrm{Sp}_n(\mathbf{K})$ are Lie subalgebras of $\mathrm{M}_n(\mathbf{K})$ in the sense defined in Remark 2.6.1. In this section we give the general Definition of Lie algebras and in section 4.4 we show that the tangent space of any Lie group has a natural structure as a Lie algebra.

Definition 4.2.1. Let $\mathfrak{v}$ be a vector space with a bilinear form

$$[\,,]\colon \mathfrak{v} \times \mathfrak{v} \to \mathfrak{v}$$

that we sometimes call the *product* on $\mathfrak{v}$ (see 1.7.1). We say that $\mathfrak{v}$ is a *Lie algebra* if the following two conditions hold for all vectors X, Y and Z of $\mathfrak{v}$:

(i) $[X, X] = 0$,
(ii) $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$.

A subalgebra $\mathfrak{v}$ of a Lie algebra $\mathfrak{w}$ is a subspace such that $[X, Y]$ is in $\mathfrak{v}$, for all X and Y of $\mathfrak{v}$.

Remark 4.2.2. Let $\mathfrak{v}$ be a Lie subalgebra of $\mathfrak{w}$. Then the product $[,]$ on $\mathfrak{w}$ induces, by definition, a product $[,]$ on $\mathfrak{v}$. With this product we have that $\mathfrak{v}$ is a Lie algebra. Indeed, this product is bilinear and satisfies the two properties of the definition 4.2.1 of Lie algebras because it does so for all elements of $\mathfrak{w}$.

Example 4.2.3. The spaces $\mathfrak{gl}_n(\mathbf{K})$, $\mathfrak{sl}_n(\mathbf{K})$, $\mathfrak{so}_n(\mathbf{K})$, and $\mathfrak{sp}_n(\mathbf{K})$ of Example 2.6.2 are all Lie subalgebras of the Lie algebra $\mathfrak{gl}_n(\mathbf{K})$ of Remark 2.6.1.

Example 4.2.4. Let A be an algebra over $\mathbf{K}$, and denote by $\mathrm{Der}_K(A, A)$ the vector space of $\mathbf{K}$ derivation on A (see Definition 3.6.2). Given derivations X and Y, we let $[X, Y]$ denote the map $(XY - YX)\colon A \to A$. We have that $[X, Y]$ is, in fact, a $\mathbf{K}$ derivation. Indeed, for all a and b in A, we have that $XY(ab) - YX(ab) = X(aYb + bYa) - Y(aXb + bXa) = XaYb + aXYb + XbYa + bXYa - YaXb - aYXb - YbXa - bYXa = (a(XY - YX)b + b(XY - YX)a)$. With this product $\mathrm{Der}_K(A, A)$ becomes a Lie algebra. Indeed the first axiom is obvious, and the second a long, but easy, calculation (see Exercise 4.2.1).

Definition 4.2.5. Let $\mathfrak{g}$ and $\mathfrak{h}$ be Lie algebras and $\varphi\colon \mathfrak{g} \to \mathfrak{h}$ a linear map. We say that φ is a *Lie algebra homomorphism* if $\varphi[X, Y] = [\varphi X, \varphi Y]$, for all X and Y of $\mathfrak{g}$. A *Lie algebra isomorphism* is a homomorphism which is an isomorphism of vector spaces.

Exercises

4.2.1. Let $\mathbf{K}$ be an arbitrary field. Show that $\mathrm{Der}_{\mathbf{K}}(A, A)$ with the product $[X, Y] = XY - YX$ is a Lie algebra.

4.3 Vector fields

In order to define a structure of Lie algebra on the tangent space of a Lie group we shall introduce vector fields on manifolds. Intuitively a vector field on a manifold M consists of a tangent vector $X(x)$ for every point x of M, such that the vectors depend analytically on the points. More precisely, for every analytic function $f\colon U \to \mathbf{K}$ defined on an open set U, the function on U sending x to $X(x)f$ should be analytic.

Definition 4.3.1. A *vector field* on a manifold M consists of a *derivation*

$$X_U\colon \mathcal{O}_M(U) \to \mathcal{O}_M(U),$$

on the ring $\mathcal{O}_M(U)$, for all open subsets U of M, such that, if V is an open subset of U, then

$$\rho_{U,V} X_U = X_V \rho_{U,V},$$

where the $\rho_{U,V}$ are the restriction maps of Remark 3.4.9.

Remark 4.3.2. A collection of maps $\varphi_U \colon \mathcal{O}_M(U) \to \mathcal{O}_M(U)$, one for each open subset U of M, such that $\rho_{U,V}\varphi_U = \varphi_V \rho_{U,V}$, for all open subsets V of U, is called a *map of the sheaf* $\mathcal{O}_M$.

4.3.3. Let X and Y be vector fields on a manifold M. We define the sum $X + Y$ of X and Y by $(X + Y)_U = X_U + Y_U$, and the product aX of a scalar a of $\mathbf{K}$ with X by $(aX)_U = aX_U$, for all open sets U of M. It is clear that the vector fields $\mathfrak{v}(M)$ on M, with these operations, become a vector space over $\mathbf{K}$.

4.3.4. Fix a point x of M, and let X be a vector field on M. The maps $X_U \colon \mathcal{O}_M(U) \to \mathcal{O}_M(U)$, for all open subsets U of M that contain x, define a $\mathbf{K}$ derivation

$$X_x \colon \mathcal{O}_{M,x} \to \mathcal{O}_{M,x}$$

on the ring $\mathcal{O}_{M,x}$. The composite of X_x with the augmentation map $\mathcal{O}_{M,x} \to \mathbf{K}$ is a $\mathbf{K}$ derivation

$$X(x) \colon \mathcal{O}_{M,x} \to \mathbf{K},$$

for the augmentation map. By the definition of $X(x)$ we have that

$$X(x)f = (Xf)(x),$$

for all functions f that are analytic in a neighborhood of x. We consequently obtain, for each point x in M, a map

$$\epsilon_{M,x} \colon \mathfrak{v}(M) \to T_x M,$$

from vector fields $\mathfrak{v}(M)$ on M to the tangent space $T_x M$ of M at x, which maps a vector field X to the derivation whose value at an analytic function f, defined in a neighborhood of x is $(Xf)(x)$ is a $\mathbf{K}$ linear map.

4.3.5. let X and Y be vector fields on a manifold M. For all open subsets U of M the composite $(XY)_U = X_U Y_U$ of X_U and Y_U defines a linear map $\mathcal{O}_M(U) \to \mathcal{O}_M(U)$, such that $\rho_{U,V}(XY)_U = (XY)_V \rho_{U,V}$, for all open subsets V of U such that V is contained in U. That is, we obtain a map XY of sheaves. This map is however, not a derivation. On the other hand the map $(XY - YX)_U \colon \mathcal{O}_M(U) \to \mathcal{O}_M(U)$ is a derivation. Indeed, we saw in Example 4.2.4 that $\mathrm{Der}_K(\mathcal{O}_M(U), \mathcal{O}_M(U))$ is a Lie algebra under the operation $[A, B] = AB - BA$, and X_U and Y_U lie in $\mathrm{Der}_K(\mathcal{O}_M(U), \mathcal{O}_M(U))$. Hence, the maps $(XY - YX)_U$, for all open sets U of M, define a vector field. We shall denote this vector field by $[X, Y]$. Since the subset of $\mathrm{Der}_K(\mathcal{O}_M(U), \mathcal{O}_M(U))$ consisting of derivations of the form X_U, where X is a vector field on M, form a Lie subalgebra, it follows that the space of vector fields on M is a Lie algebra with product $[\,,]$.

Definition 4.3.6. We denote the Lie algebra of vector fields on a manifold M by $\mathfrak{v}(M)$.

Remark 4.3.7. Let $\Phi\colon M \to N$ be a homomorphism of analytic manifolds. For each point x of M we have maps

$$\epsilon_{M,x}\colon \mathfrak{v}(M) \to T_xM, \quad T_x\Phi\colon T_xM \to T_{\Phi(x)}N, \quad \text{and } \epsilon_{N,\Phi(x)}\colon \mathfrak{v}(N) \to T_{\Phi(x)}N.$$

There is no natural map from $\mathfrak{v}(M)$ to $\mathfrak{v}(N)$. However, we can relate vector fields on M and N in the following way:

Let X and Y be vector fields on M respectively N and let let f be a function which is analytic in an open subset V of N. It follows from the definitions that the following equalities are equivalent:

(i) $T_x\Phi\epsilon_{M,x}X = \epsilon_{N,\Phi(x)}Y$, for x in $\Phi^{-1}(V)$,
(ii) $(\epsilon_{M,x}X)(f\Phi) = (\epsilon_{N,\Phi(x)})Yf$, for x in $\Phi^{-1}(V)$,
(iii) $X(f\Phi) = (Yf)\Phi$ on $\Phi^{-1}(V)$,
(iv) $X(x)f\Phi = Y(\Phi(x))f$, for all x in $\Phi^{-1}(V)$.

Lemma 4.3.8. *Let $\Phi\colon M \to N$ be an analytic map of manifolds. Given vector fields X_i on M, and Y_i on N, for $i = 1, 2$. Assume that $T_x\Phi\epsilon_{M,x}X_i = \epsilon_{N,\Phi(x)}Y_i$, for all x in M, and for $i =, 1, 2$. Then we have that*

$$T_x\Phi\epsilon_{M,x}[X_1, X_2] = \epsilon_{N,\Phi(x)}[Y_1, Y_2].$$

Proof: It follows from Remark 4.3.7 that the condition of the lemma is equivalent to asserting that we, for every function f that is analytic in a neighborhood of a point $\Phi(x)$, have that $X(f\Phi) = (Yf)\Phi$, in a neighborhood of x. The proof now consists in unraveling the definitions involved as follows:

$$\begin{aligned}
T_x\Phi\epsilon_{M,x}[X_1, X_2]f &= [X_1, X_2]f\Phi(x) = (X_1X_2)(f\Phi)(x) - (X_2X_1)(f\Phi)(x)\\
&= X_1(X_2(f\Phi))(x) - X_2(X_1(f\Phi))(x)\\
&= X_1((Y_2f)\Phi)(x) - X_2((Y_1f)\Phi))(x)\\
&= Y_1Y_2f(\Phi(x)) - Y_2Y_1f(\Phi(x)) = [Y_1, Y_2]f(\Phi(x)) = \epsilon_{N,\Phi(x)}[Y_1, Y_2].
\end{aligned}$$

□

4.4 The Lie algebra of a Lie group

In this section we shall show that the tangent space of a Lie group has a structure of a Lie algebra, and that a homomorphism of Lie groups induces a homomorphism of Lie algebras.

Definition 4.4.1. Let G be a Lie group. We shall say that a vector field X on G is *left invariant* if, for every point a of G and every analytic function $f\colon U \to \mathbf{K}$ on an open subset U of G, we have that

$$(Xf)\lambda_a = X(f\lambda_a)$$

on the open subset $a^{-1}U$ of G. That is, for all b in G such that $ab \in U$, we have that $(Xf)(ab) = X(f\lambda_a)(b)$. Here λ_a is the left translation of Definition 4.1.7.

4.4.2. The left invariant vector fields on a Lie group G form a Lie subalgebra of $\mathfrak{v}(G)$. Indeed, if X and Y are left invariant vector fields on G, we have that

$$\begin{aligned}[X,Y]f(ab) &= XYf(ab) - YXf(ab) = X((Yf)\lambda_a)(b) - Y((Xf)\lambda_a)(b)\\ &= X(Y(f\lambda_a))(b) - Y(X(f\lambda_a))(b) = [X,Y](f\lambda_a)(b).\end{aligned}$$

Hence we have that $[X,Y]$ is left invariant.

Definition 4.4.3. The Lie algebra of left invariant vector fields on G is called the *Lie algebra* of G and is denoted by . The map $\epsilon_{G,e}\colon \mathfrak{v}(G) \to T_e(G)$ of Paragraph 4.3.4 induces a map

$$\lambda_G\colon \mathfrak{g} \to T_e(G).$$

Remark 4.4.4. Let G be a Lie group and let $\varphi\colon U \to G$ be a chart. The multiplication map $G\times G \to G$ is continuous. Consequently, we can choose charts $\psi\colon V \to G$ and $\chi\colon W \to G$, such that $\chi(0) = e$, and such that the multiplication induces as map

$$\mu\colon V \times W \to U.$$

Choose coordinates $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, and $z = (z_1, \dots, z_n)$ in U, V, and W respectively. We write $\mu(y,z) = (\mu_1(y,z), \dots, \mu_n(y,z))$.

Given a tangent D in $T_e(G)$, we can write

$$D = T_o\chi(c_1\frac{\partial}{\partial z_1} + \cdots + c_n\frac{\partial}{\partial z_n}),$$

for some $c_1, \dots, c_n$ in $\mathbf{K}$. For each a in V and each analytic function f defined on an open subset of G containing $\psi(a) = \varphi(\mu(a,0))$ we obtain equations

$$D(f\varphi\mu(a,z)) = \sum_{i=1}^n c_i\frac{\partial(f\varphi)\mu(a,z)}{\partial z_i}(a,0) = \sum_{i=1}^n\sum_{j=1}^n c_i\frac{\partial(f\varphi)}{\partial x_j}\mu(a,0)\frac{\partial\mu_j}{\partial z_i}(a,0).$$

The map $\mu(y,z)$ is analytic in the variables y and z. Consequently, we have that $\Psi_{ij}(y) = \frac{\partial\mu_j}{\partial z_i}(y,0)$ is analytic in y, and we have proved the expression

$$D(f\lambda_a) = \sum_{i=1}^n\sum_{j=1}^n c_i\Psi_{ij}(a)\frac{\partial(f\varphi)}{\partial x_j}(a), \tag{4.4.4.1}$$

with $\Psi_{ij}(y)$ analytic.

Lemma 4.4.5. *Let G be a Lie group and D a tangent vector in T_eG. For each analytic map $f\colon U \to \mathbf{K}$, on an open subset U of G, the function $(Xf)_U\colon U \to \mathbf{K}$, defined by $(Xf)_U(a) = D(f\lambda_a)$, is analytic. The map*

$$X_U\colon \mathcal{O}_G(U) \to \mathcal{O}_G(U),$$

obtained in this way, for each open subset U of G, define a left invariant vector field on G.

Proof: We have that $D(f\lambda_a)$ depends only on the value of the function $f\lambda_a$ near the unit e of G. Choose charts and coordinates as in Remark 4.4.4. We obtain from Equation 4.4.4.1

$$Xf(a) = D(f\lambda_a) = D(f\mu(a,z)) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i \Psi_{ij}(a)\frac{\partial(f\varphi)}{\partial x_j}(a),$$

with the $\Psi_{ij}(a)$ analytic in a. We obtain that $D(f\mu(a,z)) = D(f\lambda_a)$ is an analytic function of a and we have proved the first part of the lemma.

It is clear, from the definition of the functions X_U and the restriction functions $\rho_{U,V}$, that $\rho_{U,V}X_U = X_V\rho_{U,V}$. Consequently, the second assertion of the lemma holds.

It remains to prove that X is left invariant. Let $f\colon U \to \mathbf{K}$ be analytic and let a be in G. We must prove that $(Xf)\lambda_a = X(f\lambda_a)$ on $a^{-1}U$. Let b be in $a^{-1}U$. We have that $(Xf)\lambda_a(b) = (Xf)(ab) = D(f\lambda_{ab}) = D(f\lambda_a\lambda_b) = D((f\lambda_a)\lambda_b) = X(f\lambda_a)(b)$. Hence, X is left invariant. □

Remark 4.4.6. Lemma 4.4.5 asserts that, to a derivation D in T_eG, we can associate a left invariant vector field X. In this way we obtain a map $\delta_G\colon T_eG \to \mathfrak{g}$. This map is clearly linear.

Choose charts and coordinates as in Remark 4.4.4. Let $X = \delta_G(D)$ be the left invariant vector field associated to $D = T_0\chi(c_1\frac{\partial}{\partial z_1} + \cdots + c_n\frac{\partial}{\partial z_n})$. That is, we have that $Xf(a) = D(f\lambda_a)$, for all analytic functions $f\colon U \to \mathbf{K}$, and all $a \in V$. The Equation 4.4.4.1 can be rewritten as

$$X = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\Psi_{ij}\frac{\partial}{\partial x_j},$$

where this expression means that $Xf(\varphi(a)) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\Psi_{ij}(a)\frac{\partial(f\varphi)}{\partial x_j}\mu(a,0)$, for all analytic functions $f\colon U \to G$ and all $a \in U$.

Proposition 4.4.7. *The map* $\epsilon_{G,e}\colon \mathfrak{v}(G) \to T_e(G)$ *of Paragraph 4.3.4 induces an isomorphism of* $\mathbf{K}$ *vector spaces*

$$\epsilon_G\colon \mathfrak{g} \to T_e(G).$$

The inverse of ϵ_G *is the map* $\delta_G\colon T_e(G) \to \mathfrak{g}$ *defined in Remark 4.4.6.*

Proof: It suffices to show that ϵ_G and the map δ_G defined in Remark 4.4.6 are inverse maps.

Let D be a vector in T_eG, and let $X = \delta_G(D)$ be the vector field associated to D in Remark 4.4.6. For f in $\mathcal{O}_{G,x}$ we have that $\varepsilon_G(x)(f) = X(e)f = Xf(e) = D(f\lambda_e) = Df$. Hence we have that $\varepsilon_G\delta_G = \mathrm{id}$.

Conversely, let X be a vector field on G, and let $D = \varepsilon_G(X) = X(e)$. Let $Y = \delta_G(D)$ be the vector field associated to D in Remark 4.4.6. For all analytic functions $f\colon U \to \mathbf{K}$ defined on an open subset U of G, and for all points a of U we have that $Xf(a) = (Xf)\lambda_a(e) = X(f\lambda_a)(e) = X(e)(f\lambda_a) = D(f\lambda_a) = Yf(a) = \delta_G\varepsilon_G(X)(a)$. Hence, we have proved the proposition. □

Remark 4.4.8. It follows from Proposition 4.4.7 that we can use the map $\delta_G\colon \mathfrak{v}(G) \to T_eG$ to give the space T_eG a structure as a Lie group.

Definition 4.4.9. Let G and H be Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$. Moreover, let $\Phi\colon G \to H$ be a homomorphism of Lie groups. Then we have a map

$$\mathfrak{l}(\Phi)\colon \mathfrak{g} \to \mathfrak{h}$$

defined by $\mathfrak{l}(\Phi) = \delta_H^{-1} T_e \Phi \delta_G$. If $\Psi\colon F \to G$ is another map of Lie groups we clearly have that $\mathfrak{l}(\Psi\Phi) = \mathfrak{l}(\Psi)\mathfrak{l}(\Phi)$.

Proposition 4.4.10. *Let $\Phi\colon G \to H$ be a homomorphism of Lie groups. The map*

$$\mathfrak{l}(\Phi)\colon \mathfrak{g} \to \mathfrak{h}$$

of the corresponding Lie algebras is a Lie algebra homomorphism.

Proof: It is clear that $\mathfrak{l}(\Phi)$ is a map of vector spaces. To show that it is a map of Lie algebras, let X_i, for $i = 1, 2$, be left invariant vector fields on G and let $Y_i = \mathfrak{l}(\Phi)X_i$. Since the maps δ_G and δ_H are induced by the maps $\epsilon_{G,e}$ and $\epsilon_{H,e}$ of Paragraph 4.3.4 and Remark 4.3.7, we have that the proposition follows from Lemma 4.3.8, once we can show that $T_e\Phi\delta_G X_i = \delta_H Y_i$. However, we have that $T_e\Phi\delta_G X_i = \delta_H \mathfrak{l}(\Phi) X_i = \delta_H Y_i$, and we have finished the proof. □

4.5 One parameter subgroups of Lie groups

In this section we shall construct one parameter subgroups of any Lie group and thus generalize the construction of one parameter subgroups of the matrix groups given in Section 2.7. For the construction we need a well known result about differential equations, which is proved for differentiable functions in any beginning course in differential equations, or in advanced calculus courses. We shall start by giving a proof of the result because we shall use it in the less frequently presented case of analytic functions.

Proposition 4.5.1. *For $p = 1, \dots, n$, we give analytic functions $f_p\colon U \to \mathbf{K}$ defined on an open subset U of $\mathbf{K}^n$. The differential equation*

$$g_p'(t) = f_p(g_1(t), \dots, g_n(t))$$

in the functions $g_1, \dots, g_n$ in the variable t, with initial conditions $g_p(0) = a_p$, for $p = 1, \dots, n$, with $a_p \in \mathbf{K}$, has a unique solution $g_1(t), \dots, g_n(t)$, for t in a neighborhood V of 0 in $\mathbf{K}$, and the functions g_p are analytic on V.

Proof: Write

$$f_p(x) = \sum_{i \in \mathcal{I}} c_{pi} x^i, \quad \text{for } p = 1, \dots, n.$$

Let

$$g_p(t) = \sum_{q=0}^{\infty} d_{pq} t^q, \quad \text{for } p = 1, \dots, n$$

be formal power series. In order to satisfy the differential equation of the proposition, we must have that

$$\sum_{q=1}^{\infty} q d_{pq} t^{q-1} = \sum_{i \in \mathcal{I}} c_{pi} \left(\sum_{q=0}^{\infty} d_{1q} t^q \right)^{i_1} \cdots \left(\sum_{q=0}^{\infty} d_{nq} t^q \right)^{i_n}.$$

Hence there are unique polynomials $Q_m(c_{pi}, d_{1q}, \dots, d_{nq})$, for $|i| < m$ and $q < m$, with positive integers as coefficients, such that

$$d_{pm} = \frac{1}{m} Q_m(c_{pi}, d_{1q}, \dots, d_{nq}), \quad \text{for } p = 1, \dots, n.$$

By induction on m, starting with the initial condition $d_{p0} = a_p$, we obtain that the d_{pm} are uniquely determined such that the formal series $g_1(t), \dots, g_n(t)$ satisfy the differential equation of the proposition.

It remains to prove that the formal series $g_1, \dots, g_n$ define analytic functions.

Assume that we have real numbers $\bar{c}_{pi}$, for $p = 1, \dots, n$ and for all i in $\mathcal{I}$, such that

$$|c_{pi}| \le \bar{c}_{pi},$$

for all p and i, and such that the functions

$$\bar{f}_p(x) = \sum_{i \in \mathcal{I}} \bar{c}_{pi} x^i, \quad \text{for } p = 1, \dots, n$$

are analytic. As we saw above, we can find unique formal series

$$\bar{g}_p(t) = \sum_{q=o}^{\infty} \bar{d}_{pq} t^q$$

that solve the differential equation

$$\bar{g}_p'(t) = \bar{f}_p(\bar{g}_1(t), \dots, \bar{g}_n(t)). \tag{4.5.1.1}$$

If the functions $\bar{g}_1, \dots, \bar{g}_n$ were analytic, the same would be true for $g_1, \dots, g_n$. Indeed, we have inequalities

$$\begin{aligned} |d_{pm}| &= \frac{1}{m} |Q_m(c_{pi}, d_{1q}, \dots, d_{nq})| \\ &\le \frac{1}{m} Q_m(|c_{pi}|, |d_{1q}|, \dots, |d_{nq}|) \le \frac{1}{n} Q_m(\bar{c}_{pi}, \bar{d}_{1q}, \dots, \bar{d}_{nq}) = \bar{d}_{pm}. \end{aligned}$$

Hence, in order to prove the proposition, it suffices to construct analytic functions $\bar{f}_p(x) = \sum_{i \in \mathcal{I}} \bar{c}_{pi} x^i$ such that $|c_{pi}| \le \bar{c}_{pi}$, for all p and i, and such that the corresponding solutions $\bar{g}_1, \dots, \bar{g}_n$ of Equation 4.5.1.1 are analytic.

To construct the $\bar{f}_p$ we note that the functions f_p are analytic on some neighborhood of 0 in $\mathbf{K}^n$. Consequently there are positive constants C and r such that

$$\sum_{i \in \mathcal{I}} |c_{pi}| r^{|i|} < C.$$

Let

$$\bar{c}_{pi} = \frac{C}{r^{|i|}}.$$

We have that $|c_{pi}|r^{|i|} \leq \sum_{i \in \mathcal{I}} |c_{pi}|r^{|i|} < C = \bar{c}_{pi} r^{|i|}$. Consequently, we have that $|c_{pi}| \leq \bar{c}_{pi}$, for all p and i. Moreover, we have that

$$\bar{f}_p(x) = \sum_{i \in \mathcal{I}} \bar{c}_{pi} x^i = C \sum_{i \in \mathcal{I}} \left(\frac{x}{r}\right)^i = \frac{C}{\prod_{q=1}^m \left(1 - \frac{x_q}{r}\right)}.$$

Hence $\bar{f}_1, \dots, \bar{f}_n$ are analytic. Moreover, the power series

$$\bar{g}_p(t) = \bar{g}(t) = r\left(1 - \left(1 - (n+1)C\frac{t}{r}\right)^{\frac{1}{n+1}}\right),$$

is analytic in a neighborhood of 0 and satisfies the differential equation 4.5.1.1, that is

$$\bar{g}'(t) = \frac{C}{\left(1 - \frac{\bar{g}(t)}{r}\right)^n}.$$

Indeed, we have that

$$\bar{g}'(t) = C\left(1 - (n+1)C\frac{t}{r}\right)^{-\frac{n}{n+1}},$$

and

$$\left(1 - (n+1)C\frac{t}{r}\right)^{\frac{n}{n+1}} = \left(1 - \frac{\bar{g}(t)}{r}\right)^n.$$

□

Definition 4.5.2. A *one parameter subgroup* of a Lie group G is an analytic mapping $\gamma \colon \mathbf{K} \to G$, which is also a group homomorphism. The *tangent* of a one parameter subgroup is the tangent $\gamma'(0)$ of the corresponding curve at the unit element, as defined in 3.6.9.

Remark 4.5.3. Let G be a Lie group and let $\gamma \colon T \to G$ be an analytic map from an open subset T of $\mathbf{K}$ containing 0. After possibly shrinking T we choose a chart $\varphi \colon U \to G$ such that $\gamma(T) \subseteq \varphi(U)$ and $\varphi(0) = \gamma(0)$. As in Remark 4.4.4 we choose charts $\psi \colon V \to G$ and $\chi \colon W \to W$, such that $\chi(0) = e$, and such that the multiplication induces as map

$$\mu \colon V \times W \to U.$$

Moreover, we let $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, and $z = (z_1, \dots, z_n)$ be coordinates in U, V, and W respectively. Write $\mu(y, z) = (\mu_1(y, z), \dots, \mu_n(y, z))$ and let $\varphi^{-1}\gamma = (\gamma_1, \dots, \gamma_n)$.

Assume that we have $\gamma(s+t) = \gamma(s)\gamma(t)$, for all s and t in T in a neighborhood of 0, that is, we have an equation $\mu_j(\gamma_1(s), \dots, \gamma_n(s), \gamma_1(t), \dots, \gamma_n(t)) = \gamma_j(s+t)$. We differentiate the latter equation with respect to t at $t = 0$, and obtain

$$\frac{d\gamma_j}{dt}(s) = \sum_{i=1}^n \frac{\partial \mu_j}{\partial y_i}(\gamma_1(s), \dots, \gamma_n(s), 0)\frac{d\gamma_i}{dt}(0) = \sum_{i=1}^n \Psi_{ij}(\gamma(s))c_i, \tag{4.5.3.1}$$

with $\Psi_{ij}(y) = \frac{\partial \mu_j}{\partial z_i}(y, 0)$ and $c_i = \frac{d\gamma_j}{dt}(0)$.

Fix a basis for $T_e(G)$. It follows from Proposition 4.5.1 that, when we fix $c_1, \dots, c_n$, the curve $\gamma\colon T \to G$ is determined uniquely in a neighborhood of 0 in $\mathbf{K}$ by the condition that $\gamma'(0) = (c_1, \dots, c_n)$ in the fixed basis of $T_e(G)$.

Proposition 4.5.4. *Let G be a Lie group and D a tangent of G at the identity e. Then there is a unique one parameter subgroup $\gamma\colon \mathbf{K} \to G$ of G whose tangent $\gamma'(0)$ at 0 is equal to D.*

Proof: It follows from Remark 4.5.3 that a one parameter subgroup of G, with derivative D at 0, is uniquely determined in a neighborhood of 0 in $\mathbf{K}$.

Choose charts and coordinates of G as in Remark 4.5.3. Let $\gamma_1(t), \dots, \gamma_n(t)$ be solutions, in a neighborhood T of 0 in $\mathbf{K}$, of the differential equation 4.5.3.1 with derivative $(c_1, \dots, c_n)$ at 0, and let $\gamma(t) = \varphi(\gamma_1(t), \dots, \gamma_n(t))$.

We shall show that the curve can be extended, in a unique way, to a one parameter subgroup of G.

First we shall show that $\gamma(s+t) = \gamma(s)\gamma(t)$, for s and t in some neighborhood of 0 in $\mathbf{K}$. Since multiplication in G is associative we have an equation $\mu_j(\mu(y, z), z') = \mu_j(y, \mu(z, z'))$. Differentiating the latter equation with respect to z'_j at $z' = 0$ we get

$$\Phi_{ij}\mu(y, z) = \frac{\partial \mu_j}{\partial z'_i}(\mu(y, z), 0) = \sum_{k=1}^{n} \frac{\partial \mu_j}{\partial z_k}(y, z)\frac{\partial \mu_k}{\partial z'_i}(z, 0) = \sum_{k=1}^{n} \Phi_{ik}(z)\frac{\partial \mu_j}{\partial z_k}(y, z). \qquad (4.5.4.1)$$

On the other hand, differentiating $\mu_j(\gamma(s), \gamma(t))$ with respect to t, we obtain

$$\frac{d\mu_j}{dt}(\gamma(s), \gamma(t)) = \sum_{k=1}^{n} \frac{\partial \mu_j}{\partial z_k}(\gamma(s), \gamma(t))\frac{d\gamma_k}{dt}(\gamma(t)) = \sum_{k=1}^{n}\sum_{i=1}^{n} \frac{\partial \mu_j}{\partial z_k}(\gamma(s), \gamma(t))\Psi_{ik}(\gamma(t))c_i.$$

It follows from the latter equation and Equation 4.5.4.1, with $y = \gamma(s)$ and $z = \gamma(t)$, that

$$\frac{d\mu_j}{dt}(\gamma(s), \gamma(t)) = \sum_{i=1}^{n} \Phi_{ij}(\mu(\gamma(s), \gamma(t)))c_i.$$

We also have that

$$\frac{d\gamma_j}{dt}(s+t) = \sum_{i=1}^{n} \Psi_{ij}(\gamma(s+t))c_i,$$

since $\gamma_j(t)$ and thus $\gamma_j(s+t)$ satisfies the differential equation 4.5.3.1 in a neighborhood of 0. Hence we have that $\mu_j(\varphi^{-1}\gamma(s), \varphi^{-1}\gamma(t))$ and $\gamma_j(s+t)$ satisfy the same differential equation 4.5.3.1, and for $t = 0$ we have that $\mu(\varphi^{-1}\gamma(s), \varphi^{-1}\gamma(0)) = \gamma(s)$. It follows from the uniqueness part of Proposition 4.5.1 that we must have that $\gamma(s)\gamma(t) = \gamma(s+t)$, for s and t in some neighborhood S of 0 in $\mathbf{K}$.

We can now extend the curve $\gamma\colon T \to G$ uniquely to a one parameter subgroup $\gamma\colon \mathbf{K} \to G$ of G. First we note that any extension is unique. Indeed, given t in $\mathbf{K}$, then $\frac{1}{m}t$ is in T for some positive integer m. Then we have that $\gamma(t) = \gamma(\frac{n}{n}t) = \gamma(\frac{1}{n}t)^n$, such that $\gamma(\frac{1}{n}t)$

determines $\gamma(t)$. To extend γ to $\mathbf{K}$ we use the same method. Given t in $\mathbf{K}$, we choose a positive integer p such that $\frac{1}{p}t$ is in S. If q is another such integer we obtain that

$$\gamma\left(\frac{1}{p}t\right)^p = \gamma\left(\frac{q}{pq}t\right)^p = \gamma\left(\frac{1}{pq}t\right)pq = \gamma\left(\frac{q}{pq}t\right)q = \gamma\left(\frac{1}{q}t\right)^q,$$

since $p\frac{1}{pq}t$ and $q\frac{1}{pq}t$ both are in T. It follows that we can define uniquely $\gamma(t)$ by $\gamma(t) = \gamma(\frac{1}{p}t)^p$, for any positive integer p such that $\frac{1}{p}t$ is in T. We can thus extend the curve to a curve $\gamma\colon \mathbf{K} \to G$ and the extension is analytic because division by p is analytic in $\mathbf{K}$, and taking p'th power is analytic in G.

Finally, we have that γ is a group homomorphism because, for any s and t in $\mathbf{K}$, we choose a positive integer p such that $\frac{1}{p}s$, $\frac{1}{p}t$ and $\frac{1}{p}(s+t)$ are in T. Then we have that

$$\gamma(s+t) = \gamma\left(\frac{1}{p}(s+t)\right)^p = \gamma\left(\frac{1}{p}s\right)^p \gamma\left(\frac{1}{p}t\right)^p = \gamma(s)\gamma(t).$$

We have proved that γ is a one parameter subgroup of G and that the condition that $\gamma'(0) = (c_1, \dots, c_n)$ in the given coordinates of $T_e(G)$ determines γ uniquely. Thus we have proved the proposition. □

Example 4.5.5. Let G be one of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, or $\mathrm{SG}_S(\mathbf{K})$. We have identified, in 2.5, the tangent space of G with a subspace $\mathfrak{g}$ of $\mathrm{M}_n(\mathbf{K})$. Given D in $\mathfrak{g}$, it follows from assertion (ii) of 2.2.8 and from 2.4.8 and that $\gamma(t) = \exp(tD)$ is an one parameter subgroup of G, and from Example 2.4.15 that the tangent $\gamma'(0)$ is D. Consequently, $\exp(tD)$ is the unique one parameter subgroup of G with tangent D.

4.6 The exponential function for Lie groups

We shall next construct an exponential function for Lie groups, generalizing the exponential function for matrix groups defined in Section 2.2.

4.6.1. For Lie groups there is a Taylor expansion of analytic functions generalizing the usual Taylor expansion in analysis. We shall in this paragraph deduce the expansion on Lie groups from that of analysis.

Let G be a Lie group and X a vector field on G. To X there correspond a unique one parameter subgroup $\gamma\colon \mathbf{K} \to G$ of G with tangent $X(e)$, that is $\gamma'(0) = X(e)$. Choose a chart $\varphi\colon U \to G$ of G such that $\varphi(0) = e$ and choose coordinates $x = (x_1, \dots, x_n)$ in U. In this chart we can write

$$X(e) = T_0\varphi\left(c_1\frac{\partial}{\partial x_1} + \cdots c_n\frac{\partial}{\partial x_n}\right),$$

for some elements $c_1, \dots, c_n$ of $\mathbf{K}$. Let $f : \varphi(U) \to \mathbf{K}$ be an analytic function. For each point x of U it follows from Remark 4.4.6 that

$$Xf(\varphi(x)) = \sum_{i=1}^n \sum_{j=1}^n c_i \Psi_{ij}(x) \frac{\partial(f\varphi)}{\partial x_j}(x), \tag{4.6.1.1}$$

where $\Psi_{ij}(y) = \frac{\partial \mu_j}{\partial z_i}(y,0)$, and where $(\mu_1,\dots,\mu_n)$ represent the multiplication of G in the chart. Write

$$\varphi^{-1}\gamma(t) = (\gamma_1(t),\dots,\gamma_n(t))$$

in a neighborhood of 0 in $\mathbf{K}$. We have that $\gamma_j(s+t) = \mu_j(\varphi^{-1}\gamma(s),\varphi^{-1}\gamma(t))$ for s and t in a neighborhood of 0 in $\mathbf{K}$. Differentiation of the latter expression with respect to t, at 0, gives

$$\frac{d\gamma_j}{dt}(t) = \sum_{i=1}^{n}\frac{\partial \mu_j}{\partial z_i}(\varphi^{-1}\gamma(x),0)\frac{d\gamma_i}{dt}(0) = \sum_{i=1}^{n}\Psi_{ij}(\varphi^{-1}\gamma(t))c_i.$$

Consequently, we have that

$$\frac{d(f\gamma)}{dt}(t) = \sum_{j=1}^{n}\frac{\partial(f\varphi)}{\partial x_j}(\varphi^{-1}\gamma(t))\frac{d\gamma_j}{dt}(t) = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial(f\varphi)}{\partial x_j}(\varphi^{-1}\gamma(t))\Psi_{ij}(\varphi^{-1}\gamma(t))c_i.$$

Comparing the latter formula with Formula 4.6.1.1 we get that

$$\frac{d(f\gamma)}{dt}(t) = Xf(\gamma(t)). \tag{4.6.1.2}$$

We obtain that

$$\frac{d^2(f\gamma)}{dt^2}(t) = \frac{d(Xf)}{dt}(\gamma(t)) = X^2f(\gamma(t)),$$

where the first equality is obtained by differentiation of both sides of Equation 4.6.1.2 and the second by applying Equation 4.6.1.2 to Xf. Iterating we obtain that

$$\frac{d^i(f\gamma)}{dt^i}(t) = X^if(\gamma(t)), \quad \text{for } i = 1,2,\dots.$$

Taylor expansion of the function $f\gamma\colon V \to \mathbf{K}$ in one variable defined in a neighborhood of 0 in $\mathbf{K}$ gives

$$f\gamma(t) = f\gamma(0) + \frac{1}{1!}\frac{d(f\gamma)}{dt}(0) + \frac{1}{2!}\frac{d^2(f\gamma)}{dt^2}(0) + \cdots.$$

We obtain that

$$f\gamma(t) = f(e) + \frac{1}{1!}Xf(e)t + \frac{1}{2!}X^2f(e)t^2 + \cdots = \left(\left((1 + \frac{1}{1!}tX + \frac{1}{2!}t^2X^2 + \cdots\right)f\right)(e),$$

which is the *Taylor expansion* of f on G, and converges in a neighborhood of 0 in $\mathbf{K}$.

Definition 4.6.2. To every left invariant vector field X on a Lie group G we have associated, in Proposition 4.5.4, a unique one parameter subgroup $\gamma\colon \mathbf{K} \to G$ of G. We write $\gamma(t) = \exp(tX)$ and define a map $\exp\colon \mathfrak{g} \to G$ on the space of left invariant vector fields by $\exp(X) = \gamma(1)$. The map exp is called the exponential function of G.

Example 4.6.3. Let G be one of the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, or $\mathrm{SG}_S(\mathbf{K})$. It follows from 4.5.5 that the exponential function maps a vector D in the tangent space $\mathfrak{g}$ of G to $\exp(D)$, where exp is the exponential function of Section 2.2. Hence, in the case of the matrix groups the exponential function, as defined in this section, is the same as the exponential map as defined in 2.2.

Example 4.6.4. Let V be a vector space. We choose a basis $v_1, \dots, v_n$ of V and consider V as a normed space, isomorphic to $\mathbf{K}^n$, via this basis (see 2.1.7). Then V is a Lie group with respect to the addition of V, and the isomorphism $\varphi\colon \mathbf{K}^n \to V$, defined by the basis is a chart. The tangent space of V at 0 is has, via this chart, a basis $\delta_1, \dots, \delta_n$ corresponding to $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$, where $x_1, \dots, x_n$ are coordinates on $\mathbf{K}^n$. Let $D = a_1\delta_1 + \cdots + a_n\delta_n$ The map $\gamma\colon \mathbf{K} \to V$ defined by $\gamma(t) = (ta_1v_1 + \cdots + ta_nv_n)$ is a one parameter subgroup whose derivative at 0 is $a_1v_1 + \cdots + a_nv_n$. Consequently, we have that $\exp(a_1\delta_1 + \cdots + a_n\delta_n) = a_1v_1 + \cdots + a_nv_n$, and we can identify V with its tangent space at 0, via the exponential map.

4.6.5. By the Taylor expansion we obtain, for each analytic function $f\colon U \to \mathbf{K}$, an expression

$$f\exp(tX) = \left((1 + \frac{1}{1!}tX + \frac{1}{2!}t^2X^2 + \cdots)f\right)(e).$$

4.6.6. Choose, as in Paragraph 4.6.1, a chart $\varphi\colon U \to G$ of G and coordinates $x_1, \dots, x_n$ of U. We define a norm on the space $T_e(G)$ via the basis $\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}$ of $T_0(U)$ (see Example 2.1.7). Denote the coordinates of $T_0(U)$ with respect to this basis by $u_1, \dots, u_n$. The space $\mathfrak{g}$ obtains a norm via the isomorphism $\epsilon_G\colon \mathfrak{g} \to T_e(G)$ of Proposition 4.4.7. It follows from Example 2.1.7 that the analytic structure on $\mathfrak{g}$ is independent of the choice of basis. We shall next show that the map $\exp\colon \mathfrak{g} \to G$ is analytic with respect to this analytic structure of $\mathfrak{g}$.

Proposition 4.6.7. *The exponential map defines an analytic map*

$$\exp\colon \mathfrak{g} \to G.$$

Moreover, the map $T_0\exp\colon T_0(\mathfrak{g}) \to T_e(G)$ *is an isomorphism. More precisely, if we identify the tangent space of* $\mathfrak{g}$ *at* 0 *with* $\mathfrak{g}$*, as in Example 4.6.4, we have that the map of left invariant vector fields associated to* exp *is the identity map.*

Proof: We shall use the same notation as in Paragraph 4.6.1. In this notation the vector $T_0\varphi\left(u_1\frac{\partial}{\partial x_1} + \cdots + u_n\frac{\partial}{\partial x_n}\right)$ of $T_e(G)$ corresponds to the left invariant vector field

$$X = \sum_{i=1}^{n}\sum_{j=1}^{n} u_i\Psi_{ij}(x)\frac{\partial}{\partial x_j}.$$

Taylors formula applied to the analytic functions $x_j\varphi^{-1}$ gives

$$\gamma_j(t) = \frac{1}{1!}(Xx_j\varphi^{-1})(e)t + \frac{1}{2!}(X^2x_j\varphi^{-1})(e)t^2 + \cdots.$$

We obtain formulas

$$(Xx_j\varphi^{-1})(\varphi(x)) = \sum_{i=1}^{n} u_i\Psi_{ij}(x), \quad \text{for } j = 1, \dots, n$$

and

$$X^2(x_j\varphi^{-1})(\varphi(x)) = X(Xx_j\varphi^{-1})(\varphi(x))$$
$$= \sum_{i=1}^n \sum_{k=1}^n u_i \Psi_{ik}(x) \frac{\partial(Xx_j\varphi^{-1})}{\partial x_k}(\varphi(x)) = \sum_{i=1}^n \sum_{k=1}^n \sum_{l=1}^n u_i u_l \Psi_{ik}(x) \frac{\partial \Psi_{lj}}{\partial x_k}(x).$$

Iterating these calculations we obtain expressions for $X^i(x_j\varphi^{-1})(\varphi(x))$, and we see that $\gamma_j(t)$ is a power series in $tu_1, \dots, tu_n$ that converges in a neighborhood of 0 in $\mathbf{K}$, for fixed $u_1, \dots, u_n$.

Fix $c = (c_1, \dots, c_n)$ such that the series $\gamma_j(t)$ converges for $|t| \le C_c$ for some positive constant C_c. Let ϵ be the smallest nonzero $|c_i|$, for $i = 1, \dots, n$. We shall show that there is an open neighborhood U_c of c in $T_0(U)$ such that, for all $u \in U_c$, the series $\gamma(t)$ converges for $t \le \frac{1}{2}$. To show this we may, by changing the coordinate system, which does not affect the analytic structure of $T_0(U)$, assume that all the c_i are nonzero. Let $\epsilon = \min_i |c_i|$. Then, for u in $U_c = B(c, \epsilon)$, we have an inequality $|u_i| \le |u_i - c_i| + |c_i| < 2|c_i|$. Consequently, we have that $|tu_i| < |2tc_i|$ and $\gamma(t)$ converges at $2tc$, when $|t| < \frac{1}{2}C_c$. Let $X = \{u \in T_0(U) \mid |u_i| \le 1\}$. Then X is closed and bounded and thus compact by Proposition 3.9.2. The sets U_c for $c \in X$ cover X and we can find a finite subcover $U_{c_1}, \dots, U_{c_m}$. For each $i = 1, \dots, m$ there is a corresponding positive constant C_i such that $\gamma_j(t)$ converges for $u \in U_c$ and for $|t| < C_i$. Let $C = \min_i C_i$. Then we have that $\gamma_j(t)$ converges for all $u \in B(0, \frac{1}{2})$ and all t such that $|t| < C$. Consequently $\gamma_j(t)$ is an analytic function of $u = (u_1, \dots, u_n)$ in some neighborhood of 0 in U. The same argument applied to $\gamma_1, \dots, \gamma_n$ shows that γ is analytic in a neighborhood of 0 in $\mathfrak{g}$.

To prove the last assertion of the Proposition we differentiate $\gamma_j(1)$, with respect to $u_1, \dots, u_n$ at 0. Since $X^i(x_j\varphi^{-1}(\varphi(x)))$ is a polynomial of degree i in $u_1, \dots, u_n$, we have that

$$\frac{\partial \gamma_j}{\partial u_i}(0) = \frac{\partial(Xx_j\varphi^{-1})}{\partial u_i}(0) = \left(\frac{\partial}{\partial u_i} \sum_{l=1}^n u_l \Psi_{lj}\right)(e) = u_i \Psi_{ij}(0).$$

However, we have that $\Psi_{ij}(0) = \frac{\partial \mu_j}{\partial w_j}(0,0)$, where the w_j are the variables corresponding to the second coordinate of the μ_j, and the maps $\mu_1(0,w), \dots, \mu_n(0,w)$ correspond to multiplication by e and is thus the identity map. Consequently, we have that $(\Psi_{ij}(0))$ is the $n \times n$ identity matrix. We obtain that

$$\frac{\partial \gamma_j}{\partial u_i}(0) = I_n,$$

as we wanted to prove. □

5 Algebraic varieties

In this chapter we shall show that there is a beautiful geometric theory for matrix groups over an arbitrary field, that is similar to the one for analytic manifolds presented in Chapter 3. To compensate for the lack of a norm on the fields, and the ensuing lack of exponential function, the inverse function theorem, and other techniques depending on the access to analytic functions, we shall make extensive use of the machinery of commutative algebra.

5.1 Affine varieties

We saw in Section 3.2 that the matrix groups are the zeroes of polynomials in some space $\mathrm{M}_n(\mathbf{K})$. The central objects of study of algebraic geometry are the zeroes of polynomials. It is therefore natural to consider the matrix groups from the point of view of algebraic geometry. In this section we shall introduce algebraic sets that form the underlying geometric objects of the theory.

5.1.1. We fix a field $\mathbf{K}$, and an inclusion $\mathbf{K} \subset \overline{\mathbf{K}}$ into an algebraically closed field $\overline{\mathbf{K}}$, that is, a field such that every polynomial $a_m x^m + a_{m-1}x^{m-1} + \cdots + a_0$ in a variable x with coefficients in $\overline{\mathbf{K}}$ has a zero in $\overline{\mathbf{K}}$ (see Exercise 5.1.3).

Remark 5.1.2. The reason why we introduce a second field is that we want to assure that all polynomials have zeroes. For example the polynomial x^2+1 does not have a zero in the real numbers $\mathbf{R}$, but it has zeroes in the algebraically closed field of complex numbers $\mathbf{C}$, containing $\mathbf{R}$. The question of zeroes of analytic function never came up in the analytic theory of Chapter 3, where the underlying sets are manifolds, and thus locally look like an open subset of $\mathbf{K}^n$. Given a field $\mathbf{K}$ we can always find an *algebraically closed* field containing $\mathbf{K}$ (see Exercises 5.1.8).

Definition 5.1.3. We denote the polynomial ring in the variables $x_1, \dots, x_n$ with coefficients in $\mathbf{K}$ by $\mathbf{K}[x_1, \dots, x_n]$. The cartesian product $\overline{\mathbf{K}}^n$ we denote by $\mathbf{A}^n_{\overline{\mathbf{K}}}$, and we call $\mathbf{A}^n_{\overline{\mathbf{K}}}$ the n dimensional *affine space* over $\overline{\mathbf{K}}$, or simply the affine n space.

We say that a subset X of $\mathbf{A}^n_{\overline{\mathbf{K}}}$ is an *affine variety* if there exists an ideal I in $\mathbf{K}[x_1, \dots, x_n]$, such that X is the set of common zeroes of the polynomials f of I. That is

$$X = \mathcal{V}(I) = \{(a_1, \dots, a_n) \in \mathbf{A}^n_{\overline{\mathbf{K}}} \quad : \quad f(a_1, \dots, a_n) = 0 \quad \text{for all} \quad f \in I\}.$$

We do not need all the polynomials in an ideal to define an affine variety. It suffices to consider certain families of polynomials that generate the ideal in a sense that we shall explain next. Later, as a consequence of Corollary 5.1.17, we shall see that it suffices to consider a finite number of polynomials.

Definition 5.1.4. Let R be a ring and I and ideal in R. A subset $\{a_i\}_{i\in\mathcal{I}}$ is called a *set of generators* for I if I is the smallest ideal containing the elements a_i, for $i \in \mathcal{I}$. Equivalently, I is generated by the set $\{a_i\}_{i\in\mathcal{I}}$ if I consists of the sums $b_1 a_{i_1} + \cdots + b_m a_{i_m}$, for all finite subsets $\{i_1, \dots, i_m\}$ of $\mathcal{I}$, and elements $b_1, \dots, b_m$ in R (see Exercise 5.1.7).

Remark 5.1.5. Let I be an ideal in $\mathbf{K}[x_1, \dots, x_n]$ generated by polynomials $\{f_i\}_{i\in\mathcal{I}}$. Then $X = \mathcal{V}(I)$ is the common zeroes $\mathcal{V}(\{f_i\}_{i\in\mathcal{I}})$ of the polynomials f_i, for $i \in \mathcal{I}$. Indeed, if a point x is a common zero for the polynomials in I, then it is a zero for the polynomials f_i. Conversely, if x is a common zero for the polynomials f_i, for all $i \in \mathcal{I}$, then x is a zero for all polynomials in I, because all polynomials in I are of the form $g_1 f_{i_1} + \cdots + g_m f_{i_m}$, for some indices $i_1, \dots, i_m$ of $\mathcal{I}$, and polynomials $g_1, \dots, g_m$ of $\mathbf{K}[x_1, \dots, x_n]$.

Example 5.1.6. As we remarked in Section 3.2 the set $\mathrm{Sl}_n(\mathbf{K})$ is the affine variety of $\mathrm{M}_n(\mathbf{K}) = \mathbf{A}^{n^2}_{\mathbf{K}}$ where the polynomial $\det(x_{ij})-1$ is zero. When S is invertible the set $\mathrm{G}_S(\mathbf{K})$ is the zeroes of the n^2 quadratic equations in the variables x_{ij} obtained by equating the n^2 coordinates on both sides of $(x_{ij})S^t(x_{ij}) = S$. Finally, $\mathrm{SG}_S(\mathbf{K})$ is the subset of $\mathrm{Gl}_n(\mathbf{K})$ which is the intersection of $\mathrm{G}_S(\mathbf{K})$ with the matrices where the polynomial $\det(x_{ij}) - 1$ vanishes. On the other hand we have that $\mathrm{Gl}_n(\mathbf{K})$ itself can be considered as the zeroes of polynomials in the affine space $\mathbf{A}^{(n+1)^2}_{\mathbf{K}}$. Indeed, we saw in Example 1.2.11 that we have an injection $\varphi\colon \mathrm{Gl}_n(\mathbf{K}) \to \mathrm{Sl}_{n+1}(\mathbf{K})$. As we just saw $\mathrm{Sl}_{n+1}(\mathbf{K})$ is the zeroes of a polynomial of degree $n+1$ in the variables x_{ij}, for $i, j = 1, \dots, n+1$, and clearly $\operatorname{im}\varphi$ is given in $\mathrm{Sl}_{n+1}(\mathbf{K})$ by the relations $x_{1i} = x_{i1} = 0$, for $i = 2, \dots, n+1$. Hence all the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$ or $\mathrm{G}_S(\mathbf{K})$, for some invertible matrix S, are affine varieties.

Example 5.1.7. Let X and Y be affine varieties in $\mathbf{A}^n_{\mathbf{K}}$ respectively $\mathbf{A}^m_{\mathbf{K}}$. Then the subset $X \times Y$ is an affine variety in $\mathbf{A}^n_{\mathbf{K}} \times \mathbf{A}^m_{\mathbf{K}}$. Indeed, let I and J, be ideals in the polynomial rings $\mathbf{K}[x_1, \dots, x_n]$ respectively $\mathbf{K}[y_1, \dots, y_m]$ such that $X = \mathcal{V}(I)$ respectively $Y = \mathcal{V}(J)$ in $\mathbf{A}^n_{\mathbf{K}}$ respectively $\mathbf{A}^m_{\mathbf{K}}$. Then $X \times Y$ is the affine variety in $\mathbf{A}^{m+n}_{\mathbf{K}} = \mathbf{A}^n_{\mathbf{K}} \times \mathbf{A}^m_{\mathbf{K}}$ defined by the smallest ideal in $\mathbf{K}[x_1, \dots, x_n, y_1, \dots, y_m]$ containing I and J. This ideal consists of all polynomials of the form $af + bg$, where f and g are in I respectively J, and a and b are in $\mathbf{K}[x_1, \dots, x_n, y_1, \dots, y_m]$. Indeed, it is clear that all the polynomials of this form are zero on $X \times Y$. Conversely, if $(a_1, \dots, a_n, b_1, \dots, b_m)$ in $\mathbf{A}^{m+n}_{\mathbf{K}}$ is not in $X \times Y$, then there is a polynomial f in I, or a polynomial g in J, such that $f(a_1, \dots, a_n) \neq 0$ or $g(b_1, \dots, b_m) \neq 0$. Consequently, the point $(a_1, \dots, a_n, b_1, \dots, b_m)$ is not in the common zeroes of all the polynomials of the form $af + bg$.

Lemma 5.1.8. *The affine varieties in $\mathbf{A}^n_{\mathbf{K}}$ have the following three properties:*

(i) *The empty set and $\mathbf{A}^n_{\mathbf{K}}$ are affine varieties.*

(ii) *Let $\{X_i\}_{i \in I}$ be a family of affine varieties. Then the intersection $\cap_{i \in \mathcal{I}} X_i$ is an affine variety.*

(iii) *Let $X_1, \dots, X_m$ be a finite family of affine varieties. Then the union $X_1 \cup \cdots \cup X_m$ is an affine variety.*

Proof: To prove the first assertion is suffices to observe that the common zeroes of the polynomials 1 and 0 are $\emptyset$ respectively $\mathbf{A}^n_{\mathbf{K}}$.

Let $X_i = \mathcal{V}(I_i)$, for $i \in \mathcal{I}$, where I_i is an ideal of $\mathbf{K}[x_1, \dots, x_n]$. Moreover, let I be the smallest ideal in $\mathbf{K}[x_1, \dots, x_n]$ that contains all the ideals I_i. That is, I is the ideal generated by all the polynomials in the ideals I_i for $i \in \mathcal{I}$, and hence consists of all sums $f_{i_1} + \cdots + f_{i_m}$, for all finite subsets $\{i_1, \dots, i_m\}$ of $\mathcal{I}$, and with $f_{i_j} \in I_{i_j}$. It is clear that $\mathcal{V}(I) = \cap_{i \in \mathcal{I}} \mathcal{V}(I_i) = \cap_{i \in \mathcal{I}} X_i$. We have proved the second assertion.

To prove the third assertion we let $X_i = \mathcal{V}(I_i)$ for some ideal I_i in $\mathbf{K}[x_1, \dots, x_n]$. Let I be the ideal in $\mathbf{K}[x_1, \dots, x_n]$ generated by the elements in the set

$$\{f_1 \cdots f_m \quad | \quad f_i \in I_i, \quad \text{for } i = 1, \dots, m\}.$$

We have an inclusion $\cup_{i=1}^m X_i = \cup_{i=1}^m \mathcal{V}(I_i) \subseteq \mathcal{V}(I)$. To prove the opposite inclusion we take a point x in $\mathbf{A}^n_{\overline{\mathbf{K}}} \setminus \cup_{i=1}^m \mathcal{V}(I_i)$ Then there exists, for $i = 1, \dots, m$ a polynomial $f_i \in I_i$ such that $f_i(x) \neq 0$. We have that $(f_1 \cdots f_m)(x) = f_1(x) \cdots f_m(x) \neq 0$, and thus $x \notin \mathcal{V}(I)$. Hence we that $\mathcal{V}(I) \subseteq \cup_{i=1}^m X_i$, and we have proved the third assertion of the lemma. □

Remark 5.1.9. The properties of Lemma 5.1.8 can be interpreted as stating that the affine varieties of $\mathbf{A}^n_{\overline{\mathbf{K}}}$ form the closed sets of a topology (see Definition 3.3.1).

Definition 5.1.10. The topology on $\mathbf{A}^n_{\overline{\mathbf{K}}}$ whose open sets are the complements of the affine varieties is called the *Zariski topology*. For each subset X of $\mathbf{A}^n_{\overline{\mathbf{K}}}$ the topology induced on X is called the *Zariski topology on* X.

When X is an affine variety in $\mathbf{A}^n_{\overline{\mathbf{K}}}$ we call the open subsets of X in the Zariski topology, *quasi affine varieties.*

Example 5.1.11. The closed sets for the Zariski topology on $\mathbf{A}^1_{\overline{\mathbf{K}}} = \overline{\mathbf{K}}$ consists of the common zeroes of polynomials in one variable with coefficients in the field $\mathbf{K}$. In the ring $\mathbf{K}[x]$ all ideals can be generated by one element (see Exercise 5.1.2) In particular, every closed set different from $\overline{\mathbf{K}}$ is finite, and consists of the zeroes of one polynomial in $\mathbf{K}[x]$.

Take $\mathbf{K}$ and $\overline{\mathbf{K}}$ to be $\mathbf{R}$ respectively $\mathbf{C}$. Then i is not a closed subset of $\mathbf{C}$ because every polynomial in $\mathbf{R}[x]$ that has i as a root also has $-i$ as a root. However, the set $\{i, -i\}$ is closed in $\mathbf{C}$, beeing the zeroes of the polynomial $x^2 + 1$.

Example 5.1.12. In the Zariski topology on $\mathbf{A}^{m+n}_{\overline{\mathbf{K}}} = \mathbf{A}^m_{\overline{\mathbf{K}}} \times \mathbf{A}^n_{\overline{\mathbf{K}}}$ the sets of the form $U \times V$, where U and V are open in $\mathbf{A}^m_{\overline{\mathbf{K}}}$ respectively $\mathbf{A}^n_{\overline{\mathbf{K}}}$ are open in $\mathbf{A}^m_{\overline{\mathbf{K}}} \times \mathbf{A}^n_{\overline{\mathbf{K}}}$. This follows immediately from Example 5.1.7

The Zariski topology on the product $\mathbf{A}^m_{\overline{\mathbf{K}}} \times \mathbf{A}^n_{\overline{\mathbf{K}}}$ is however, not the product topology of the topologies on $\mathbf{A}^m_{\overline{\mathbf{K}}}$ and $\mathbf{A}^n_{\overline{\mathbf{K}}}$ as defined in Example 3.3.3. Indeed, for example when $m = n = 1$ the diagonal of $\mathbf{A}^1_{\overline{\mathbf{K}}} \times \mathbf{A}^1_{\overline{\mathbf{K}}}$ is closed. However, the closed sets in $\mathbf{A}^1_{\overline{\mathbf{K}}}$ are finite, or the whole space (see Example 5.1.11). Hence the sets $X \times Y$ with X closed in $\mathbf{A}^m_{\overline{\mathbf{K}}}$ and Y is closed in $\mathbf{A}^n_{\overline{\mathbf{K}}}$, can not be the diagonal.

Some open sets, often called *principal*, are particularly important for the Zariski topology.

Definition 5.1.13. Let f be a polynomial in $\mathbf{K}[x_1, \dots, x_n]$. The set

$$\mathcal{V}(f) = \{x \in \mathbf{A}^n_{\overline{\mathbf{K}}} \quad | \quad f(x) = 0\},$$

where f vanishes, is a closed subset of $\mathbf{A}^n_{\overline{\mathbf{K}}}$. For each affine variety X of $\mathbf{A}^n_{\overline{\mathbf{K}}}$ we denote by X_f the open subset $X \setminus \mathcal{V}(f)$ of X. We call the subsets of the form X_f, the *principal* open subsets.

Lemma 5.1.14. *Let X be an affine variety in $\mathbf{A}^n_{\overline{\mathbf{K}}}$ and let U be an open subset of X. For each point x of U there is a polynomial f in $\mathbf{K}[x_1, \dots, x_n]$ such that $x \in X_f$, and $X_f \subseteq U$.*

Proof: We have that $X \setminus U$ is a closed subset of $\mathbf{A}^n_{\mathbf{K}}$. Consequently there is an ideal I of $\mathbf{K}[x_1, \dots, x_n]$ such that $X \setminus U = \mathcal{V}(I)$. Since $x \in U$ there is a polynomial f in I such that $f(x) \neq 0$. Then $x \in X_f$, by the definition of X_f. Since $f \in I$, we have that $\mathcal{V}(I) \subseteq \mathcal{V}(f)$. Hence $X \setminus U \subseteq X \cap \mathcal{V}(f)$, or equivalently, $X_f \subseteq U$ □

It is interesting, and useful, to notice that every affine variety is the common zeroes of a finite number of polynomials. Before we prove this important fact we shall introduce some terminology.

Definition 5.1.15. We say that a ring R is *noetherian* if every ideal in R is *finitely generated*. That is, let I be an ideal in R, then there is a finite set of elements $a_1, \dots, a_m$ such that I is the smallest ideal of R containing $a_1, \dots, a_m$. Equivalently, the elements of I consists of all elements of the form $a_1 b_1 + \cdots + a_m b_m$, for all elements $b_1, \dots, b_m$ of R.

Proposition 5.1.16. *Let R be a noetherian ring. Then the ring $R[x]$ of polynomials in the variable x with coefficients in R is also noetherian.*

Proof: Let J be an ideal of $R[x]$, and let I be the subset of R consisting of all elements a in R, such that $ax^m + a_{m-1}x^{m-1} + \cdots + a_0$ is in J for some nonnegative integer m and some elements $a_0, \dots, a_{m-1}$ in R. We have that I is an ideal of R. Indeed, if $a \in I$, then $ba \in I$, for all $b \in R$, because there is a polynomial $f(x) = ax^p + a_{p-1}x^{p-1} + \cdots + a_0$ in J, and then $bf(x) = bax^p + ba_{p-1}x^{p-1} + \cdots + ba_0$ is in J. Moreover, if b is in I then $a + b$ is in I because some $g(x) = bx^p + b_{p-1}x^{p-1} + \cdots + b_0$ is in J. Assume that $q \geq p$. Then $f(x) - x^{q-p}g(x) = (a+b)x^q + (a_{p-1}+b_{q-1})x^{q-1} + \cdots + (a_0+b_{q-p})x^{q-p} + b_{q-p-1}x^{q-p-1} + \cdots + b_0$ is in J, and thus $a + b \in I$. A similar argument shows that $a + b$ is in I when $p \leq q$.

In a similar way we show that the set I_i consisting of the coefficient of x^i of all polynomials of J of degree at most i, is an ideal.

We have that R is noetherian, by assumption, so all the ideals I, and I_i are finitely generated. Choose generators $a_1, \dots, a_m$ for I, and $b_{i1}, \dots, b_{im_i}$ for I_i, for $i = 0, 1, \dots$. Moreover, we choose polynomials in $R[x]$ whose highest nonzero coefficient is $a_1, \dots, a_m$, respectively. Multiplying with an appropriate power of x we can assume that all these polynomials have the same degree p. Hence we can choose polynomials $f_i(x) = a_i x^p + a_{i(p-1)}x^{p-1} + \cdots + a_{i0}$, for $i = 1, \dots, m$. Moreover, we choose polynomials $f_{ij}(x) = b_{ij}x^i + b_{ij(i-1)}x^{i-1} + \cdots + b_{ij0}$ in J, for $i = 0, 1, \dots$ and $j = 1, \dots, m_i$.

We shall show that the polynomials $S = \{f_1, \dots, f_m\} \cup \{f_{i_1}, \dots, f_{im_i} : i = 0, \dots, p-1\}$, generate J. It is clear that all polynomials in J of degree 0 is in the ideal generated by the polynomials in S. We proceed by induction on the degree of the polynomials of J. Assume that all polynomials of degree strictly less than p in J lie in the ideal generated by the elements of S. Let $f(x) = bx^q + b_{q-1}x^{q-1} + \cdots + b_0$ be in J. Assume that $q \geq p$. We have that $b \in I$. Hence $b = c_1 a_1 + \cdots + c_m a_m$, for some $c_1, \dots, c_m$ in R. Then $g(x) = f(x) - c_1 x^{q-p} f_1(x) - \cdots - c_m x^{q-p} f_m(x)$ is in J and is of degree strictly less than q. Hence $g(x)$ is in the ideal generated by the elements of S by the induction assumption. Since $f(x) = c_1 f_1(x) + \cdots + c_m f_m(x) + g(x)$, we have proved that all polynomials of degree at least equal to p are in the ideal generated by the elements in S.

When $q < p$ we reason in a similar way, using $b_{q1}, \dots, b_{qm_q}$ and $f_{q1}, \dots, f_{qm_q}$, to write $f(x)$ as a sum of an element in J of degree strictly less than q and an element that is in the ideal generated by the elements $\{f_{i_j}\}$. By induction we obtain, in this case, that all polynomials in J of degree less than p are in the ideal generated by S, and we have proved the proposition. □

Corollary 5.1.17. (The Hilbert basis theorem) *The ring* $\mathbf{K}[x_1, \dots, x_n]$ *is noetherian.*

Proof: The field $\mathbf{K}$ has only the ideals (0) and $\mathbf{K}$ (see Exercise 1.3.2), and consequently is noetherian. It follows from the Proposition, by induction on n, that $\mathbf{K}[x_1, \dots, x_n]$ is noetherian. □

5.1.18. The Zariski topology is different, in many important respects, from metric topologies. We shall next show that the quasi affine varieties are compact and that they have a unique decomposition into particular, irreducible, closed sets.

Proposition 5.1.19. *All quasi affine varieties are compact topological spaces.*

Proof: Let X be an algebraic subset of $\mathbf{A}^n_{\mathbf{K}}$, and let U be an open subset of X. Moreover choose an open subset V of $\mathbf{A}^n_{\mathbf{K}}$ such that $V \cap X = U$. Let $U = \cup_{i \in \mathcal{I}} U_i$ be a covering of U by open subsets U_i. Choose open subsets V_i of $\mathbf{A}^n_{\mathbf{K}}$ such that $U_i = X \cap V_i$. Then $V = (V \setminus X) \cup \cup_{i \in \mathcal{I}}(V_i \cap V)$ is a covering of V by open sets of $\mathbf{A}^n_{\mathbf{K}}$. If we can find a finite subcover $V_{i_1} \cap V, \dots, V_{i_m} \cap V, V \setminus X$, then $U_{i_1}, \dots, U_{i_m}$ is an open cover of U. Consequently it suffices to show that every open subset V of $\mathbf{A}^n_{\mathbf{K}}$ is compact.

It follows from Lemma 5.1.14 that it suffices to prove that every covering $V = \cup_{i \in \mathcal{I}} (\mathbf{A}^n_{\mathbf{K}})_{f_i}$ of V by principal open sets $(\mathbf{A}^n_{\mathbf{K}})_{f_i}$ has a finite subcover. Let I be the ideal generated by the the elements f_i, for $i \in \mathcal{I}$. It follows from Corollary 5.1.17 that I is generated by a finite number of elements $g_1, \dots, g_m$. Since $(\mathbf{A}^n_{\mathbf{K}})_{f_i} \subseteq V$, we have that $\mathcal{V}(f_i) \supseteq \mathbf{A}^n_{\mathbf{K}} \setminus V$, for all $i \in \mathcal{I}$. Consequently, we have, for all $g \in I$, that $\mathcal{V}(g) \supseteq \mathbf{A}^n_{\mathbf{K}} \setminus V$, that is $(\mathbf{A}^n_{\mathbf{K}})_g \subseteq V$. Moreover, since the $(\mathbf{A}^n_{\mathbf{K}})_{f_i}$ cover V, we have that for each $x \in V$ there is an f_i such that $f_i(x) \neq 0$. Consequently, there is a g_j such that $g_j(x) \neq 0$. Hence we have that the sets $(\mathbf{A}^n_{\mathbf{K}})_{g_i}$, for $i = 1, \dots, m$, cover V, and we have proved the proposition. □

Corollary 5.1.20. *Every sequence* $U \supseteq X_1 \supseteq X_2 \supseteq \cdots$ *of closed subsets of a quasi affine variety* U *is stationary, that is, for some positive integer* m *we have that* $X_m = X_{m+1} = \cdots$.

Proof: Let $X = \cap_{i=1}^{\infty} X_i$. Then X is a closed subset of U and $V = U \setminus X$ is a quasi affine variety. Let $U_i = U \setminus X_i$. Then we have a covering $V = \cup_{i=1}^{n} U_i$ of V by open sets U_i, where $U_1 \subseteq U_2 \subseteq \cdots \subseteq V$. By Proposition 5.1.19 we can find a finite subcover $U_{i_1}, \dots, U_{i_r}$. Let $m = \max\{i_1, \dots, i_r\}$. Then $U_m = U_{m+1} = \cdots = V$, and we have that $X_m = X_{m+1} = \cdots = X$. □

Definition 5.1.21. A topological space X is called *noetherian* if every sequence of closed subspaces $X \supseteq X_1 \supseteq X_2 \supseteq \cdots$ is *stationary*, that is, for some positive integer m we have that $X_m = X_{m+1} = \cdots$.

We say that a topological space X is *irreducible* if it can not be written as a union of two proper closed subsets.

Remark 5.1.22. A topological space is noetherian if and only if every family $\{X_i\}_{i\in\mathcal{I}}$ of closed sets has a *minimal element*, that is, an element that is not properly contained in any other member of the family. Indeed, if every family has a minimal element, a chain $X \supseteq X_1 \supseteq X_2 \supseteq \cdots$ has a minimal element X_m. Then $X_m = X_{m+1} = \cdots$. Conversely, if X is noetherian and $\{X_i\}_{i\in\mathcal{I}}$ is a family of closed sets, then we can construct, by induction on m, a sequence of sets $X_{i_1} \supset X_{i_2} \supset \cdots \supset X_{i_m}$, by taking X_{i_1} arbitrary, and, if X_{i_m} is not minimal, choosing $X_{i_{m+1}}$ to be a proper subset contained in the family. Since the space is assumed to be noetherian we must end up with a minimal element of the family.

Remark 5.1.23. A space X is clearly irreducible if and only if two nonempty open sets of X always intersect. Consequently, if X is irreducible, then all open subsets of X are irreducible.

Lemma 5.1.24. *Let X be a noetherian topological space. Then we can write X as a union $X = X_1 \cup \cdots \cup X_m$, where $X_1, \dots, X_m$ are irreducible closed subsets of X, and no two of these sets are contained in each other. The sets $X_1, \dots, X_m$ are unique, up to order.*

Proof: We shall show that the family $\{Y_i\}_{i\in I}$ of closed subsets of X for which the first part of the lemma does not hold is empty. If not, it has a minimal element Y_j since X is noetherian. Then Y_j can not be irreducible and hence must be the union of two proper closed subsets. Each of these can, by the minimality of Y_j, be written as a finite union of irreducible closed subsets, and hence, so can Y_j, which is impossible. Hence the family must be empty, and hence X can be written as a finite union of closed irreducible subsets. Cancelling the biggest of the sets when two of the ireeducible sets are contained in each other we arrive at a decomposition of the type described in the first part of the lemma.

We shall show that the decomposition is unique. Assume that we have two decompositions $X = X_1 \cup \cdots \cup X_p = Y_1 \cup \cdots \cup Y_q$. Then $X_i = (X_i \cap Y_1) \cup \cdots \cup (X_i \cap Y_q)$. Since X_i is irreducible we have that, either the intersection $X_i \cap Y_j$ is equal to X_i, or it is empty. At least one of the intersections must be equal to X_i. Then we have that $X_i \subseteq Y_{\sigma(i)}$, for some index $\sigma(i)$. Reasoning in a similar way for $Y_{\sigma(i)}$, we obtain that $Y_{\sigma(i)} \subseteq X_j$, for some index j. But then we have that $X_i \subseteq Y_{\sigma(i)} \subseteq X_k$. Consequently $i = k$ and $X_i = Y_{\sigma(i)}$. Since the latter relation must hold for all i, the second part of the lemma has been proved. □

Definition 5.1.25. Let R be a ring. An ideal I of R is *prime* if, given two elements a and b of R, not in I, then the product ab is not in I.

Proposition 5.1.26. *Let X be an affine variety in $\mathbf{A}^n_{\mathbf{K}}$. Then X is irreducible if and only if the ideal*

$$\mathcal{I}(X) = \{f \in \mathbf{K}[x_1, \dots, x_n] : f(a_1, \dots, a_n) = 0, \quad \textit{for all } (a_1, \dots, a_n) \in X\}$$

of polynomials vanishing on X is a prime ideal in $\mathbf{K}[x_1, \dots, x_n]$.

Proof: It follows from Lemma 5.1.14 that it suffices to prove that any two principal open subsets intersect. Note that for f in $\mathbf{K}[x_1,\dots,x_n]$ we have that $X_f \cap X \neq \emptyset$ if and only if f is not in $\mathcal{I}(X)$, because, for $(a_1,\dots,a_n)$ in $X_f \cap X$, we have that $f(a_1,\dots,a_n) \neq 0$ and $g(a_1,\dots,a_n) = 0$, for all g in $\mathcal{I}(X)$. Let f and g in $\mathbf{K}[x_1,\dots,x_n]$ be polynomials not in $\mathcal{I}(X)$, we have that fg is not in $\mathcal{I}(X)$ if and only if $X_{fg} \cap X \neq \emptyset$. Clearly, we have that $X_{fg} = X_f \cap X_g$, so $X_{fg} \cap X \neq \emptyset$ if and only if $(X_f \cap X) \cap (X_g \cap X) \neq \emptyset$. Consequently, we have that fg is not in $\mathcal{I}(X)$ if and only if $(X_f \cap X)$ and $(X_g \cap X)$ meet. We have proved the proposition. □

Example 5.1.27. Since $\overline{\mathbf{K}}$ is infinte (see Excercise 5.1.5) the only polynomial that vanishes on $\mathbf{A}^n_{\overline{\mathbf{K}}}$ is the zero polynomial. Consequently, we have that $\mathcal{I}(\mathbf{A}^n_{\overline{\mathbf{K}}}) = (0)$ in $\mathbf{K}[x_1,\dots,x_n]$, and $\mathbf{A}^n_{\overline{\mathbf{K}}}$ is irreducible.

Remark 5.1.28. The above results illustrate the difference between the Zariski topology and the metric topology on $\mathbf{A}^n_{\mathbf{R}}$ and $\mathbf{A}^n_{\mathbf{C}}$. In the Zariski topology all open sets are compact and two open subsets always meet. In the metric topology, no open sets are compact (see Proposition 3.9.2), and the space is Hausdorff (see Exercise 3.3.1), so two distinct points always have open neighbourhoods that do not intersect.

Exercises

5.1.1. Let $\mathbf{K}[x]$ be the ring of polynomials in the variable x with coefficients in $\mathbf{K}$. Let $f(x)$ and $g(x)$ be polynomials in $\mathbf{K}[x]$. Show that there are polynomials $q(x)$ and $r(x)$ with $\deg r(x) < \deg f(x)$, such that $g(x) = q(x)f(x) + r(x)$.

Hint: Keep the polynomial $f(x)$ fixed and use induction on the degree of $g(x)$.

5.1.2. Show that every ideal I in the ring $\mathbf{K}[x]$ of polynomials in the variable x with coefficients in $\mathbf{K}$ can be generated by one element.

Hint: Use Exercise 5.1.1 to prove that every polynomial of lowest degree in I will generate I.

5.1.3. Let $\mathbf{K}[x]$ be the ring of polynomials in the variable x with coefficients in $\mathbf{K}$. Moreover let $f(x)$ be a polynomial. Show that an element a of $\mathbf{K}$ is a *root* of $f(x)$, that is $f(a) = 0$, if and only if $f(x) = (x - a)g(x)$.

5.1.4. Let $\mathbf{K}[x]$ be the ring of polynomials in one variable x with coefficients in $\mathbf{K}$. We say that a polynomial $f(x)$ *divides* a polynomial $g(x)$ if there is a polynomial $q(x)$ such that $g(x) = q(x)f(x)$. A polynomial is *irreducible* if it can not be divided by a nonconstant polynomial of strictly lower degree than itself. Two polynomials are *relatively prime* if they have no common divisors except the *constants*, that is, the elements of $\mathbf{K}$.

(a) Use Exercise 5.1.2 to show that if $f(x)$ and $g(x)$ are relatively prime polynomials in $\mathbf{K}[x]$, then $\mathbf{K}[x]$ is the smallest ideal that contains $f(x)$ and $g(x)$.
(b) Show that if $f(x)$ and $g(x)$ are polynomials, and $f(x)$ is irreducible, then, either $f(x)$ and $g(x)$ are relatively prime, or $f(x)$ divides $g(x)$.
(c) Let $f(x)$, $g(x)$, and $h(x)$ be polynomials in $\mathbf{K}[x]$, with $f(x)$ irreducible. Use assertion (a) and (b) to show that, if $f(x)$ divides $g(x)h(x)$, but does not divide $g(x)$, then $f(x)$ divides $h(x)$.

(d) Show that every polynomial $f(x)$ can be written as a product $f(x) = f_1(x)\cdots f_m(x)$, where the polynomials $f_i(x)$ are irreducible, not necessarily disinct, and use (c) to show that the $f_i(x)$ are unique, up to order and multiplication with a constant.

(e) Show that there are infinitely many irreducible polynomials in $\mathbf{K}[x]$ that can not be obtained from each other by multiplication by elements of $\mathbf{K}$.

Hint: Assume that there is a finite number of irreducible polynomials $f_1(x), \dots, f_m(x)$ up to multiplication by constants. Show that each irreducible factor of $(f_1(x)\cdots f_m(x))+1$ is relatively prime to $f_1(x), \dots, f_m(x)$.

5.1.5. Let $\mathbf{K}$ be a field.

(a) Show that a $\mathbf{K}$ is algebraically closed if and only if all irreducible polynomials (see Exercise 5.1.4) in one variable x with coefficients in $\mathbf{K}$ are of degree 1.

(b) Use Exercise 5.1.4 (e) to show that an algebraically closed field has infinitely many elements.

5.1.6. Let $\mathbf{K}[x]$ be the ring of polynomials in one variable x with coefficients in $\mathbf{K}$. Let $f(x)$ be a polynomial and $I = (f(x))$ the smallest ideal in $\mathbf{K}[x]$ containing $f(x)$. Show that the residue ring $\mathbf{K}[x]/I$ is a field, if and only if $f(x)$ is irreducible.

Hint: Use Exercise 5.1.4 (a).

5.1.7. Let R be a ring and I and ideal in R. Show that a subset $\{a_i\}_{i\in\mathcal{I}}$ of R generates I if and only if I consists of the sums $b_1a_{i_1} + \cdots + b_ma_{i_m}$, for all finite subsets $\{i_1, \dots, i_m\}$ of $\mathcal{I}$, and elements $b_1, \dots, b_m$ in R.

5.1.8. Show that every field has an algebraic extension which is algebraically closed.

Hint: Construct a field E_1 that contains F and where all polynomials in $F[x]$ of degree a least 1 have a root. In order to accomplish this choose for every polynomial $f \in F[x]$ of degree at least 1 a symbo X_f, and let $F[S]$ be the polynomial ring having these symbols as variables. The ideal $\mathfrak{I}$ in $F[S]$ generated by the elements $f(X_f)$ is not the whole ring, because then we would have that $g_1f_1(X_{f_1}) + \cdots + g_nf_n(X_{f_n}) = 1$ for some elements $g_1, \dots, g_n$ in $F[S]$. However, these equations involve only a finite number of variables X_f, and we can always find an extension E of F with elements $\alpha_1, \dots, \alpha_n$ such that $f_1(\alpha_1) = \cdots = f_n(\alpha_n) = 0$. Let $\varphi\colon F[X] \to E$ be the map sending the variable X_{f_i} to α_i for $i = 1, \dots, n$ and sending the remaining variables to 0. We get $1 = \varphi(g_1f_1(X_{f_1}) + \cdots + g_nf_n(X_{f_n})) = \varphi(g_1)f_1(\alpha_1) + \cdots + \varphi(g_n)f_n(\alpha_n) = 0$, which is impossible. Hence $\mathfrak{I}$ is not the whole ring.

Let $\mathcal{M}$ be a maximal ideal that contain $\mathfrak{I}$ and let $E_1' = F[X]/\mathcal{M}$. The field F maps injectively into the field E_1'. Replacing E_1' by the set $E_1 = F \cup (E_1 \setminus \operatorname{im} F)$ and giving the latter set the field structure given by the canonical bijection of sets $E_1' \to E_1$ we have that E_1 is a field containing the field F. Every polynomial f in $F[x]$ has a root in E_1 since $f(X_f)$ maps to zero in $F[x]/\mathcal{M}$.

Define inductively $E_1 \subset E_2 \subset \cdots$ such that every polynomial of degree at least 1 in $E_n[x]$ has a root in E_{n+1}, and let E be the union of these fields. Then E is clearly a field that contains F. Every polynomial in $E[x]$ has coefficients in E_n for some n, and consequently it has a root in $E_{n+1} \subseteq E$.

Let $\overline{F}$ be the algebraic closure of F in E. For each polynomial f in $\overline{F}[x]$ we have that f has a root α in E. The element α is then algebraic over $\overline{F}$, and consequently algebraic over $F(\alpha_1, \dots, \alpha_n)$ for some elements $\alpha_1, \dots, \alpha_n$ in $\overline{F}$. Consequently we have that α is algebraic over F.

We have proved that $\overline{F}$ is an algebraic extension of F which is algebraically closed.

5.2 Irreducibility of the matrix groups

Recall that a topological space is *irreducible* if two nonempty open subsets always intersect (see Remark 5.1.23). Hence irreducible spaces are connected. For quasi affine varieties it is therefore more interesting to check irreducibility than to check connectedness. In this section we shall determine which of the matrix groups that are irreducible.

Lemma 5.2.1. *Let Y be a topological space. Assume that for every pair of points x and y of Y there is an irreducible topological space X and a continous map $f\colon X \to Y$, such that $f(X)$ contains x and y. Then Y is irreducible.*

Proof: Let U and V be open non-empty subsets of Y. Assume that $U \cap V = \emptyset$. Choose x in U and y in V, and let $f\colon X \to Y$ be a map such that x and y are in $f(X)$. We then have that $f^{-1}(U)$ and $f^{-1}(V)$ are nonempty open sets of X that do not intersect. This contradicts the irreducibility of X. Consequently each pair of open sets intersect and we have that Y is connected. □

Proposition 5.2.2. *We have that* $\mathrm{Gl}_n(\mathbf{K})$ *and* $\mathrm{Sl}_n(\mathbf{K})$ *are irreducible.*

Proof: If follows from Proposition 1.5.2 that every element A of $\mathrm{Gl}_n(\mathbf{K})$ can be written in the form $A = E_{i_1,j_1}(a_1)\cdots E(a)\cdots E_{i_n,j_n}(a_n)$, where $E(a)$ is the matrix 1.5.2.1. We obtain a continous map $f\colon \mathbf{K}^n \times (\mathbf{K} \setminus 0) \to \mathrm{Gl}_n(\mathbf{K})$ such that the image of the point $(a_1,\dots,a_n,a)$ is the matrix $E_{i_1,j_1}(a_1)\cdots E(a)\cdots E_{i_n,j_n}(a_n)$. Clearly $f(0,\dots,0,1) = I_n$ and $f(a_1,\dots,a_n,a) = A$. We have that $\mathbf{K}^n \times \mathbf{K} \setminus 0$ is an open subset (see 5.1.12) of the irreducible set $\mathbf{A}^{n+1}_{kropp} = \mathbf{K}^n \times \mathbf{K}$ (see 5.1.27). Hence $\mathbf{K}^n \times \mathbf{K} \setminus 0$ is irreducible (see 5.1.23). It follows from Lemma 5.2.1 that $\mathrm{Gl}_n(\mathbf{K})$ is irreducible. In the case of the groups $\mathrm{Sl}_n(\mathbf{K})$. We have that $a = 1$ and we can use the map $f\colon \mathbf{K}^n \to \mathrm{Sl}_n(\mathbf{K})$ sending the point $(a_1,\dots,a_n)$ to the matrix $E_{i_1,j_1}(a_1)\cdots E_{i_n,j_n}(a_n)$. We conclude, as above, that $\mathrm{Sl}_n(\mathbf{K})$ is irreducible. □

Proposition 5.2.3. *Let $\mathbf{K}$ be a field such that $2 \neq 0$ we have that* $\mathrm{SO}_n(\mathbf{K})$ *is irreducible and* $\mathrm{O}_n(\mathbf{K})$ *is not irreducible.*

Proof: The determinant gives a surjective map $\det \mathrm{O}_n(\mathbf{K}) \to \{\pm 1\}$. Since $\{\pm 1\}$ is not irreducible when $2 \neq 0$ it follows from Lemma 5.2.1 that $\mathrm{O}_n(\mathbf{K})$ is not irreducible.

It follows from Proposition 1.9.4 that every element A in $\mathrm{SO}_n(\mathbf{K})$ can be written in the form $A = s_{x_1}s_{y_1}\cdots s_{x_m}s_{y_m}$, for some m, with $\langle x_i, x_k\rangle \neq 0 \neq \langle y_i, y_i\rangle$. Consider $\mathbf{K}^{mn}$ as the space consisting of m vectors in $\mathbf{K}^n$, that is as points $(a_{1,1},\dots,a_{1,n},\dots,a_{m,1},\dots,a_{mn})$. Let U_i be the open set in $\mathbf{K}^{mn}$ consisting of points $x = (a_{i,1},\dots,a_{i,n})$ such that $\langle x, x\rangle \neq 0$. Then $\cap_{i=1}^m U_i$ is open and it is nonempty because $\mathbf{K}$ has infinitely many elements (see Exercise 3.3.1). We define a map $\gamma\colon \cap_{i=1}^m U_i \to \mathrm{SO}_n(\mathbf{K})$ by $\gamma(z_1,\dots,z_m) = s_{x_1}s_{z_1}\cdots s_{x_m}s_{z_m}$. Clearly the map is continuous and we have that $\gamma(x_1,\dots,x_m) = I_n$ and $\gamma(y_1,\dots,y_m) = A$. Since $\cap_{i=1}^m U_i$ is an open subset of $\mathbf{K}^{mn}$ it is follows from Example 5.1.27 and Remark 5.1.23 that it is irreducible. It follows from Lemma 5.2.1 that $\mathrm{SO}_n(\mathbf{K})$ is irreducible. □

Proposition 5.2.4. *We have that* $\mathrm{Sp}_n(\mathbf{K})$ *is irreducible.*

Proof: It follows from Proposition 1.9.9 that every element A in $\mathrm{Sp}_n(\mathbf{K})$ can be written in the form $A = \tau(x_1, a_1) \cdots \tau(x_m, a_m)$, for some m. The map $f\colon \mathbf{K}^n \to \mathrm{Sp}_n(\mathbf{K})$ which is defined by $f(b_1, \dots, b_m) = \tau(x_1, b_1) \dots \tau(x_m, b_m)$ is clearly continous and maps $(0, \dots, 0)$ to I_n and $(a_1, \dots, a_m)$ to A. It follows from Lemma 5.2.1 that $\mathrm{Sp}_n(\mathbf{K})$ is irreducible. □

5.3 Regular functions

The only natural functions on affine varieties are functions induced by polynomials or quotients of polynomials. We shall, in this section, define polynomial functions on quasi affine varieties.

Definition 5.3.1. Let X be an affine variety in $\mathbf{A}^n_{\mathbf{K}}$. Denote by

$$\mathcal{I}(X) = \{f \in \mathbf{K}[x_1, \dots, x_n] : f(x) = 0, \quad \text{for all } x \in X\},$$

the set of polynomials in $\mathbf{K}[x_1, \dots, x_n]$ that vanish on X.

5.3.2. Since the sum of two polynomials that both vanish on X, and the product of a polynomial that vanish on X with an arbitrary polynomial, vanish on X, we have that $\mathcal{I}(X)$ is an ideal. This ideal has the property that, if f is a polynomial in $\mathbf{K}[x_1, \dots, x_n]$ such that f^m is in $\mathcal{I}(X)$, for some positive integer m, then f is in $\mathbf{K}[x_1, \dots, x_n]$. Indeed, if $f(x)^m = 0$, for all x in X, then $f(x) = 0$, for all x in X. We say that the ideal $\mathcal{I}(X)$ is *radical.*

Definition 5.3.3. Let R be a ring. For each ideal I of R we let

$$\sqrt{I} = \{a \in R | a^m \in I, \quad \text{for some positive integer } m\}.$$

We call $\sqrt{I}$ the *radical* of I, and we say that the ideal I is *radical* if $\sqrt{I} = I$.

Remark 5.3.4. The radical of an ideal I contains I, and is itself an ideal. Indeed, if a is in $\sqrt{I}$, then a^m is in I for some positive integer m. Hence, for all b in I, we have that $b^m a^m = (ba)^m$ is in I. Consequently ba is in $\sqrt{I}$. Moreover, if a and b are in $\sqrt{I}$, then a^p and b^q are in I for some positive integers p and q. Let $m = p + q - 1$. Then $(a+b)^m = \sum_{i=0}^m \binom{m}{i} a^i b^{m-1}$. For $i = 0, \dots, m$, we have that, either $i \geq p$ or $m - i \geq q$, and consequently each term $\binom{m}{i} a^i b^{m-i}$ is in I. Hence $(a+b)^m$ is in I, and we have that $a + b$ is in $\sqrt{I}$.

We note that if I is a *proper ideal* of R, that is, if it is different from R, then $\sqrt{I}$ is proper. Indeed, the element 1 can not be in $\sqrt{I}$.

5.3.5. Let f be a polynomial in $\mathbf{K}[x_1, \dots, x_n]$. We define a function

$$\varphi_f\colon \mathbf{A}^n_{\mathbf{K}} \to \overline{\mathbf{K}},$$

by $\varphi_f(a_1, \dots, a_n) = f(a_1, \dots, a_n)$. For each affine subvariety X of $\mathbf{A}^n_{\mathbf{K}}$ we obtain by restriction a function

$$\varphi_f|_X\colon X \to \mathbf{K}.$$

Let g be another polynomial in $\mathbf{K}[x_1, \dots, x_n]$. By the definition of the ideal $\mathcal{I}(X)$, we have that $\varphi_f|_X = \varphi_g|_X$ if and only if $f - g$ is in $\mathcal{I}(X)$. It is therefore natural to consider the residue ring $\mathbf{K}[x_1,\ldots,x_n]/\mathcal{I}(X)$ of $\mathbf{K}[x_1,\ldots,x_n]$ by the ideal $\mathcal{I}(X)$ (see Example 3.5.2) to be the ring of polynomial functions on X.

Definition 5.3.6. Let X be an algebraic variety in $\mathbf{A}^n_{\overline{\mathbf{K}}}$. We denote the residue ring $\mathbf{K}[x_1,\ldots,x_n]/\mathcal{I}(X)$ by $\mathbf{K}[X]$, and call $\mathbf{K}[X]$ the *coordinate ring* of X. Let f be an element of $\mathbf{K}[X]$. We saw in Paragraph 5.3.5 that all the polynomials F in $\mathbf{K}[x_1,\ldots,x_n]$ whose residue is f take the same value $F(x)$ at each point x of X. The common value we denote by $f(x)$. We say that f is the function *induced* by F, and define X_f to be the set $\{x \in X \quad | \quad f(x) \neq 0\}$, or equivalently $X_f = X_F$.

Example 5.3.7. We have that the coordinate ring $\mathbf{K}[\mathbf{A}^n_{\overline{\mathbf{K}}}]$ of the affine variety $\mathbf{A}^n_{\overline{\mathbf{K}}}$ is equal to $\mathbf{K}[x_1,\ldots,x_n]$. Indeed, the only polynomial that is zero on $\mathbf{A}^n_{\overline{\mathbf{K}}}$ is 0 (see Example 5.1.27).

Definition 5.3.8. Let U be a quasi affine variety in $\mathbf{A}^n_{\overline{\mathbf{K}}}$. A function

$$\varphi\colon U \to \overline{\mathbf{K}}$$

is *regular* if there, for every point x in U, exists a neighbourhood V of x contained in U and polynomials $f(x_1,\ldots,x_m)$ and $g(x_1,\ldots,x_n)$ in $\mathbf{K}[x_1,\ldots,x_n]$ such that $g(a_1,\ldots,a_n) \neq 0$ and $\varphi(a_1,\ldots,a_n) = \frac{f(a_1,\ldots,a_n)}{g(a_1,\ldots,a_n)}$, for all $(a_1,\ldots,a_n)$ in V.

Let V be a quasi affine variety in $\mathbf{A}^m_{\overline{\mathbf{K}}}$. A map

$$\Phi\colon U \to V$$

is a *regular map* if there are regular functions $\varphi_1,\ldots,\varphi_m$ on U such that

$$\Phi(a_1,\ldots,a_n) = (\varphi_1(a_1,\ldots,a_n),\ldots,\varphi_m(a_1,\ldots,a_n)),$$

for all $(a_1,\ldots,a_n)$ in U.

Example 5.3.9. Let U and V be quasi affine varieties in $\mathbf{A}^m_{\overline{\mathbf{K}}}$ respectively $\mathbf{A}^n_{\overline{\mathbf{K}}}$, and let $f_1(x_1,\ldots,x_n),\ldots,f_m(x_1,\ldots,x_n)$ be polynomials in $\mathbf{K}[x_1,\ldots,x_n]$. The map

$$\Phi\colon U \to \mathbf{A}^m_{\overline{\mathbf{K}}}$$

defined by

$$\Phi(a_1,\ldots,a_n) = (\varphi_1(a_1,\ldots,a_n),\ldots,\varphi_m(a_1,\ldots,a_n))$$

is regular. If $\Phi(a_1,\ldots,a_n)$ is in V, for all $(a_1,\ldots,a_n)$ in U, we have that Φ induces a regular map

$$\Phi|_U\colon U \to V.$$

Since the multiplication map $\mathrm{Gl}_n(\mathbf{K}) \times \mathrm{Gl}_n(\mathbf{K}) \to \mathrm{Gl}_n(\mathbf{K})$, is given by polynomials, it is a regular map. Here $\mathrm{Gl}_n(\mathbf{K}) \times \mathrm{Gl}_n(\mathbf{K})$ is the product of quasi affine varieties in $\mathbf{A}^{n^2}_{\overline{\mathbf{K}}} \times \mathbf{A}^{n^2}_{\overline{\mathbf{K}}}$, given in Example 5.1.12. It follows that the product maps of all the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, or $\mathrm{SG}_S(\mathbf{K})$, for some invertible matrix S, are regular maps.

Example 5.3.10. The inverse map $\mathrm{Gl}_n(\mathbf{K}) \to \mathrm{Gl}_n(\mathbf{K})$ which sends a matrix A to A^{-1} is regular. Indeed, we have that $A^{-1} = \frac{1}{\det A}B$, where B is the adjoint matrix (see Section 1.4 and Exercise 1.4.2). Every coordinate of A^{-1} is a polynomial in the variables x_{ij} divided by the polynomial $\det(x_{ij})$, which is nonzero on all points of $\mathrm{Gl}_n(\mathbf{K})$. Consequently, the inverse map on $\mathrm{Gl}_n(\mathbf{K})$ is regular. It follows that the inverse map is regular for the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, or $\mathrm{SG}_S(\mathbf{K})$, for all invertible matrices S.

Example 5.3.11. Let $f(x_1, \dots, x_n)$ be an element in the polynomial ring $\mathbf{K}[x_1, \dots, x_n]$, and let X be an affine algebraic variety in $\mathbf{A}^n_{\mathbf{K}}$. The map

$$\Phi\colon X_f \to \mathcal{V}(1 - x_{n+1}f(x_1, \dots, x_n)),$$

defined by $\Phi(a_1, \dots, a_n) = (a_1, \dots, a_n, \frac{1}{f(a_1,\dots,a_n)})$, from the *principal* set X_f to the zeroes in $\mathbf{A}^{n+1}_{\mathbf{K}}$ of the polynomial $1 - x_{n+1}f(x_1, \dots, x_n)$ in $\mathbf{K}[x_1, \dots, x_{n+1}]$, is regular, and is given by the polynomials $x_1, \dots, x_n, \frac{1}{f(x_1,\dots,x_n)}$. The map

$$\Psi\colon \mathcal{V}(1 - x_{n+1}f(x_1, \dots, x_n)) \to X_f,$$

given by $\Psi(a_1, \dots, a_{n+1}) = (a_1, \dots, a_n)$ is also regular. The regular maps Φ and Ψ are inverses.

Lemma 5.3.12. *A regular map between quasi affine varieties is continous.*

Proof: Let U and V be quasi affine varieties in $\mathbf{A}^n_{\mathbf{K}}$ respectively $\mathbf{A}^m_{\mathbf{K}}$, where V is an open subset of an affine variety Y, and let $\Phi\colon U \to V$ be a regular map. It follows from Lemma 5.1.14 that it suffices to prove that $\Phi^{-1}(Y_g)$ is open in U, for all polynomials $g(y_1, \dots, y_m)$ in the m variables $y_1, \dots, y_m$ with coordinates in $\mathbf{K}$, such that Y_g is contained in V.

Let x be a point of $\Phi^{-1}(Y_g)$. Since Φ is regular there are open neighbourhoods U_i of x in U and polynomials f_i, g_i in $\mathbf{K}[x_1, \dots, x_n]$, such that $g_i(a_1, \dots, a_n) \neq 0$, and such that $\Phi(a_1, \dots, a_n) = \left(\frac{f(a_1,\dots,a_n)}{g_1(a_1,\dots,a_n)}, \dots, \frac{f_m(a_1,\dots,a_n)}{g_m(a_1,\dots,a_n)}\right)$, for all $(a_1, \dots, a_n)$ in $W_x = \cap_{i=1}^m U_i$. Write $f(x_1, \dots, x_n) = g\left(\frac{f_1(x_1,\dots,x_n)}{g_1(x_1,\dots,x_n)}, \dots, \frac{f_m(x_1,\dots,x_n)}{g_m(x_1,\dots,x_n)}\right)$. For a sufficiently big integer d we have that

$$h(x_1, \dots, x_n) = (g_1(x_1, \dots x_n) \cdots g_m(x_1, \dots, x_n))^d f(x_1, \dots, x_n)$$

is a polynomial in $x_1, \dots, x_n$. We have that

$$\begin{aligned}(\Phi|_{W_x})^{-1}(Y_g) &= W_x \cap \{(a_1, \dots, a_n) : f(a_1, \dots, a_n) \neq 0\} \\ &= W_x \cap \{(a_1, \dots, a_n) : h(a_1, \dots, a_n) \neq 0\} = W_x \cap X_h.\end{aligned}$$

Hence, for each x in U, the set $\Phi^{-1}(Y_g)$ contains an open subset $W_x \cap X_h$ containing x, and hence it is open. □

5.4 The Hilbert Nullstellensatz

The most fundamental result about regular functions is that the ring of regular functions on an affine algebraic set is canonically isomorphic to the coordinate ring. In order to prove this result we shall need the Hilbert Nullstellensatz. The algebraic prerequisites that we need to prove the Hilbert Nullstellensatz are quite extensive, and we have devoted the next section to the prerequisites, and to a proof of a generalized version of the Hilbert Nullstellensatz.

5.4.1. In Section 5.1 we associated an affine variety $\mathcal{V}(I)$ of $\mathbf{A}^n_{\mathbf{K}}$ to every ideal I in $\mathbf{K}[x_1, \dots, x_n]$. Conversely, in Section 5.3 we saw how we can associate a radical ideal $\mathcal{I}(X)$ to every affine variety of $\mathbf{A}^n_{\mathbf{K}}$. For every affine variety we have that

$$\mathcal{V}\mathcal{I}(X) = X.$$

Indeed, the inclusion $X \subseteq \mathcal{V}\mathcal{I}(X)$ is clear. To prove the opposite inclusion we take a point x of $\mathbf{A}^n_{\mathbf{K}} \setminus X$. Since X is an affine variety there is an ideal I of $\mathbf{K}[x_1, \dots, x_n]$ such that $X = \mathcal{V}(I)$. Clearly, we have that $I \subseteq \mathcal{I}(X)$ and since $x \notin X$, there is a polynomial f in I such that $f(x) \neq 0$. Consequently, x is not in $\mathcal{V}\mathcal{I}(X)$, and we have proved that the inclusion $\mathcal{V}\mathcal{I}(X) \subseteq X$ holds.

For every ideal I of $\mathbf{K}[x_1, \dots, x_n]$ it is clear that we have an inclusion

$$\sqrt{I} \subseteq \mathcal{I}\mathcal{V}(I).$$

The *Hilbert Nullstellensats* asserts that the opposite inclusion holds. In particular we must have that, if I is a proper ideal in $\mathbf{K}[x_1, \dots, x_n]$, then $\mathcal{V}(I)$ is not empty. Indeed, if $\mathcal{V}(I)$ were empty, then $\mathcal{I}\mathcal{V}(I)$ must be the whole of $\mathbf{K}[x_1, \dots, x_n]$. However, if I is a proper ideal, then so is $\sqrt{I}$ (see Remark 5.3.4).

The Hilbert Nullstellensats states that, if $f(x_1, \dots, x_n)$ is a polynomial in the ring $\mathbf{K}[x_1, \dots, x_n]$ which vanishes on the common zeroes $\mathcal{V}(f_1, \dots, f_n)$ of a family of polynomials $f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)$, then there is a positive integer d and polynomials $g_1, \dots, g_m$ such that

$$f^d = g_1 f_1 + \cdots + g_m f_m.$$

Indeed, let I be the ideal generated by the polynomials $f_1, \dots, f_m$. Let f be a polynomial that vanishes on $\mathcal{V}(I)$, that is, the polynomial f is in $\mathcal{I}\mathcal{V}(I)$. Then, by the Hilbert Nullstellensatz f is in $\sqrt{I}$ so that f^d is in I for som positive integer d. That is, we have $f^d = g_1 f_1 + \cdots + g_m f_m$ for some polynomials $g_1, \dots, g_m$.

The Hilbert Nullstellensats is a fundamental result in algebraic geometry and has many uses. We shall therefore present a proof of a very useful generalization. Before we start the proof we need some algebraic preliminaries.

Definition 5.4.2. An element a of a ring R is a *zero divisor* if there is a nonzero element b in R such that $ab = 0$. We say that R is an *integral domain* if 0 is the only zero divisor of R.

Definition 5.4.3. Let R be a ring. We say that a ring S is an *R-algebra* if R is a subring of S. A *homomorphism* $\Phi\colon S \to T$ of R algebras is a ring homomorphism which is the identity on R.

Given an R algebra S and elements $a_1, \dots, a_n$ of S. Then there is a unique R algebra homomorphism

$$\Phi\colon R[x_1, \dots, x_n] \to S,$$

from the ring of polynomials in the variables $x_1, \dots, x_n$ with coefficients in R, such that $\Phi(x_i) = a_i$, for $i = 1, \dots, n$. We have that $\Phi(f(x_1, \dots, x_n)) = f(a_1, \dots, a_n)$, for all polynomials $f(x_1, \dots, x_n)$ of $R[x_1, \dots, x_n]$.

The image of Φ is an R subalgebra of S that we denote by $R[a_1, \dots, a_n]$. We call $R[a_1, \dots, a_n]$ the R algebra *generated by the elements* $a_1, \dots, a_n$. When $S = R[a_1, \dots, a_n]$, we say that S is *finitely generated*, and that the elements $a_1, \dots, a_n$ are generators of the R algebra S. By definition $R[a_1, \dots, a_n]$ consists of all elements of the form $f(a_1, \dots, a_n)$, where f is a polynomial in $R[x_1, \dots, x_n]$, and it is clearly the smallest R subalgebra of S containing the elements $a_1, \dots, a_n$.

Let S be an R algebra which is an integral domain. We say that an element a of S is *algebraic* over R if there is a nonzero polynomial $f(x) = a_m x^m + \dots + a_0$, in the variable x with coefficients in R, such that $f(a) = 0$. An element of S which is not algebraic is called *transcendental*.

Lemma 5.4.4. *Let R be an integral domain and let $R[x]$ be the ring of polynomials in one variable x with coefficients in R. Moreover, let $f(x) = bx^m + b_{m-1}x^{m-1} + \dots + b_0$ be a polynomial with $b \neq 0$. For every polynomial $g(x)$ of $R[x]$ there is a nonnegative integer d and polynomials $q(x)$ and $r(x)$, with $\deg r < \deg f$, such that*

$$b^d g(x) = q(x)f(x) + r(x).$$

Proof: The assertion of the Lemma holds for all polynomials g such that $\deg g < \deg f$. Indeed, we can then take $d = 0$, $q = 0$ and $r = g$. We shall prove the lemma by induction on the degree of g. Assume that the assertion of the lemma holds for all polynomials of degree strictly less than p for some integer $p \geq \deg f = m$. Let $g(x) = cx^p + c_{p-1}x^{p-1} + \dots + c_0$. We have that $h(x) = bg(x) - cx^{p-m}f(x)$ is of degree less than p. By the induction assumption, we can find an integer d and polynomials q_1 and r such that $b^d h(x) = q_1(x)f(x) + r(x)$, with $\deg r < \deg f$. Consequently we have that $b^{d+1}g(x) = b^d h(x) + cb^d x^{p-m} f(x) = (q_1(x) + cb^d x^{p-m})f(x) + r(x)$, and we have proved that the assertion of the lemma holds for $g(x)$. □

Remark 5.4.5. We note that an element a of an R algebra S which is an integral domain is trancendental if and only if the surjection $\Phi\colon R[x] \to R[a]$ is an isomorphism. Indeed, the nonzero elements of the kernel of Φ consists of the nonzero polynomials $f(x)$ such that $f(a) = 0$. To determine the kernel of Φ when a is algebraic we choose a nonzero polynomial $f(x) = bx^m + b_{m-1}x^{m-1} + \dots + b_0$ of smallest possible degree m such that $f(a) = 0$. Then the kernel of Φ is equal to

$$\{g(x) \in R[x] \quad | \quad b^d g(x) = q(x)f(x), \text{where } d \in \mathbf{Z}, \text{with } d > 0, \text{and } q(x) \in R[x]\}.$$

Indeed, if $b^d g(x) = q(x)f(x)$ we have that $b^d g(a) = q(a)f(a) = 0$, and hence that $g(a) = 0$. Conversely, assume that $g(a) = 0$. It follows from Lemma 5.4.4 that $b^d g(x) = q(x)f(x) + r(x)$, for some nonnegative integer d and polynomials $q(x)$ and $r(x)$ with $\deg r(x) < \deg f(x)$. We obtain that $r(a) = b^d g(a) - q(a)f(a) = 0$. However, we have chosen f to be a nonzero polynomial of lowest degree such that $f(a) = 0$. It follows that $r(x) = 0$ in $R[x]$, and consequently $b^d g(x) = q(x)f(x)$.

Proposition 5.4.6. *Let R be a ring and S an R algebra which is an integral domain, and let a be an element of S. Moreover, let T be an integral domain and $\varphi\colon R \to T$ a ring homomorphism. Finally, let c be an element of T. The following two assertions hold:*

(i) *Assume that a is transcendental over R. Then there exists a unique ring homomorphism $\psi\colon R[a] \to T$ such that $\psi(a) = c$, and $\psi|_R = \varphi$.*

(ii) *Assume that a is algebraic and let $f(x) = bx^m + b_{m-1}x^{m-1} + \cdots + b_0$ be a polynomial of lowest possible degree in the variable x with coefficients in R such that $f(a) = 0$. If $\varphi(b) \neq 0$ and $\varphi(b)c^m + \varphi(b_{m-1})c^{m-1} + \cdots + \varphi(b_0) = 0$, there exists a unique homomorphism $\psi\colon R[a] \to T$ such that $\psi(a) = c$ and $\psi|_R = \varphi$.*

Proof: Let $\psi\colon R[a] \to T$ be a ring homomorphism such that $\psi(a) = c$, and $\psi|_R = \varphi$. For every element $g(a) = c_p a^p + \cdots + c_0$ of $R[a]$ we have that $\psi(g(a)) = \psi(c_p a^p + \cdots + c_0) = \psi(c_p a^p) + \cdots + \psi(c_0) = \psi(c_p)\psi(a)^p + \cdots + \psi(c_0) = \varphi(c_p)\psi(a)^p + \cdots + \varphi(c_0) = \varphi(c_p)c^p + \cdots + \varphi(c_0)$. Hence ψ is uniquely determined by the conditions that $\psi(a) = c$, and $\psi|_R = \varphi$.

Assume that a is transcendental. Then every element of $R[a]$ has an expression $g(a) = c_p a^p + \cdots + c_0$, where p, and $c_0, \dots, c_p$ are uniquely determined. Hence we can define a map $\psi\colon R[a] \to T$ by $\psi(g(a)) = \varphi(c_p)c^p + \cdots + \varphi(c_0)$. Clearly, ψ is a ring homomorphism such that $\psi(a) = c$, and $\psi|_R = \varphi$. Hence we have proved the first assertion of the proposition when a is trancendental.

Assume that a is algebraic. Then every element of $R[a]$ can be written in the form $g(a) = c_p a^p + \cdots + c_0$, for some polynomial $g(x) = c_p x^p + \cdots + c_0$ in the variable x with coefficients in R. Let $h(x) = d_q x^q + \cdots + d_0$. It follows from Remark 5.4.5 that $g(a) = h(a)$ if and only if $b^d(g(x) - h(x)) = q(x)f(x)$, for some nonnegative integer d, and some polynomial $q(x)$ in $R[x]$. Hence we can define a map $\psi\colon R[a] \to T$ by $\psi(e_r a^r + \cdots + e_0) = \varphi(e_r)c^r + \cdots + \varphi(e_0)$, if we can prove that $b^e(e_r x^r + \cdots + e_0) = p(x)f(x)$, for some nonegative integer e and some polynomial $p(x) = f_s x^s + \cdots + f_0$ implies that $\varphi(e_r)c^r + \cdots + \varphi(e_0) = 0$. Assume that $b^e(e_r x^r + \cdots + e_0) = p(x)f(x)$. We use φ on the coefficients of the monomials x^i in both sides of $b^e(e_r x^r + \cdots + e_0) = p(x)f(x)$, and substitute c for x. Then we obtain that $\varphi(b)^e(\varphi(e_r)c^r + \cdots + \varphi(e_0)) = (\varphi(f_s)c^s + \cdots + \varphi(f_0))(\varphi(b)c^m + \varphi(b_{m-1})c^{m-1} + \cdots + \varphi(b_0)) = 0$. Since $\varphi(b) \neq 0$ by assumption, and T is an integral domain by assumption, we obtain that $\varphi(e_r)c^r + \cdots + \varphi(e_0) = 0$. Thus we have proved that we can define a map $\psi\colon R[a] \to T$ by $\psi(e_r a^r + \cdots + e_0) = \varphi(e_r)c^r + \cdots + \varphi(e_0)$. We clearly have that ψ is a ring homomorphism, that $\psi(a) = c$, and that $\psi|_R = \varphi$. □

Lemma 5.4.7. *Let S be an R algebra which is an integral domain. Let a be an element of S which is algebraic over R and let $f(x) = bx^m + b_{m-1}x^{m-1} + \cdots + b_0$ be a polynomial of smallest possible degree m such that $f(a) = 0$. For all polynomials $g(x)$ such that $g(a) \neq 0$ there exists polynomials $p(x)$ and $q(x)$ such that*

$$p(x)f(x) + q(x)g(x) = c,$$

where c is a non zero element c of R.

Proof: Let $r(x)$ be a nonzero polynomial of the lowest possible degree such that $p(x)f(x) + q(x)g(x) = r(x)$ for some polynomials $p(x)$ and $q(x)$, and such that $r(a) \neq 0$. Such a polynomial exists since it follows from Lemma 5.4.4 that we can find a non negative integer

d, and polynomials $s(x)$ and $q(x)$ such that $\deg s < \deg f$ and such that $-q(x)f(x) + a^d g(x) = s(x)$. Here $s(a) = -q(a)f(a) + a^d g(a) = a^d g(a) \neq 0$.

We shall show that $r(x)$, in fact, has degree 0. Assume that $\deg r > 1$, and write $r(x) = cx^q + c_{q-1}x^{q-1} + \cdots + c_0$, with $c \neq 0$ and $q > 1$. It follows from Lemma 5.4.4 that we can write $c^d f(x) = q_1(x)r(x) + r_1(x)$, for some nonnegative integer d and polynomials $q_1(x)$ and $r_1(x)$, with $\deg r_1 < \deg r$. We have that $r_1(a) \neq 0$ because, if $r_1(a) = 0$, then $0 = c^d f(a) = q_1(a)r(a) + r_1(a) = q_1(a)r(a)$, and hence either $q_1(a) = 0$ or $r(a) = 0$. However, since $\deg f > \deg r \geq 1$, both $q_1(x)$ and $r(x)$ have lower degree than f and can not be zero at a because $f(x)$ is chosen of mininal degree such that $f(a) = 0$. We have that

$$\begin{aligned} r_1(x) &= c^d f(x) - q_1(x)r(x) = c^d f(x) - q_1(x)p(x)f(x) - q_1(x)q(x)g(x) \\ &= \left(c^d - q_1(x)p(x)\right) f(x) - q_1(x)q(x)g(x). \end{aligned}$$

The last equation, together with the observation that $r_1(a) \neq 0$ contradicts the minimality of the degree of $r(x)$. Hence we can not have that $\deg r \geq 1$, and we have proved the lemma. □

Theorem 5.4.8. *Let R be a ring and S an R algebra which is an integral domain. Moreover let $a_1, \dots, a_n$ be elements in S, and b be a nonzero element of $R[a_1, \dots, a_n]$. Then there is an element a in R such that, for every ring homomorphism $\varphi\colon R \to \overline{\mathbf{K}}$ such that $\varphi(a) \neq 0$ there is a ring homomorphism $\psi\colon R[a_1, \dots, a_n] \to \overline{\mathbf{K}}$ such that $\psi(b) \neq 0$, and such that $\psi|_R = \varphi$.*

Proof: We shall prove the theorem by induction on the number n of generators $a_1, \dots, a_n$ of $R[a_1, \dots, a_{n-1}]$. Let $R' = R[a_1, \dots, a_n]$. Then $R'[a_n] = R[a_1, \dots, a_n]$. We shall first prove the theorem with R' and $R'[a_n]$, for R and S. In particular we prove the theorem for the case $n = 1$.

Assume first that a_n is transcentdental over R'. Then $b = a'a_n^p + f_{p-1}a_n^{p-1} + \cdots + f_0$, for some elements $a', f_{p-1}, \dots, f_0$ of R'. For every homorphism $\varphi'\colon R' \to \overline{\mathbf{K}}$ such that $\varphi'(a') \neq 0$, we choose an element c of $\overline{\mathbf{K}}$ such that $\varphi'(a')c^p + \varphi'(f_{p-1})c^{p-1} + \cdots + \varphi'(f_0) \neq 0$. This is possible because $\overline{\mathbf{K}}$ is infinite (see Exercise 5.1.5), so that there are elements of $\overline{\mathbf{K}}$ that are not roots in the polynomial $\varphi'(a')x^p + \varphi'(f_{p-1})x^{p-1} + \cdots + \varphi'(f_0)$. It follows from the first assertion of Proposition 5.4.6 that there is a unique ring homomorphism $\psi\colon R'[a_n] \to \overline{\mathbf{K}}$ such that $\psi(a_n) = c$. We have that $\psi(b) = \psi(a'a_n^p + f_{p-1}a_n^{p-1} + \cdots + f_0) = \varphi'(a')\psi(a_n)^p + \varphi'(f_{p-1})\psi(a_n)^{p-1} + \cdots + \varphi'(f_0) = \varphi'(a')c^p + \varphi'(f_{p-1})c^{p-1} + \cdots + \varphi'(f_0) \neq 0$, and $\psi|_{R'} = \varphi'$, and we have proved the case $n = 1$ of the theorem when a_n is transcentdental.

Assume that a_n is algebraic over R'. Let $f(x) = cx^m + b_{m-1}x^{n-1} + \cdots + b_0$ be a polynomial in x with coefficients in R' of lowest degree m such that $f(a_m) = 0$. There is a polynomial $g(x) = c_p x^p + \cdots + c_0$ such that $b = g(a_n) \neq 0$. It follows from Lemma 5.4.7 that we can find polynomials $p(x)$ and $q(x)$ in $R'[x]$ such that $p(x)f(x) + q(x)g(x) = d$ is a nonzero element of R'. Let $a' = cd$, and let $\varphi'\colon R' \to \overline{\mathbf{K}}$ be a ring homomorphism such that $\varphi'(a') \neq 0$. Then $\varphi'(c) \neq 0$. Since $\overline{\mathbf{K}}$ is algebraically closed we can find a root e in $\overline{\mathbf{K}}$ of the polynomial

$\varphi'(c)x^m + \varphi'(b_{m-1})x^{m-1} + \cdots + \varphi'(b_0)$. It follows from part two of Proposition 5.4.6 that there is a ring homomorphism $\psi\colon R'[a_n] \to R'$ such that $\psi(a_n) = e$, and $\psi|_{R'} = \varphi'$. We have that $\psi(q(a_n))\psi(g(a_n)) = \psi(q(a_n)g(a_n)) = \psi(q(a_n)g(a_n)+p(a_n)f(a_n)) = \psi(d) = \varphi'(d)$, which is not zero because $\varphi'(a') = \varphi'(cd) \neq 0$. Hence we have that $\psi(g(a_n)) \neq 0$ and we have proved the case $n = 1$ of the theorem when a_n is algebraic.

We have proved the theorem in the case $n = 1$ and proceed by induction on n. Assume that the theorem holds for an algebra with $n-1$ generators. We use the induction assumption on the R algebra $R' = R[a_1, \dots, a_{n-1}]$ and the element a' of $R[a_1, \dots, a_{n-1}]$, used above. By the theorem we can find an element a of R such that every ring homomorphism $\varphi\colon R \to \overline{\mathbf{K}}$ such that $\varphi(a) \neq 0$ can be extended to a ring homomorphsim $\varphi'\colon R' \to \overline{\mathbf{K}}$ such that $\varphi'(a') \neq 0$, and such that $\varphi'|_{R'} = \varphi$. However, we have from the case $n = 1$ above that there is a ring homorphism $\psi\colon R[a_1, \dots, a_n] \to R$ such that $\psi(b) \neq 0$, and such that $\psi|_{R'} = \varphi'$. We have that $\psi|_R = \varphi'|_R = \varphi$, and we have proved the theorem. □

The Hilbert Nullstellensatz is a direct consequence of Theorem 5.4.8. In order to deduce the Hilbert Nullstellensatz from the Theorem it is however, convenient to use another characterization of the radical of an ideal, that illustrates why the radical is an ideal.

Lemma 5.4.9. *Let R be a ring and let I be an ideal of R. The radical of I is the intersection of all prime ideals in R that contain I.*

Proof: Let P be a prime ideal containing I. If a is in $\sqrt{I}$ we have that a^m is in I, and hence in P, for some positive integer m. Since P is prime it follows that either a or a^{m-1} is in P. By descending induction on m it follows that a is in P. Consequently, the radical of I is contained in the intersection of the primes containing I.

Conversely, let a be an element of R that is not contained in the radical of I. We shall show that there is a prime ideal containing I, but not a. Let $\{I_i\}_{i\in\mathcal{I}}$ be the family of ideals in R that contain I and that do not contain any power $1, a, a^2, \dots$ of a. Given any *chain of ideals* $\{I_i\}_{i\in\mathcal{J}}$, that is a subset of the family $\{I_i\}_{i\in\mathcal{I}}$, such that $I_i \subseteq I_j$ or $I_j \subseteq I_i$ for all i and j in $\mathcal{J}$. We have that $\cup_{i\in\mathcal{J}} I_i$ is an ideal that contains I and does not contain any power of a. Since every chain contains a maximal element the family $\{I_i\}_{i\in\mathcal{I}}$ contains a maximal element J (see Remark 5.4.10). We shall show that J is a prime ideal. Let b and c be elements of R that are not in J. The smallest ideals (b, J) and (c, J) that contain b and J, respectively c and J must contain a power of a, by the maximality of J. Consequently, $bb' + i = a^p$ and $cc' + j = a^q$, for some elements b' and c' of R and i and j of J. We take the product of the left and right hand sides of these expressions and obtain that $b'c'bc + cc'i + bb'j + ij = a^{p+q}$. Since $cc'i + bb'j + ij$ is in J, we obtain that, if bc were in J, then a^{p+q} would be in J, contrary to the assumption. Consequently, we have that bc is not in J, and J is prime. Thus, for every element a in R not contained in the radical of I we have a prime ideal J containing I, but not a. Hence the intersection of all prime ideals containing I is contained in the radical. □

Remark 5.4.10. In the proof or Lemma 5.4.9 we used Zorn's Lemma stating that, if every chain in a family $\{I_i\}_{i\in\mathcal{I}}$ of sets has a maximal element, then the family itself has maximal elements. For noetherian rings we can avoid the use of Zorn's Lemma by noting that a ring

R is noetherian, if and only if, every sequence $I_1 \subseteq I_2 \subseteq \cdots$ of ideals is *stationary*, that is $I_m = I_{m+1} = \cdots$, for some positive integer m. To prove this equvialence we first assume that R is noetherian and consider a sequence $I_1 \subseteq I_2 \subseteq \cdots$ of ideals. Let $I = \cup_{i=1}^{\infty} I_i$. Then I is an ideal, and thus generated by a finite number of elements $a_1, \dots, a_p$. Clearly we must have that all the generators must be in one of the ideals in the sequence, say I_m. Then we have that $I_m = I_{m+1} = \cdots = I$, and the sequence is stationary. Conversely, assume that every sequence is stationary. Given an ideal I of R and let $\{a_i\}_{i \in \mathcal{I}}$ be a set of generators. Choose ideals $I_1 \subset I_2 \subset \cdots$, where I_p is generated by $a_{i_1}, \dots, a_{i_p}$, by induction as follows:

We take I_1 to be the ideal generated by one of the generators a_{i_1}. Assume that we have chosen I_p, then, if $I_p \neq I$, we choose a generator $a_{i_{p+1}}$ that is not in I_p, and let I_{p+1} be the ideal generate by $a_{i_1}, \cdots, a_{i_{p+1}}$. Since, the chain must stop, by assumption, we must have that $I = I_m$, for some m, and thus that I is generated by $a_{i_1}, \dots, a_{i_m}$.

Theorem 5.4.11. (The Hilbert Nullstellensatz)*Let I be a proper ideal in the polynomial ring $\mathbf{K}[x_1, \dots, x_n]$. Then $\sqrt{I} = \mathcal{I}\mathcal{V}(I)$.*

Proof: We observed in Paragraph 5.4.1 that $\sqrt{I} \subseteq \mathcal{I}\mathcal{V}(I)$. To prove that the opposite inclusion holds, we take an element a not in $\sqrt{I}$ and shall find a point x in $\mathcal{V}(I)$ such that $a(x) \neq 0$. From the alternative description of the radical of Lemma 5.4.9 we can find a prime ideal P of $\mathbf{K}[x_1, \dots, x_n]$ which contains I and does not contain a. We have that the $\mathbf{K}$ algebra $S = \mathbf{K}[x_1, \dots, x_n]/P$ is an integral domain (see Exercise 5.5.1). Let g be the image of a in S by the residue map $\chi\colon \mathbf{K}[x_1, \dots, x_n] \to S$. Then $g \neq 0$. It follows from Theorem 5.4.8 with $\mathbf{K} = R$ and φ being the inclusion $\mathbf{K} \subseteq \overline{\mathbf{K}}$, that there is a $\mathbf{K}$ algebra homomorphism $\psi\colon S \to \overline{\mathbf{K}}$ such that $\psi(g) \neq 0$. Let a_i be the image of x_i by the composite $\zeta\colon \mathbf{K}[x_1, \dots, x_n] \to \overline{\mathbf{K}}$ of the residue map χ and ψ, for $i = 1, \dots, n$. Then $(a_1, \dots, a_n)$ is a point in $\mathbf{A}^n_{\overline{\mathbf{K}}}$. For each polynomial $f(x_1, \dots, x_n) = \sum_{i_1, \dots, i_n} a_{i_1 \dots i_n} x_1^{i_1} \cdots x_n^{i_n}$ in $\mathbf{K}[x_1, \dots, x_n]$ we have that $\zeta f(x_1, \dots, x_n) = \sum_{i_1, \dots, i_n} a_{i_1 \dots i_n} a_1^{i_1} \cdots a_n^{i_n} = f(a_1, \dots, a_n)$. When $f(x_1, \dots, x_n)$ is in I we have that it is in P, and hence that $\chi(f(x_1, \dots, x_n)) = 0$. Consequently we have that $\zeta(f) = 0$, that is $f(a_1, \dots, a_n) = 0$, or equivalently, $(a_1, \dots, a_n)$ is a zero for f. Hence we have that $(a_1, \dots, a_n)$ is in $\mathcal{V}(I)$. However, $\psi(g) \neq 0$, so $a(a_1, \dots, a_n) = \zeta(a(x_1, \dots, x_n)) = \psi\chi(a(x_1, \dots, x_n)) = \psi(g) \neq 0$. Hence, we have found a point $x = (a_1, \dots, a_n)$ in $\mathbf{A}^n_{\overline{\mathbf{K}}}$ which is a zero for I, and such that $a(x) \neq 0$, and we have proved the theorem. □

5.5 Prevarieties

A manifold is a topological space that *locally* looks like an open subset of $\mathbf{K}^n$, for some n, when $\mathbf{K}$ is the field of real or complex numbers. Analogously we define, in this section, prevarieties over arbitrary fields $\mathbf{K}$ as topological spaces that *locally* look like quasi affine varieties.

Definition 5.5.1. Let X be a topological space. An *algebraic chart* of X consists of an open set U of X, an open quasi affine variety V in some affine space $\mathbf{A}^n_{\mathbf{K}}$, and a homeomorphism $\varphi\colon V \to U$ of topological spaces. A family of algebraic charts $\{(\varphi_i, V_i, U_i)\}_{i \in I}$ is called an *algebraic atlas* if the open sets $\{U_i\}_{i\in I}$ cover X and if the map $\varphi_j^{-1}\varphi_i\colon \varphi_i^{-1}(U_i \cap U_j) \to \varphi_j(U_i \cap U_j)$ is regular, when $U_i \cap U_j$ is nonempty, for all indices i and j in $\mathcal{I}$. Here, and in the following, we write, for simplicity, $\varphi_j^{-1}\varphi_i$ for the map $(\varphi_j|(U_i \cap U_j))^{-1}\varphi_i|\varphi_i^{-1}(U_i \cap U_j)$. The set $\varphi_i^{-1}(U_i \cap U_j)$ where $\varphi_j^{-1}\varphi_i$ is defined will be clear from the context.

A compact topological space X together with an algebraic atlas is called an *algebraic prevariety*, or simply a *prevariety*. It is often convenient to include in the atlas all the homeomorphisms $\varphi\colon V \to U$, from an open quasi affine set in some $\mathbf{A}^n_{\overline{\mathbf{K}}}$ to an open subset in X, such that, for all $x \in U$ and some U_i in the chart that contain x, the homeomorphism $\varphi_i^{-1}\varphi$ is regular on $\varphi^{-1}(U \cap U_i)$. The condition then holds for all charts containing x. Such a maximal chart is called an *algebraic structure*.

For each open subset U of X the charts $\varphi_i\colon \varphi_i^{-1}(U \cap U_i) \to U \cap U_i$ define a structure as an algebraic prevariety on U, called the *induced structure*.

Example 5.5.2. The quasi affine varieties are themselves algebraic prevarieties with the identity map as a chart. In particular, all the matrix groups $\mathrm{Gl}_n(\mathbf{K})$, $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, or $\mathrm{SG}_S(\mathbf{K})$ for some invertible matrix S, are algebraic prevarieties (see Example 5.1.6).

Example 5.5.3. Let $S = \overline{\mathbf{K}}^{n+1} \setminus (0)$. Defining $(a_0, \dots, a_n)$ and $(b_0, \dots, b_n)$ to be related, if there is an a of $\mathbf{K}$ such that $a_i = ab_i$, for $i = 0, \dots, n$, we obtain a relation on S. This relation clearly is an equivalence relation. The set $(\overline{\mathbf{K}}^{n+1} \setminus (0))/\equiv$ is denoted $\mathbf{P}^n_{\overline{\mathbf{K}}}$, and is called the *projective space* of dimension n over $\mathbf{K}$. We have a canonical map

$$\Phi\colon \overline{\mathbf{K}}^{n+1} \to \mathbf{P}^n_{\overline{\mathbf{K}}},$$

and write $\Phi(a_1, \dots, a_n) = (a_1; \dots; a_n)$.

The sets U in $\mathbf{P}^n_{\overline{\mathbf{K}}}$ such that $\Phi^{-1}(U)$ is open in the Zariski topology on $\overline{\mathbf{K}}^{n+1}$, are the open sets in a topology on $\mathbf{P}^n_{\overline{\mathbf{K}}}$. By definition, the map Φ is continuous with respect to this topology and the Zariski topology on $\overline{\mathbf{K}}^n$.

For $i = 0, \dots, n$ we denote by H_i the subset of $\mathbf{P}^n_{\overline{\mathbf{K}}}$ consisting of points of the form $(a_0; \dots; a_{i-1}; 0; a_{i+1}; \dots; a_n)$. Then H_i is closed in the topology. Let $U_i = \mathbf{P}^n_{\overline{\mathbf{K}}} \setminus H_i$. Then the sets U_i, for $i = 0, \dots n$, form an open covering of $\mathbf{P}^n_{\overline{\mathbf{K}}}$. Let

$$\varphi_i\colon \mathbf{A}^n_{\overline{\mathbf{K}}} \to \mathbf{P}^n_{\overline{\mathbf{K}}}$$

be the map defined by $\varphi_i(a_1, \dots, a_n) = (a_1; \dots; a_{i-1}; 1; a_i; \dots; a_n)$. Then φ_i is a homeomorphism of $\mathbf{A}^n_{\overline{\mathbf{K}}}$ onto the open subset U_i of $\mathbf{P}^n(\mathbf{K})$. We have that the map $\varphi_j^{-1}\varphi_i$ is defined on the set $\varphi_i^{-1}(U_i \cap U_j)$ and is given by $\varphi_j^{-1}\varphi_i(a_1, \dots, a_n) = (\frac{a_1}{a_j}, \dots, \frac{a_{j-1}}{a_j}, \frac{a_{j+1}}{a_j}, \dots, \frac{a_n}{a_j})$, where $a_j \neq 0$ because $\varphi_i(a_1, \dots, a_n)$ is in $U_i \cap U_j$. Hence the map is regular. We see that (U_i, φ_i), for $i = 0, \dots, n$, define an algebraic chart on $\mathbf{P}^n_{\overline{\mathbf{K}}}$, which makes $\mathbf{P}^n_{\overline{\mathbf{K}}}$ into a prevariety.

Remark 5.5.4. Since every quasi affine variety is compact by Paragraph 5.1.19, we have that X is a prevariety if and only if there is an atlas consisting of a finite number of charts. Hence the condition that a prevariety is compact is not a serious restriction.

Note that an algebraic prevariety is covered by quasi affine subsets of some space $\mathbf{A}^n_{\overline{\mathbf{K}}}$. Such a quasi algebraic subset will also be quasi algebraic in any space $\mathbf{A}^m_{\overline{\mathbf{K}}}$ such that $\mathbf{A}^n_{\overline{\mathbf{K}}}$

is contained in $\mathbf{A}^m_{\mathbf{K}}$ as a closed subset. Hence the numbers n that appear in the definition of an algebraic prevariety are not determined by the algebraic prevariety. We shall later define the dimension of an algebraic prevariety.

Definition 5.5.5. Let X be an algebraic prevariety and U an open subset. A function $f\colon U \to \overline{\mathbf{K}}$ is *regular* if for every x in U and some chart $\varphi_i\colon V_i \to U_i$, such that x is contained in U_i, we have that the map $\varphi\varphi_i^{-1}$ is regular on $\varphi_i^{-1}(U \cap U_i)$. The condition then holds for all such charts. We denote by $\mathcal{O}_X(U)$ the set of all regular functions on U.

Remark 5.5.6. The set $\mathcal{O}_X(U)$ is clearly a ring, and for an open subset V of X contained in U there is a natural ring homomorphism $\rho_{U,V}\colon \mathcal{O}_X(U) \to \mathcal{O}_X(V)$ mapping a function f to its restriction $f|V$. The following two fundamental properties hold:

(i) If $f \in \mathcal{O}_X(U)$ and there is an open cover $\{U_i\}_{i\in I}$ of U such that $\rho_{U,U_i}(f) = 0$, for all $i \in I$, we have that $f = 0$.

(ii) If $\{U_i\}_{i\in I}$ is an open covering of U and $\{f_i\}_{i\in I}$ is a collection of functions $f_i \in \mathcal{O}_X(U_i)$ such that $\rho_{U_i,U_i\cap U_j}(f_i) = \rho_{U_j,U_i\cap U_j}(f_j)$, for all i and j, there is a function $f \in \mathcal{O}_X(U)$ such that $\rho_{U,U_i}(f) = f_i$, for all $i \in I$.

Consequently $\mathcal{O}_X$ is a *sheaf* on X (see Remark 3.4.9).

Example 5.5.7. Let X be a prevariety and x a point of X. Let S be the set consisting of pairs (U, f), where U is an open neighbourhood of x and f a regular function on U. We give a relation on S by defining (U, f) to be related to (V, g) if there is an open neighbourhood W of x, contained in $U \cap V$ such that $f|W = g|W$. Clearly this relation is an equivalence relation. The residual set $S/\equiv$ is denoted by $\mathcal{O}_{X,x}$. The elements of $\mathcal{O}_{X,x}$ can be added and multiplied by the rules $[(U, f)] + [(V, g)] = [(U \cap V, (f + g)|U \cap V)]$ and $[(U, f)][(V, g)] = [(U \cap V, (fg)|U \cap V)]$. Clearly $O_{X,x}$ becomes a ring with this addition and multiplication, zero being the element $[(X, 0)]$ and the unity the element $[(X, 1)]$.

For every open neighbourhood U of x we obtain a ring homomorphism

$$\mathcal{O}_X(U) \to \mathcal{O}_{X,x},$$

mapping (U, f) to $[(U, f)]$. The ring $\mathcal{O}_{X,x}$ is called the *ring of germs* of regular functions at x. We also have a ring homomorphism

$$\mathcal{O}_{X,x} \to \overline{\mathbf{K}}, \tag{5.5.7.1}$$

mapping $[(U, f)]$ to $f(x)$. This map is called the *augmentation map* at x.

Remark 5.5.8. Let U be an open neighbourhood of x. Then we have that the natural restriction map

$$\mathcal{O}_{X,x} \to \mathcal{O}_{U,x}$$

is an isomorphism.

Given a map $\Phi\colon Y \to X$ of prevarieties, we have a natural ring homomorphism

$$\Phi^*_x\colon \mathcal{O}_{X,f(x)} \to \mathcal{O}_Y, x \tag{5.5.8.1}$$

defined by $\Phi^*_x[(U, g)] = [(\Phi^{-1}(U), g\Phi)]$.

Definition 5.5.9. Let X and Y be prevarieties and $\Phi\colon Y \to X$ a continous map. We say that Φ is a *morphism* if, for every open subset U of X and every regular function $f\colon U \to \overline{\mathbf{K}}$ on U, we have that $f\Phi$ is regular on $\Phi^{-1}(U)$. When Φ has an inverse, which is also a morphism, we say that Φ is an *isomorphism* of prevarieties.

Remark 5.5.10. It follows immediately from the definition that if $\Psi\colon Z \to Y$ is another morphism of prevarieties, then $\Phi\Psi\colon Z \to X$ is also a morphism.

Let X be a topological space and U an open subset. We denote by $\mathcal{C}_X(U)$ the ring og all continous functions $U \to \overline{\mathbf{K}}$. It follows from Lemma 5.3.12 that, if X is a prevariety, the ring $\mathcal{O}_X(U)$ is a subring of $\mathcal{C}_X(U)$, for all open subsets U of X. A continous map $\Phi\colon Y \to X$ of topological spaces induces, for all open subsets U of X, a ring homomorphism $\mathcal{C}_X(U) \to \mathcal{C}_Y(\Phi^{-1}(U))$, which maps a function $g\colon U \to \overline{\mathbf{K}}$ to the composite $g\Phi\colon \Phi^{-1}(U) \to \overline{\mathbf{K}}$. When X and Y are prevarieties, this map is a morphism if and only if it induces a map $\Phi^*(U)\colon \mathcal{O}_X(U) \to \mathcal{O}_Y(\Phi^{-1}(U))$, on the subrings of regular functions. Clearly $\Phi^*(U)$ is a ring homomorphism and, when V is an open subset of U, we have that $\Phi^*(V)\rho_{U,V} = \rho_{\Phi^{-1}(U),\Phi^{-1}(V)}\Phi^*(U)$.

Remark 5.5.11. When U and V are quasi affine varieties a map $\Phi\colon V \to U$ is a morphism if and only if it is regular. Indeed, if Φ is regular, then, for every regular function $f\colon V \to \overline{\mathbf{K}}$ we have that $f\Phi\colon U \to \overline{\mathbf{K}}$ is regular. Consequently Φ is a morphism. Conversely, let $\Phi\colon V \to U$ be a morphism, and assume that V is a quasi affine variety in $\mathbf{A}^n_{\overline{\mathbf{K}}}$. Then the restriction to U of the coordinate functions $x_i\colon \mathbf{A}^n_{\overline{\mathbf{K}}} \to \overline{\mathbf{K}}$, that maps $(a_1, \dots, a_n)$ to a_i, is regular. Hence, the composite map $x_i|_U\colon V \to \overline{\mathbf{K}}$ is regular. Let $f_i = (x_i|_U)\Phi$, for $i = 1, \dots, n$. We have that $f_i(b_1, \dots, b_n) = (x_i|_U)\Phi(b_1, \dots, b_n)$. Consequently we have that $\Phi(b_1, \dots, b_m) = (f_1(b_1, \dots, b_n), \dots, f_n(b_1, \dots, b_m))$, and Φ is regular.

Example 5.5.12. Let X be an affine algebraic variety $\mathbf{A}^n_{\overline{\mathbf{K}}}$, and let $f(x_1, \dots, x_n)$ be a polynomial in $\mathbf{K}[x_1, \dots, x_n]$. We saw in Example 5.3.11 that the map

$$\Phi\colon X_f \to \mathcal{V}(1 - x_{n+1}f(x_1, \dots, x_n))$$

defined by $\Phi(a_1, \dots, a_n) = (a_1, \dots, a_n, \frac{1}{f(a_1,\dots,a_n)})$ is an isomorphism of the quasi affine variety X_f of $\mathbf{A}^n_{\overline{\mathbf{K}}}$, with the affine variety $\mathcal{V}(1 - x_{n+1}f(x_1, \dots, x_n))$ of $\mathbf{A}^{n+1}_{\overline{\mathbf{K}}}$.

In particular it follows from Lemma 5.1.14 that a prevariety can be covered by open subsets that are affine varieties.

A fundamental result, that we shall prove next, is that there is a natural isomorphism between the coordinate ring of an affine variety, and the ring of regular functions on the variety. Although it is not necessary for the proof, or for the apparent generalization given below, it is extremely convenient to introduce the localization of a ring with respct to multiplicatively closed subsets.

Definition 5.5.13. Let R be a ring. We call a subset S *multiplicatively closed*, if ab is in S, for all pairs of elements a, b in S. Let S be a multiplicatively closed subset and let $R \times S$ be the set of pairs (a, b), with a in R and b in S. We say that two pairs (a, b) and (c, d) in $R \times S$ are *related*, if $ead = ebc$, for some element e of S. This relation is an equivalence relation. Indeed, it is clearly reflexive and symmetric. To prove that it is transitive, let (f, g) be an element related to (c, d). Then there is an element h of S such that $hfd = hcg$. We obtain that $hedag = hebcg = hedbf$, where h, e and d, and thus hed, are contained in S. Consequently, (a, b) is related to (f, g). We denote by $S^{-1}R$ the set of equivalence classes. The class in $S^{-1}R$ of the pair (a, b) in $R \times S$ we denote by $\frac{a}{b}$.

We define addition and multiplication of elements in $S^{-1}R$ by the formulas:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \text{and} \quad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$

It is easily checked that these operations are *well defined*, that is, they are independent of the representative we chose of each equivalence class, and that $S^{-1}R$, with these operations become a ring with 0 and 1 given by $\frac{0}{a}$, respectively $\frac{a}{a}$, for any a in S. Moreover, we have a natural ring homomorphism

$$R \to S^{-1}R,$$

which maps a to $\frac{ab}{b}$, for any b in S. The homomorphism is not always injective. For example, if the zero element is in S, then $S^{-1}R = 0$, because (a, b) is equivalent to $(0, 0)$, for all a in R and b in S. We call the ring $S^{-1}R$ the *localization* of R with respect to the multiplicatively closed subset S.

Let a be an element of R, and let $S = \{1, a, a^2, \dots\}$. Clearly S is multiplicatively closed. In this case we let $S^{-1}R = R_a$. The map $R \to R_a$ is injective if and only if there does not exist a nozero element b in R such that $a^m b = 0$, for some positive integer m. It follows, by descending induction on m, that the condition holds if and only if there is no element b of R such that $ab = 0$.

Let P be a prime ideal of R. By the definition of a prime ideal, the set $S = R \setminus P$ is multiplicatively closed. We let $S^{-1}R = R_P$.

Let S be the set of non zero divisors of R. Then S is multiplicatively closed. Indeed, if a and b are not zero divisors, and c is a nonzero element such that $abc = 0$, then, either $bc = 0$, in which case b is a zero divisor, or $bc \neq 0$, and then $a(bc) = 0$, in which case a is a zero divisor. Hence ab is not a zero divisor. We denote the resulting ring $S^{-1}R$ by $K(R)$ and call $K(R)$ the *total quotient ring* of R. The map

$$R \to K(R)$$

is injective because, if a is an element that maps to $\frac{a}{1} = 0$, then there is a nonzero divisor b such that $ba = 0$. Consequently we have that $a = 0$. When R is an integral domain, then $K(R)$ is a field. Indeed, the inverse of a nonzero element $\frac{a}{b}$ of $K(R)$ is $\frac{b}{a}$.

Definition 5.5.14. Let X be an affine variety. For every nonzero element f in the coordinate ring $\mathbf{K}[X]$ we have a natural map

$$\mathbf{K}[X]_f \to \mathcal{O}_X(X_f),$$

which maps a quotient $\frac{g}{f^m}$ in $\mathbf{K}[X]_f$ to the function $X_f \to \overline{\mathbf{K}}$, which maps the point x to $\frac{g(x)}{f(x)^m}$.

For each point x of X we have a $\mathbf{K}$ algebra homomorphism

$$\mathbf{K}[X] \to \overline{\mathbf{K}}, \tag{5.5.14.1}$$

which sends an element f to $f(x)$. We call this map the *augmentation* at x. Let

$$\mathcal{M}_{X,x} = \{f \in \mathbf{K}[X] : f(x) = 0\}$$

be the kernel of the augmentation at x. It is clear that $\mathcal{M}_{X,x}$ is a prime ideal. It is also maximal, because if I were an ideal strictly containing $\mathcal{M}_{X,x}$ then it follows from Hilberts Nullstellensatz that I has a zero, this zero must then be x. Thus I must be the radical of $\mathcal{M}_{X,x}$, and thus $I = \mathcal{M}_{X,x}$, since $\mathcal{M}_{X,x}$ is prime.

We have a natural map

$$\mathbf{K}[X]_{\mathcal{M}_{X,x}} \to \mathcal{O}_{X,x},$$

which sends a quotient $\frac{f}{g}$ in $\mathbf{K}[X]_{\mathcal{M}_{X,x}}$, to the class of the function $X_g \to \overline{\mathbf{K}}$ that sends a point x to $\frac{f(x)}{g(x)}$.

Proposition 5.5.15. *Let X be an affine variety. For every element f in $\mathbf{K}[X]$, and point x of X the maps $\mathbf{K}[X]_f \to \mathcal{O}_X(X_f)$, and $\mathbf{K}[X]_{\mathcal{M}_{X,x}} \to \mathcal{O}_{X,x}$, are isomorphisms.*

Proof: We first show that the map $\mathbf{K}[X]_f \to \mathcal{O}_X(X_f)$ is injective. Assume that a quotient $\frac{g}{f^m}$ maps to zero in $\mathcal{O}_X(X_f)$. Then $g(x) = 0$ for x in X_f. However, then $fg(x) = 0$ for all x in X. That is $fg = 0$ in $\mathbf{K}[X]$. Hence $\frac{g}{f} = 0$ in $\mathbf{K}[X]_f$. The proof that $\mathbf{K}[X]_{\mathcal{M}_{X,x}} \to \mathcal{O}_{X,x}$ is injective is similar.

We next show that the map $\mathbf{K}[X]_f \to \mathcal{O}_X(X_f)$ is surjective. Let X be a closed subset of $\mathbf{A}^n_{\overline{\mathbf{K}}}$, and let s be an element in $\mathcal{O}_X(X_f)$. By definition there is an open covering $X_f = \cup_{i\in\mathcal{I}} U_i$ of X_f by open sets U_i, and polynomials f_i and g_i in $\mathbf{K}[x_1, \dots, x_n]$ such that $g_i(x) \neq 0$, and $s(x) = \frac{f_i(x)}{g_i(x)}$, for x in U_i. It follows from Lemma 5.1.14 that, refining the covering if necessary, we may assume that $U_i = X_{h_i}$, for some h_i in $\mathbf{K}[x_1, \dots, x_n]$. Since the sets $U_i = X_{h_i}$ cover X_f we have that, if $f(x) \neq 0$, for some x in X, there is an index i in $\mathcal{I}$ such that $h_i(x) \neq 0$, or equvalently $g_i(x)h_i(x) \neq 0$. That is, if $(g_ih_i)(x) = 0$, for all i in $\mathcal{I}$, then $f(x) = 0$. It follows from the Hilbert Nullstellensatz, applied to the ideal generated by the elements g_ih_i, that there is a finite subset $i_1, \dots, i_r$ of $\mathcal{I}$, elements $k_1, \dots, k_r$ of $\mathbf{K}[x_1, \dots, x_n]$, and a nonnegative integer m, such that

$$f^m = g_{i_1}h_{i_1}k_1 + \cdots + g_{i_r}h_{i_r}k_r.$$

Let

$$g = f_{i_1}h_{i_1}k_1 + \cdots + f_{i_r}h_{i_r}k_r.$$

For each point x in X_f there is an index j such that $h_{i_j}(x) \neq 0$. We obtain that

$$g(x) = \frac{f_{i_j}(x)}{g_{i_j}(x)} g_{i_1}(x)h_{i_1}(x)k_1(x) + \cdots + \frac{f_{i_j}(x)}{g_{i_j}(x)} g_{i_r}(x)h_{i_r}(x)k_r(x).$$

Indeed, on the one hand we have that , if x is in $X_{h_{i_l}}$, then $s(x) = \frac{f_{i_j}(x)}{g_{i_j}(x)} = \frac{f_{i_k}(x)}{g_{i_k}(x)}$, such that $\frac{f_{i_j}(x)}{g_{i_j}(x)} g_{i_l}(x) h_{i_l}(x) k_l(x) = f_{i_l}(x) h_{i_l}(x) k_l(x)$, and, on the other hand, if x is not in $X_{g_{i_l}}$, then $h_{i_l}(x) = 0$, such that $f_{i_l}(x) h_{i_l}(x) k_l(x) = 0 = \frac{f_{i_j}(x)}{g_{i_j}(x)} g_{i_l}(x) h_{i_l}(x) k_l(x)$. Consequently we have that

$$g(x) = \frac{f_{i_j}(x)}{g_{i_j}(x)} \left(g_{i_1}(x) h_{i_1}(x) k_1(x) + \cdots + g_{i_r}(x) h_{i_r}(x) k_r(x) \right) = \frac{f_{i_j}(x)}{g_{i_j}(x)} f^m(x).$$

We have proved that $\frac{f(x)}{g^m(x)} = s(x)$, for all x in X, and consequently, that the map $\mathbf{K}[X]_f \to \mathcal{O}_X(X_f)$ is surjective.

To show that the map $\mathbf{K}[X]_{\mathcal{M}_{X,x}} \to \mathcal{O}_{X,x}$ is surjective it suffices to observe that an element of $\mathcal{O}_{X,x}$, comes from an element of $\mathcal{O}_X(X_f)$, for some neighbourhood X_f of x. However, the latter element comes from an element of $\mathbf{K}[X]_f$, by what we just proved, and the last element clearly maps onto the first by the map $\mathbf{K}[X]_{\mathcal{M}_{X,x}} \to \mathcal{O}_{X,x}$. □

Remark 5.5.16. We note that with $f = 1$ we obtain, from Proposition 5.5.15, a natural isomorphism $\mathbf{K}[X] \to \mathcal{O}_X(X)$, for all affine varieties X. Given a morphism $\Phi\colon Y \to X$, the map $\mathcal{O}_X(X) \to \mathcal{O}_Y(Y)$ on regular functions, give a natural homomorphism $\Phi^*\colon \mathbf{K}[X] \to \mathbf{K}[Y]$ of $\mathbf{K}$ algebras.

The next result gives the fundamental connection between algebra and geometry on which algebraic geometry rests.

Proposition 5.5.17. *Let X be an affine variety and Y a prevariety. The correspondence that to a morphism*

$$\Phi\colon Y \to X$$

associates the $\mathbf{K}$ algebra homomorphism

$$\Phi^*\colon \mathbf{K}[X] \to \mathcal{O}_Y(Y),$$

obtained by composing the isomorphism $\mathbf{K}[X] \to \mathcal{O}_X(X)$ of Proposition 5.5.15 with the natural map $\mathcal{O}_X(X) \to \mathcal{O}_Y(Y)$, gives a bijection between the morphisms from Y to X and the $\mathbf{K}$ algebra homomorhpisms from $\mathbf{K}[X]$ to $\mathcal{O}_Y(Y)$.

In particular we have that X and Y are isomorphic affine varieties if and only if $\mathbf{K}[X]$ and $\mathbf{K}[Y]$ are isomorphic $\mathbf{K}$ algebras.

Proof: Given a $\mathbf{K}$ algebra homomorphism

$$\Psi\colon \mathbf{K}[X] \to \mathcal{O}_Y(Y).$$

We shall define a morphism

$$\Phi\colon Y \to X,$$

such that $\Phi^* = \Psi$. To this end we cover Y by open affine varieties $\{Y_i\}_{i \in \mathcal{I}}$. Assume that X is an affine variety in $\mathbf{A}^n_{\mathbf{K}}$ and that Y_i is an affine variety in $\mathbf{A}^m_{\mathbf{K}}$. Let $\rho_x\colon \mathbf{K}[x_1, \dots, x_n] \to \mathbf{K}[X]$, and $\rho_{Y_i}\colon \mathbf{K}[y_1, \dots, y_m] \to \mathbf{K}[Y_i]$ be the residue maps. Moreover let $\psi_i\colon \mathbf{K}[X] \to$

$\mathbf{K}[Y_i]$ be the composite of ψ with the map $\mathcal{O}_Y(Y) \to \mathcal{O}_Y(Y_i) = \mathcal{O}_{Y_i}(Y_i)$, and the inverse of the isomorphism $\mathbf{K}[Y_i] \to \mathcal{O}_{Y_i}(Y_i)$.

Choose polynomials $g_1(y_1, \dots, y_m), \dots, g_n(y_1, \dots, y_m)$ in the ring $\mathbf{K}[y_1, \dots, y_m]$ such that $\psi_i \rho_X x_j = \rho_{Y_i} g_j(y_1, \dots, y_m)$, for $j = 1, \dots, n$. Then we have an equality

$$\psi_j \rho_X(x_j)(b_1, \dots, b_m) = g_j(b_1, \dots, b_m),$$

for $j = 1, \dots, m$, and all $(b_1, \dots, b_m)$ in Y_i. Since $\psi_i \rho_X$ is a $\mathbf{K}$ algebra homomorphism we obtain that

$$\psi_i \rho_X(f(x_1, \dots, x_n)) = f(\psi \rho_X(x_1), \dots, \psi \rho_X(x_n)),$$

for all polynomials $f(x_1, \dots, x_n)$ in $\mathbf{K}[x_1, \dots, x_n]$. Hence we have that

$$\psi_i \rho_X(f)(b_1, \dots, b_n) = f(g_1(b_1, \dots, b_m), \dots, g_n(b_1, \dots, b_m)), \tag{5.5.17.1}$$

for all $(b_1, \dots, b_m)$ in Y_i. In particular, for all f in $\mathcal{I}(X)$, and all $(b_1, \dots, b_m)$ in Y_i, we have

$$f(g_1(b_1, \dots, b_m), \dots, g_n(b_1, \dots, b_m)) = 0.$$

Hence $(g_1(b_1, \dots, b_m), \dots, g_n(b_1, \dots, b_m))$ is in X for all $(b_1, \dots, b_m)$ in Y_i. Consequently, we can define a morphism

$$\Phi_i \colon Y_i \to X$$

by $\Phi_i(b_1, \dots, b_m) = (g_1, (b_1, \dots, b_m), \dots, g_n(b_1, \dots, b_m))$, for all $(b_1, \dots, b_m)$ in Y_i. It follows from Equation 5.5.17.1 that, for all $(b_1, \dots, b_m)$ in Y_i, and f in $\mathbf{K}[x_1, \dots, x_n]$, we have

$$\psi_i \rho_X(f)(b_1, \dots, b_m) = f \psi_i(b_1, \dots, b_m) = \Psi^* \rho_X(f).$$

Consequently, we have that $\Psi_i = \Phi_i^*$. Moreover, the map $Y_i \to X$ associated to Φ_i^* is Φ_i.

Given two open affine varieties Y_i and Y_j of Y, and let W be an affine variety that is an open subset of $Y_i \cap Y_j$. The composite of the map $\psi \colon \mathbf{K}[X] \to \mathbf{K}[Y_i]$ and $\psi_j \colon \mathbf{K}[X] \to \mathbf{K}[Y_j]$ with the map $\mathbf{K}[Y_i] \to \mathbf{K}[W]$, respectively $\mathbf{K}[Y_j] \to \mathbf{K}[W]$, obtained from $\mathcal{O}_{Y_i}(Y_i) \to \mathcal{O}_{Y_i}(W) = \mathcal{O}_W(W)$, respectively $\mathcal{O}_{Y_j}(Y_j) \to \mathcal{O}_{Y_j}(W) = \mathcal{O}_W(W)$, are the same. Consequently, the construction gives maps Φ_i and Φ_j that coincide on W. It follows that the maps $\Phi_i \colon Y_i \to X$, for all i in $\mathcal{I}$, induce a map $\Phi \colon Y \to X$, such that $\Phi|_{Y_i} = \Phi_i$, for all i. It is clear that $\Phi^* = \Psi$.

Conversely let $\Phi : Y \to X$ be a morphism. The above construction applied to the $\mathbf{K}$-algebra homomorphism $\Phi^* : \mathbf{K}[Y] \to \mathcal{O}_Y(Y)$ gives a morphism $Y \to X$ which clearly is Φ. Hence we have a natural bijection between the algebraic maps from Y to X and the $\mathbf{K}$ algebra homomorphisms $\mathbf{K}[X] \to \mathcal{O}_Y(Y)$, which associates Φ^* to Φ. □

Exercises

5.5.1. Let R be a ring. Show that an ideal I of R is prime if and only if the residue ring R/I is an integral domain.

5.5.2. Show that the total quotient ring $K(\mathbf{Z})$ of the integers is canonically isomorphic to the rational numbers $\mathbf{Q}$.

5.5.3. Show that the total quotient ring $K(\mathbf{K}[x_1,\dots,x_n])$ of the polynomial ring $\mathbf{K}[x_1,\dots,x_n]$ is the field of *rational functions*, that is, the field of all quotients $\frac{f(x_1,\dots,x_n)}{g(x_1,\dots,x_n)}$ of polynomials in $\mathbf{K}[x_1,\dots,x_n]$, with $g \neq 0$.

5.6 Subvarieties

In Sections 5.1 and 5.3 we defined affine varieties, coordinate rings and regular fuctions with respect to a fixed imbedding into an affine space. We proved that the coordinate ring, and regular functions, are independent of the imbedding. In this section we go one step further to liberate the consepts from the ambient spaces.

Definition 5.6.1. Let X and Y be prevarieties and assume that Y is a closed subset of X. We say that Y is a closed *sub prevariety* of X if the inclusion map $\iota : Y \to X$ of Y in X is a morphism, and if, for each point x of Y, we have that the map

$$\mathcal{O}_{X,x} \to \mathcal{O}_{Y,x}$$

of germs of regular functions at x that is induced by ι, is surjective. When X is an affine variety we say that Y is a *subvariety*.

Example 5.6.2. Let X be an affine variety in $\mathbf{A}^n_{\mathbf{K}}$, and Y a closed subset of X. Then Y is an affine variety as a closed subset of $\mathbf{A}^n_{\mathbf{K}}$, and the inclusion map of Y in X is a morphism. We have an inclusion $\mathcal{I}(X) \subseteq \mathcal{I}(Y)$ of ideals in $\mathbf{K}[x_1,\dots,x_n]$ and thus a surjection

$$\varphi \colon \mathbf{K}[X] \to \mathbf{K}[Y].$$

For each point x of Y we have a map

$$\varphi_y \colon \mathbf{K}[X]_{\mathcal{M}_{X,x}} \to \mathbf{K}[Y]_{\mathcal{M}_{Y,x}}$$

defined by $\varphi_y(\frac{f}{g}) = \frac{\varphi(f)}{\varphi(g)}$. This map is well defined because, if $g(x) \neq 0$, then $\varphi(g)(x) = g(x) \neq 0$, and it is surjective because φ is surjective. It follows from Proposition 5.5.15 that the map $\mathcal{O}_{X,x} \to \mathcal{O}_{Y,x}$ is surjective. Hence Y is a closed subvariety of X. It follows from Example 5.1.6 that the matrix groups $\mathrm{Sl}_n(\mathbf{K})$, $\mathrm{G}_S(\mathbf{K})$, and $\mathrm{SG}_S(\mathbf{K})$, for all invertible S, are closed subvarieties of the affine variety $\mathrm{Gl}_n(\mathbf{K})$.

Example 5.6.3. Let Y be a closed subset of a prevariety X. For each open affine subvariety U of X we have, by Example 5.6.2 that $U \cap Y$ is a subvarity of X with coordinate ring equal to $\mathbf{K}[U]/I$, where I is the ideal of elements f in $\mathbf{K}[U]$ such that $f(x) = 0$, for x in $U \cap Y$. Hence we can cover Y by algebraic charts of the type $U \cap Y$, where U is an affine variety which is open in X. These charts constitute an atlas on Y. Indeed, let $\varphi_i \colon V_i \to U_i$, for $i = 1, 2$, be two charts on X, and let W be an open affine subvariety of X, containing x and contained in $U_1 \cap U_2$. Then $\varphi_2^{-1}\varphi_1$ defines an isomorphism $\psi \colon \varphi_1^{-1}(W) \to \varphi_2^{-1}(W)$ which induces a homomorphism $\varphi_1^{-1}(W \cap Y) \to \varphi_2^{-1}(W \cap Y)$. Consequently, the homomorphism $\psi^*\colon \mathbf{K}[\varphi_2^{-1}(W)] \to \mathbf{K}[\varphi_1^{-1}(W)]$ induces a bijection between the ideal of functions vanishing on the closed set $\varphi_2^{-1}(W \cap Y)$ of $\varphi_2^{-1}(W)$ with the ideal of functions vanishing on the closed subset $\varphi_1^{-1}(W \cap Y)$ of $\varphi_1^{-1}(W)$. Hence ψ^* induces an isomorphism of coordinate rings $\mathbf{K}[\varphi_2^{-1}(W \cap Y)] \to \mathbf{K}[\varphi_1^{-1}(W \cap Y)]$. It follows from Proposition 5.5.17 that the corresponding morphism $\varphi_1^{-1}(W \cap Y) \to \varphi_2^{-1}(W \cap Y)$ is an isomorphism of affine varieties. It follows that the map $\varphi_1^{-1}(U_1 \cap Y) \to \varphi_2^{-1}(U_2 \cap Y)$ is an isomorphism. Consequently the charts defined by $\varphi_1|_{U_1 \cap Y}$ and $\varphi_2|_{U_2 \cap Y}$ are part of an atlas. Hence the same is true for any two of the charts we have defined on Y, and Y is a prevariety.

We saw in Example 5.6.2 that, for all affine subsets U of X, the map $U \cap Y \to U$ is a morphism and the map

$$\mathcal{O}_{U,x} \to \mathcal{O}_{U \cap Y,x}$$

of germs of regular functions at x is surjective for all points x of $U \cap Y$. However, the regular functions of a variety at a point is the same as that for an open neighbourhood of the point. Hence the map

$$\mathcal{O}_{X,x} \to \mathcal{O}_{Y,x}$$

is also surjective, and Y is a closed sub prevariety of X.

Proposition 5.6.4. *Let X and Y be prevarieties. Assume that Y is a closed subset of X and that the inclusion makes Y into a closed sub prevariety of X. Then the inclusion map induces an isomorphism between the prevariety Y and the prevariety induced, as in Excercise 5.6.3, on the closed subset underlying Y.*

Proof: Denote by Z the prevariety induced on the underlying closed set of Y in Excercise 5.6.3. It suffices to consider the structures on open subsets of X, so we may assume that X is an affine variety. We then have a surjection $\mathbf{K}[X] \to \mathbf{K}[Z]$ of coordinate rings given by the induced structure on Z as in Excample 5.6.3. Corresponding to the map $Y \to X$ it follows from Proposition 5.5.17 that we have a map of rings $\mathbf{K}[X] \to \mathcal{O}_Y(Y)$. Since the prevarieties Y and Z have the same underlying set the kernel of the maps $\mathbf{K}[X] \to \mathbf{K}[Z]$ and $\mathbf{K}[X] \to \mathcal{O}_Y(Y)$ are the same, and equal the elements f of $\mathbf{K}[X]$ that vanish on $Y = Z$. Consequently the surjective map $\mathbf{K}[X] \to \mathbf{K}[Z]$ gives rise to an injective map

$$\psi\colon \mathbf{K}[Z] \to \mathcal{O}_Y(Y),$$

and hence it follows form Proposition 5.5.17 that the inclusion map $\iota\colon Y \to Z$ is a morphism of prevarieties. It follows from Proposition 5.5.17 that the inclusion map induces an isomorphism if and only if the map ψ is an isomorphism. Hence, to prove the proposition, it remains to prove that ψ is surjective. The composite map

$$\mathcal{O}_{X,x} \to \mathcal{O}_{Z,x} \to \mathcal{O}_{Y,x}$$

is surjective, for all x in $Y = Z$, by assumption. Hence the right hand map is also surjective. This map is also an injection, for if a class $[(U, f)]$ in $\mathcal{O}_{Z,x}$ is mapped to zero, then $f\iota(x) = f(x) = 0$, for all x in a neighbourhood of x in Y, or, which is the same because ι is a homeomorphism, in a neighbourhood of x in Z. The same reasoning shows that the map $\mathcal{O}_Z(W) \to \mathcal{O}_Y(W)$ is injective, for all open sets W of $Y = Z$.

Let g be an element of $\mathcal{O}_Y(Y)$. For all x in Y there is a unique element s_x in $\mathcal{O}_{Z,x}$ that maps to the class g_x of g in $\mathcal{O}_{Y,x}$. We have that s_x is the class of a regular function f_V defined on a neighbourhood V of x in Z. The function $(f_V)\iota$ on the neighbourhood V of x considered in Y maps to g_x in $\mathcal{O}_{Y,x}$. Consequently, we have that g and $f_V\iota$ are equal in a neighbourhood W of x in Y. Hence $f_W = f_V|_W$ maps to $g|_W$ by the map

$$\mathcal{O}_Z(W) \to \mathcal{O}_Y(W).$$

Since the latter map is injective we have that f_W is uniquely defined. Hence the elements f_W, for each point x in X, define a function f in $\mathcal{O}_Z(Z) = \mathbf{K}[Z]$, that maps to g, and we have proved the proposition. □

5.6.5. A topological space can have several structures as a prevariety. We shall show that a morphism $\Phi\colon Y \to X$ of prevarieties which is a homeomorphism of topological spaces is not necessarily an isomorphism of prevarieties.

Example 5.6.6. Let $\mathbf{K} = \overline{\mathbf{K}}$ and assume that $2 = 0$ in $\mathbf{K}$. Moreover let $\Phi\colon \mathbf{A}^1_{\overline{\mathbf{K}}} \to \mathbf{A}^1_{\overline{\mathbf{K}}}$ be the map defined by $\Phi(a) = a^2$. This map is clearly a morphism. As the field $\mathbf{K}$ contains square roots of all of its elements it is onto, and it is injective because, if $\Phi(a) = \Phi(b)$, then $0 = a^2 - b^2 = (a-b)^2$, since $2 = 0$, and hence $a = b$. The map is a homeomorphism because, it sends finite sets to finite sets, and the open sets are the complements of finite sets (see Example 5.1.11). However, it is not an isomorphism because the corresponding map of coordinate rings $\mathbf{K}[x_1] \to \mathbf{K}[x_1]$ maps x_1 to x_1^2, and therefore is not surjective.

5.7 The tangent space of prevarieties

The tangent spaces of prevarieties are introduced in analogy with those for manifolds. They have similar properties and can be computed in the same way as the tangen spaces for manifolds.

Let X be a prevariety and x a point of X. We have an augmentation map 5.5.7.1 from the ring $\mathcal{O}_{X,x}$ of germs of regular functions at x to $\overline{\mathbf{K}}$, that maps a class $[(U, f)]$ to $f(x)$. Similarly, when X is an affine variety we have an augmentation map $\mathbf{K}[X] \to \overline{\mathbf{K}}$ as in 5.5.14.1 that maps f to $f(x)$.

Definition 5.7.1. The *tangent space* $T_x(X)$ of the prevariety X at the point x is the space of derivations

$$\delta\colon \mathcal{O}_{X,x} \to \overline{\mathbf{K}},$$

for the augmentation map at x.

Remark 5.7.2. The tangent space is a vector space over $\overline{\mathbf{K}}$, where addition $\delta + \varepsilon$ of two derivations δ and ε is given by $(\delta + \varepsilon)f = \delta f + \varepsilon f$, and multiplication $a\delta$ with an element a of $\overline{\mathbf{K}}$ is given by $(a\delta)f = a\delta(f)$.

Let U be an open subset of X containing x. The restriction $\mathcal{O}_{X,x} \to \mathcal{O}_{U,x}$ is an isomorphism. Consequently we have an isomorphism $T_x(U) \to T_x(X)$.

Let $\Phi\colon Y \to X$ be a morphism of prevarieties. From the natural map 5.5.8.1

$$\Phi_y^*\colon \mathcal{O}_{X,\Phi(y)} \to \mathcal{O}_{Y,y}$$

we obtain a map

$$T_y\Phi\colon T_y(Y) \to T_{\Phi(y)}(X),$$

for all y in Y, which maps the derivative $\delta_y^*\colon \mathcal{O}_{Y,y} \to \overline{\mathbf{K}}$ to the derivative $\delta\Phi\colon \mathcal{O}_{X,\Phi(y)} \to \overline{\mathbf{K}}$. When Y is a closed sub prevariety of X we have, by definition, that Φ_y^* is a surjection. Hence, if $\delta_y^*\Phi = 0$ we have that $\delta = 0$, and thus $T_y\Phi$ is injective.

5.8 Derivations

Before we show how to compute the tangent spaces of prevarieties we shall give some of the fundamental properties of derivations.

Recall (see 3.6.2) that given $\mathbf{K}$ algebras R and S, and a $\mathbf{K}$ algebra homomorphism $\varphi\colon R \to S$, we say that a $\mathbf{K}$ linear map

$$\delta\colon R \to S$$

is a derivation with respect to φ if

$$\delta(ab) = \varphi(a)\delta b + \varphi(b)\delta b, \tag{5.8.0.1}$$

for all a and b in R. The set $\mathrm{Der}_\varphi(R,S)$ of all derivations is a vector space over $\mathbf{K}$, with addition $\delta + \varepsilon$ of two derivations δ and ε given by $(\delta+\varepsilon)f = \delta f + \varepsilon f$, and multiplication $a\delta$ by an element a of $\mathbf{K}$ given by $(a\delta)f = a\delta f$.

Let T be a third $\mathbf{K}$ algebra, and let $\psi\colon S \to T$ be another $\mathbf{K}$ algebra homomorphism. Then we have a linear map

$$\mathrm{Der}_\psi(S,T) \to \mathrm{Der}_{\psi\varphi}(R,T), \tag{5.8.0.2}$$

which maps a derivation $\delta\colon S \to T$, for φ, to the derivation $\delta\varphi\colon R \to T$, for $\psi\varphi$. When φ is surjective we have that the map $\mathrm{Der}\,\varphi\colon \mathrm{Der}_\psi(S,T) \to \mathrm{Der}_{\psi\varphi}(R,T)$ is injective, because, if $\delta\varphi = 0$, then $\delta = 0$. When φ is surjective we identigy $\mathrm{Der}_\psi(S,T)$ with its image by 5.8.0.2. Then

$$\mathrm{Der}_\psi(S,T) = \{\delta \in \mathrm{Der}_{\psi\varphi}(R,T) : \delta a = 0, \quad \text{for all } a \in \ker\varphi\}.$$

Indeed, if δ in $\mathrm{Der}_\psi(S,T)$ then $\delta a = \delta\varphi(a) = 0$, for all a in $\ker\varphi$. Conversely, if δ in $\mathrm{Der}_{\psi\varphi}(R,T)$ and $\delta a = 0$, for all a in $\ker\varphi$, we can define a homomorphism $\varepsilon\colon S \to T$ by $\varepsilon b = \delta a$, for any a such that $\varphi(a) = b$. Indeed, if $\varphi(a_1) = \varphi(a_2) = b$, then $a_1 - a_2$ is in $\ker\varphi$, so $\delta(a_1 - a_2) = \delta a_1 - \delta a_2 = 0$, and consequently $\delta a_1 = \delta a_2$. It is clear that ε is a derivation for ψ and that it maps to δ by the map 5.8.0.2.

Assume that $\varphi : R \to S$ is surjective. Then $\mathrm{Der}_\psi(S,T)$ is determined by the value of the elements of $\mathrm{Der}_{\psi\varphi}(R,T)$ on the generators of $\ker\varphi$. In fact, let $(a_i)_{i\in I}$ be generators of $\mathrm{Der}\,\varphi$. For $\delta \in \mathrm{Der}_\psi(S,T)$ we have that $\delta a_i = 0$, for all $i \in I$. Conversely, if $\delta a_i = 0$, for all $i \in I$, and $a = b_1 a_{i_1} + \cdots + b_m a_{i_m}$ with $i_j \in I$ is in $\ker\varphi$, we have that $\delta a = \varphi(a_{i_1})\delta b_1 + \cdots + \varphi(a_{i_m})\delta b_m + \varphi(b_1)\delta a_{i_1} + \cdots + \varphi(b_m)\delta a_{i_m} = 0$, since $\varphi(a_{i_j}) = 0$ and $\delta a_{i_j} = 0$. Consequently, we have that

$$\mathrm{Der}_\psi(S,T) = \{\delta \in \mathrm{Der}_{\psi\varphi}(R,T) : \delta a_i = 0, \quad \text{for } i = 1,\dots,m\}.$$

We next show that $\mathbf{K}$-derivations on finitely generated $\mathbf{K}$-algebras are determined by their values on the generators. Let $\Psi\colon \mathbf{K}[a_1,\dots,a_n] \to R$ be a homomorphism of $\mathbf{K}$-algebras to a $\mathbf{K}$-algebra R from the $\mathbf{K}$-algebra $\mathbf{K}[a_1,\dots,a_n]$ generated by the elements $a_1,\dots,a_n$. A derivation

$$\delta\colon \mathbf{K}[a_1,\dots,a_n] \to R$$

is uniquely determined by the elements $\delta a_1,\dots,\delta a_n$. Indeed, since δ is $\mathbf{K}$ linear, we only have to show that $\delta(a_1^{i_1}\cdots a_n^{i_n})$ is determined by the elements $\delta a_1,\dots,\delta a_n$ for all monomials $a_1^{i_1}\cdots a_n^{i_n}$. However, this is clear since we by repeated use of the derivation rule 5.8.0.1 obtain that

$$\delta(a_1^{i_1}\cdots a_n^{i_n}) = \sum_{i_j\geq 1} i_j \Psi(a_1)^{i_1}\cdots \Psi(a_j)^{i_j-1}\cdots \Psi(a_n)^{i_n}\delta a_j.$$

5.9 Partial derivatives

Let $\varphi : \mathbf{K}[x_1,\dots,x_n] \to R$ be a $\mathbf{K}$-algebra homomorphism from the polynomial ring over $\mathbf{K}$ in the variables $x_1,\dots,x_n$ to the $\mathbf{K}$-algebra R. We denote by $\frac{\partial}{\partial x_i}$ the map

$$\frac{\partial}{\partial x_i}\colon \mathbf{K}[x_1,\dots,x_n] \to R,$$

defined by

$$\frac{\partial}{\partial x_j}(x_1^{i_1}\cdots x_n^{i_n}) = i_j\varphi(a_1)^{i_1}\cdots\varphi(a_j)^{i_j-1}\cdots\varphi(a_n)^{i_n},$$

if $i_j \geq 1$, and 0 otherwise. The reference to φ will be omitted below because the dependence of φ will be clear from the context. It is clear that $\frac{\partial}{\partial x_i}$ is a derivation. Let $\mathbf{K}[a_1,\dots,a_n]$ be a $\mathbf{K}$-algebra generated by the elements $a_1,\dots,a_n$ and let

$$\psi : \mathbf{K}[x_1,\dots,x_n] \to \mathbf{K}[a_1,\dots,a_n]$$

be the $\mathbf{K}$-algebra homomorphism uniquely defined by $\psi(x_i) = a_i$ for $i = 1,\dots,n$. We see that for any derivation $\delta\colon \mathbf{K}[a_1,\dots,a_n] \to R$ we have that

$$\delta\psi(f) = \sum_{i=1}^{n} \delta a_i \frac{\partial f}{\partial x_i},$$

for all f in $\mathbf{K}[x_1,\dots,x_n]$.

For the polynomial ring $\mathbf{K}[x_1,\dots,x_n]$ we obtain that all derivations δ can be written uniquely in the form

$$\delta = \sum_{i=1}^{n} \delta x_i \frac{\partial}{\partial x_i}.$$

Since ψ is surjective we can identify $\mathrm{Der}_\psi(\mathbf{K}[a_1,\dots,a_n], R)$ with the subset

$$\{\delta \in \mathrm{Der}_{\psi\varphi}(\mathbf{K}[x_1,\dots,x_n], R) : \delta a = 0 \text{ for all } a \in \ker\varphi\}.$$

With this identification have that

$$\mathrm{Der}_\varphi(\mathbf{K}[a_1,\dots,a_n],R)=\{b_1\frac{\partial}{\partial x_1}+\cdots+b_n\frac{\partial}{\partial x_n}: b_i\in R,\ \text{and}$$
$$b_1\frac{\partial f}{\partial x_1}+\cdots b_n\frac{\partial f}{\partial x_n}=0, \text{for all } f \text{ in a generator set of } \ker\varphi\}.$$

In particular, if $(c_1,\dots,c_n)$ is a point of $\mathbf{A}^n_{\overline{\mathbf{K}}}$ such that $f(c_1,\dots,c_n)=0$, for all f in $\ker\varphi$ a homomorphism $\eta\colon \mathbf{K}[a_1,\dots,a_n]\to\overline{\mathbf{K}}$, which maps the element $\psi f(x_1,\dots,x_n)$ to $f(c_1,\dots,c_n)$, , we have the augmentation map $\varphi\psi\colon \mathbf{K}[x_1,\dots,x_n]\to\overline{\mathbf{K}}$ that mapps $f(x_1,\dots,x_n)$ to $f(c_1,\dots,c_n)$, and. We obtain that

$$\mathrm{Der}_{\overline{\mathbf{K}}}(\mathbf{K}[a_1,\dots,a_n],\overline{\mathbf{K}})=\{b_1\frac{\partial}{\partial x_1}+\cdots+b_n\frac{\partial}{\partial x_n}: b_i\in\overline{\mathbf{K}},\ \text{and}$$
$$b_1\frac{\partial f}{\partial x_1}(c_1,\dots,c_n)+\cdots+b_n\frac{\partial f}{\partial x_n}(c_1,\dots,c_n)=0, \text{for all } f\in\ker\varphi\}.$$

Lemma 5.9.1. *Let $\varphi\colon R\to\overline{\mathbf{K}}$ be a* $\mathbf{K}$ *algebra homomorphism, and S a multiplicatively closed subset of R, such that $\varphi(a)\neq 0$, for all a in S. There exists a unique* $\mathbf{K}$ *algebra homomorphism $\psi\colon S^{-1}R\to\overline{\mathbf{K}}$ such that $\psi(\frac{a}{1})=\varphi(a)$, for all $a\in R$.*

Let $\delta\colon R\to\overline{\mathbf{K}}$ be a derivation for φ. Then there is a unique derivation $\varepsilon\colon S^{-1}R\to\overline{\mathbf{K}}$, for ψ such that $\varepsilon(\frac{a}{1})=\delta(a)$, for all $a\in R$.

Proof: We can define a map $\psi\colon S^{-1}R\to\overline{\mathbf{K}}$ by $\psi(\frac{a}{b})=\frac{\varphi(a)}{\varphi(b)}$, for all $a\in R$ and $b\in S$. Indeed, since $b\in S$, we have, by assumption, that $\varphi(b)\neq 0$, and, if $\frac{a}{b}=\frac{a'}{b'}$, there is a $c\in S$, such that $cab'=ca'b$. Hence $\varphi(c)\varphi(a)\varphi(b')=\varphi(c)\varphi(a')\varphi(b)$ in $\overline{\mathbf{K}}$, with $\varphi(c)\neq 0$. Thus $\frac{\varphi(a)}{\varphi(b)}=\frac{\varphi(a')}{\varphi(b')}$. Clearly we have that ψ is a $\mathbf{K}$ algebra homomorphism, and, by definition, $\psi(\frac{a}{1})=\varphi(a)$.

Similarly, we can define a derivation $\varepsilon\colon S^{-1}R\to\overline{\mathbf{K}}$ by $\varepsilon(\frac{a}{b})=\frac{\delta a}{\varphi(b)}-\frac{\varphi(a)}{\varphi(b)^2}\delta b$, for all $a\in\mathbf{K}$, and $b\in S$. Indeed, since $b\in S$ we have that $\varphi(b)\neq 0$, by assumption, and if $\frac{a}{b}=\frac{a'}{b'}$, there is a $c\in S$ such that $cab'=ca'b$. We obtain that $\varphi(ca)\delta b'+\varphi(cb')\delta a+\varphi(ab')\delta c=\varphi(ca')\delta b+\varphi(cb)\delta a'+\varphi(a'b)\delta c$. We divide by $\varphi(c)\varphi(b')\varphi(b)$ and obtain

$$\frac{\varphi(a)}{\varphi(b)}\frac{\delta b'}{\varphi(b')}+\frac{\delta a}{\varphi(b)}+\frac{\varphi(a)\delta c}{\varphi(b)\varphi(c)}=\frac{\varphi(a')\delta b}{\varphi(b')\varphi(b)}+\frac{\delta a'}{\varphi(b')}+\frac{\varphi(a')\delta c}{\varphi(b)\varphi(c)}.$$

Since $\frac{\varphi(a)}{\varphi(b)}=\frac{\varphi(a')}{\varphi(b')}$, we get $\varepsilon(\frac{a}{b})=\varepsilon(\frac{a'}{b'})$. It is clear that ε is a derivation for ψ and that $\varepsilon(\frac{a}{1}0=\delta(a)$ for all $a\in R$. □

Proposition 5.9.2. *Let X be an affine variety and x a point of X. Denote by $\varphi_x\colon \mathbf{K}[X]\to\overline{\mathbf{K}}$ the augmentation map. Then we have a canonical isomorphism*

$$\mathrm{Der}_{\varphi_x}(\mathbf{K}[X],\overline{\mathbf{K}})\to T_x(X).$$

Proof: It follows from Proposition 5.5.15 that we have an isomorphism $\mathbf{K}[X]_{\mathcal{M}_{X,x}} \to \mathcal{O}_{X,x}$, where $\mathcal{M}_{X,x}$ is the kernel of φ_x. The proposition is therefore a consequence of Lemma 5.9.1. □

Example 5.9.3. It follows from Proposition 5.9.2 that, for x in $\mathbf{A}^n_{\overline{\mathbf{K}}}$, we have that $T_x(\mathbf{A}^n_{\overline{\mathbf{K}}})$ is canonically isomophic to the n dimensional vector space of derivations $\mathbf{K}[x_1, \dots, x_n] \to \overline{\mathbf{K}}$, for the augmentation map $\varphi_x \colon \mathbf{K}[x_1, \dots, x_n] \to \overline{\mathbf{K}}$. As we saw in 5.7.2 we have a basis of this vector space consisting of the derivations

$$\frac{\partial}{\partial x_i} \colon \mathbf{K}[x_1, \dots, x_n] \to \overline{\mathbf{K}}, \quad \text{for } i = 1, \dots n,$$

where $\frac{\partial x_j}{\partial x_i}$ is 1 for $i = j$ and 0 otherwise.

Example 5.9.4. Let X be the subvariety $\mathcal{V}(x_2^2 - x_1^2 - x_1^3)$ of $\mathbf{A}^2_{\overline{\mathbf{K}}}$. The kernel of the canonical map $\mathbf{K}[x_1, x_2] \to \mathbf{K}[X]$ is the ideal generated by $f = x_2^2 - x_1^2 - x_1^3$. Indeed, the kernel $\mathcal{I}(X)$ contains f, and since, by Hilberts Nullstellensatz, we have that every g in $\mathcal{I}(X)$ can be written as $g^d = hf$, for some positive integer d and polynomial h in $\mathbf{K}[x_1, x_2]$. Since f can not be written as a product of two polynomials of positive degree less than 3 it is possible to show that we can take $d = 1$.

For $x = (a_1, a_2)$ in $\mathbf{A}^2_{\overline{\mathbf{K}}}$ we have that $T_x(X)$ is the subspace of the vector space with basis $\frac{\partial}{\partial x_1}$ and $\frac{\partial}{\partial x_2}$ consisting of derivations such that $a_1 \frac{\partial}{\partial x_1} + a_2 \frac{\partial}{\partial x_2} f = 2a_2 \frac{\partial}{\partial x_2} - 2a_1 \frac{\partial}{\partial x_1} - 3a_1^2 \frac{\partial}{\partial x_1} = 0$. If $x = (a_1, a_2) \neq (0, 0)$, this space has dimension one, spanned by $\frac{\partial}{\partial x_1}$ when $a_2 \neq 0$, and by $\frac{\partial}{\partial x_2}$ if $a_2 = 0$.

On the other hand, when $x = (0, 0)$, we have that $T_x(X)$ two dimensional and thus equal to $T_x(\mathbf{A}^2_{\overline{\mathbf{K}}})$.

We see from Example 5.9.4 that the tangent space to a prevariety can have different dimension at different points. A prevariety is therefore not a good analogue of manifolds. We shall later introduce smooth manifolds that will have properties similar to those of manifolds.

5.10 Tangent spaces for zeroes of polynomials

We shall in this section present the *epsilon calculus* for prevarieties. The treatment is analogous to that for manifolds in Section 3.7.

Let X be an affine variety in $\mathbf{A}^n_{\overline{\mathbf{K}}}$. Choose generators $f_1, \dots, f_m$ for the ideal $\mathcal{I}(X)$. We saw in Section 5.7 that, for all points x in X, the tangent space $T_x(X)$ is isomorphic to the subspace of the n dimensional space $T_x(\mathbf{A}^n_{\overline{\mathbf{K}}})$ with basis

$$\frac{\partial}{\partial x_i} \colon \mathbf{K}[x_1, \dots, x_n] \to \overline{\mathbf{K}}, \text{ for } i = 1, \dots, n,$$

consisting of vectors $\delta = a_1 \frac{\partial}{\partial x_1} + \cdots + a_n \frac{\partial}{\partial x_n}$ such that $\delta(f_i) = 0$ for $i = 1, \dots, m$.

Lemma 5.10.1. *Let x be a point of an affine variety X, and let $\varphi_x\colon \mathbf{K}[X] \to \overline{\mathbf{K}}$ be the augmentation map. The map*

$$\psi\colon\ \mathrm{Der}_{\overline{\mathbf{K}}}(\mathbf{K}[X], \overline{\mathbf{K}}) \to \mathrm{Hom}_{\varphi_x - alg}(\mathbf{K}[X], \overline{\mathbf{K}}[\varepsilon]),$$

such that $\varphi(\delta)(f) = f(x) + \delta f \varepsilon$ is a bijection from the derivaties for the evaluation map, to the $\mathbf{K}$-algebra homomorphisms $\zeta\colon \mathbf{K}[X] \to \overline{\mathbf{K}}[\varepsilon]$, into the ring $\overline{\mathbf{K}}[\varepsilon]$ of dual numbers that are of the form $\zeta(f) = f(x) + \delta_\zeta(f)\varepsilon$, for some map $\delta_\zeta\colon \mathbf{K}[X] \to \overline{\mathbf{K}}$.

Proof: Given a derivation $\delta\colon \mathbf{K}[X] \to \overline{\mathbf{K}}$, for the evaluation at x. The map $\zeta\colon \mathbf{K}[X] \to \overline{\mathbf{K}}[\varepsilon]$ defined by $\zeta(f) = f(x) + \delta(f)\varepsilon$ is clearly $\mathbf{K}$ linear, and it is a $\mathbf{K}$-algebra homomorphism because $\zeta(gf) = (fg)(x) + \delta(fg)\varepsilon = f(x)g(x) + (f(x)\delta g + g(x)\delta f)\varepsilon = (f(x) + \delta f\varepsilon)(g(x) + \delta g\varepsilon) = \zeta(f)\zeta(g)$.

Conversely, given a $\mathbf{K}$ algebra homomorphism $\zeta\colon \mathbf{K}[X] \to \mathbf{K}[\varepsilon]$ such that $\zeta(f) = f(x) + \delta_\zeta(f)\varepsilon$. Then the map $\delta_\zeta\colon \mathbf{K}[X] \to \overline{\mathbf{K}}$ is clearly $\mathbf{K}$ linear and it is a derivation because $f(x)g(x) + \delta_\zeta(fg)\varepsilon = \zeta(fg) = \zeta(f)\zeta(g) = (f(x) + \delta_\zeta(f)\varepsilon)(g(x) + \delta_\zeta(g)\varepsilon) = f(x)g(x) + (f(x)\delta_\zeta(g) + g(x)\delta_\zeta(g)\varepsilon) = f(x)g(x) + (f(x)\delta_\zeta(g) + g(x)\delta_\zeta(f))\varepsilon$, and thus $\delta_\zeta(fg) = f(x)\delta_\zeta(g) + g(x)\delta_\zeta(f)$. □

Remark 5.10.2. A $\mathbf{K}$ algebra homomorphism $\varphi\colon \mathbf{K}[x_1, \dots, x_n] \to \overline{\mathbf{K}}[\varepsilon]$ such that $\varphi(f) = f(x) + \delta_\varphi(f)\varepsilon$ is completely determined by the values $\varphi(x_i) = a_i + b_i\varepsilon$, for $i = 1, \dots, n$, where $x = (a_1, \dots, a_n)$ and $v = (b_1, \dots, b_n)$ are in $\mathbf{A}^n_{\overline{\mathbf{K}}}$, as we have seen in Remark 5.7.2. With this notation we have that $\varphi(f) = f(x + \varepsilon v)$. It follows from the binomial formula that

$$(a_1 + \varepsilon b_1)^{i_1} \cdots (a_n + \varepsilon b_n)^{i_n}$$
$$= a_1^{i_1} \cdots a_n^{i_n} + \sum_{i_j \neq 1} i_j a_1^{i_1} \cdots a_j^{i_j - 1} \cdots a_n^{i_n} b_j \varepsilon = a_1^{i_1} \cdots a_n^{i_n} + \sum_{j=1}^{n} b_j \frac{\partial(x_1^{i_1} \cdots x_n^{i_n})}{\partial x_j} \varepsilon.$$

Hence we obtain, for all f in $\mathbf{K}[x_1, \dots, x_n]$ that, for all $v = (v_1, \dots, v_n) \in \mathbf{A}^n_{\overline{\mathbf{K}}}$ we have that

$$f(x + \varepsilon v) = f(x) + \sum_{j=1}^{n} b_j \frac{\partial f}{\partial x_j} \varepsilon.$$

It follows from Remark 5.7.2 that

$$T_x(X) = \{v \in \mathbf{A}^n_{\overline{\mathbf{K}}} : f(x + \varepsilon v) = f(x) = 0, \quad \text{for } f \in \mathcal{I}(X)\}.$$

Example 5.10.3. We have that $T_{I_n}(\mathrm{Gl}_n(\mathbf{K})) = T_{I_n}(\mathrm{M}_n(\mathbf{K}))$, and thus $T_{I_n}(\mathrm{Gl}_n(\mathbf{K})) = \mathbf{A}^{n^2}_{\overline{\mathbf{K}}}$.

Example 5.10.4. We have already seen, in Example 5.10.3, that the tangent space of $\mathrm{Gl}_n(\mathbf{K})$ at I_n is equal to $\mathrm{M}_n(\mathbf{K})$. To find the tangent space of $\mathrm{Sl}_n(\mathbf{K})$ at I_n we use that $\mathrm{Sl}_n(\mathbf{K})$ is the subset of $\mathrm{Gl}_n(\mathbf{K})$ defined by the polynomial $\det(x_{i,j})$ of degree n in the n^2 variables $x_{i,j}$, for $i, j = 1, \dots, n$. Consequently, the tangent space $T_{I_n}(\mathrm{Sl}_n(\mathbf{K}))$ of $\mathrm{Sl}_n(\mathbf{K})$ at the unity I_n is equal to

$$\{A \in \mathrm{M}_n(\mathbf{K})\colon\ \det(I_n + \varepsilon A) - \det I_n = 0\}.$$

A short calculation shows that $\det(I_n + \varepsilon A) = 1 + \sum_{i=1}^n a_{i,i}\varepsilon$ (see Problem 3.7.2). Consequently, we have that

$$T_{I_n}(\mathrm{Sl}_n(\mathbf{K})) = \{(a_{i,j}) \in \mathrm{M}_n(\mathbf{K})\colon \sum_{i=1}^n a_{i,i} = 0\}.$$

In particular $a_{ii} = 0$ for $i = 1, \dots, n$ since $2 \neq 0$ in $\mathbf{K}$. That is, $T_{I_n}(\mathrm{Sl}_n(\mathbf{K}))$ consists of all matrices of *trace* equal to zero. In particular we have that the tangent space, and hence $\mathrm{Sl}_n(\mathbf{K})$ both have dimension $n^2 - 1$ (see Problem 2.5.4).

Example 5.10.5. Assume that $2 \neq 0$ in $\mathbf{K}$. The group $\mathrm{O}_n(\mathbf{K})$ is the subset of $\mathrm{Gl}_n(\mathbf{K})$ defined by the n^2 polynomials, in n^2 variables, that are the coefficients in the matrix $X^tX - I_n$. Consequently, the tangent space $T_{I_n}(\mathrm{O}_n(\mathbf{K}))$ is equal to

$$\{A \in \mathrm{M}_n(\mathbf{K})\colon (I_n + A\varepsilon)^t(I_n + A\varepsilon) - I_n = 0\}.$$

We have that $(I_n + A\varepsilon)^t(I_n + A\varepsilon) - I_n = (I_n + A\varepsilon)({}^tI_n + {}^tA\varepsilon) - I_n = I_n + A\varepsilon + {}^tA\varepsilon - I_n = (A + {}^tA)\varepsilon$. Consequently,

$$T_{I_n}(\mathrm{O}_n(\mathbf{K})) = \{A \in \mathrm{M}_n(\mathbf{K})\colon A + {}^tA = 0\}.$$

That is, $T_{I_n}(\mathrm{O}_n(\mathbf{K}))$ consists of all *antisymmetric* matrices. In particular, we have that the tangent space, and hence $\mathrm{O}_n(\mathbf{K})$ both have dimension $\frac{n(n-1)}{2}$ (see Exercise 2.5.5).

The subspace $\mathrm{SO}_n(\mathbf{K})$ is defined in $\mathrm{M}_n(\mathbf{K})$ by the same equations as $\mathrm{O}_n(\mathbf{K})$ plus the equation $\det(x_{i,j}) - 1 = 0$. As in Example 3.7.1 we see that this gives the condition that the matrices of $T_{I_n}(\mathrm{SO}_n(\mathbf{K}))$ have trace 0. Consequently, we have that $T_{I_n}(\mathrm{SO}_n(\mathbf{K})) = T_{I_n}(\mathrm{O}_n(\mathbf{K}))$, and the dimension of $\mathrm{SO}_n(\mathbf{K})$ is $\frac{n(n-1)}{2}$.

Example 5.10.6. The symplectic group $\mathrm{Sp}_n(\mathbf{K})$ is the subset of $\mathrm{M}_n(\mathbf{K})$ of common zeroes of the n^2 polynomials in n^2 variables that are the coefficients in the matrix $XS^tX - S$. We obtain that the tangent space $T_{I_n}(\mathrm{Sp}_n(\mathbf{K}))$ of $\mathrm{Sp}_n(\mathbf{K})$ in I_n is

$$\{A \in \mathrm{M}_n(\mathbf{K})\colon (I_n + A\varepsilon)S^t(I_n + A\varepsilon) = S\}.$$

We have that $(I_n + A\varepsilon)S^t(I_n + A\varepsilon) - S = S + AS\varepsilon + S^tA\varepsilon - S$. Consequently, we have that

$$T_{I_n}(\mathrm{Sp}_n(\mathbf{K})) = \{A \in \mathrm{M}_n(\mathbf{K}))\colon AS + S^tA = 0\}.$$

However $AS + S^tA = AS - {}^tS^tA = AS - {}^t(AS)$. Consequently, the isomorphism of vector spaces $\mathrm{M}_n(\mathbf{K}) \to \mathrm{M}_n(\mathbf{K})$, which sends a matrix A to AS (see Problem 2.5.6), maps $T_{I_n}(\mathrm{Sp}_n(\mathbf{K}))$ isomorphically onto the subspace of $\mathrm{M}_n(\mathbf{K})$ consisting of symmetric matrices. In particular the tangent space, and the space $\mathrm{Sp}_n(\mathbf{K})$, both have dimension $\frac{n(n+1)}{2}$ (see Problem 2.5.7).

References

[1] P.M. Cohn, *Lie groups*, Cambridge tracts in mathematics and mathematical physics, vol. 46, Cambridge University Press, Cambridge 1961.
[2] M.L. Curtis, *Matrix groups*, Universitexts, (second edition), Springer Verlag, Berlin, New York, 1984.
[3] J. Dieudonné, *Sur les groupes classiques*, Actualités scientifiques et industrielles, vol. 1040, Herman, 1958.
[4] P. Grittiths and J. Harris, *Principles of algebraic geometry*, Wiley, New York, 1978.
[5] S. Lang, *Algebra*, Advanced Book Program (third edition), Addison Wesley Pub. Co., Menlo Park, Calif., 1993.
[6] J-P. Serre, *Lie algebras and Lie groups*, W.A. Benjamin, Inc., Amsterdam, New York, 1965.
[7] F.W. Warner, *Foundation of differentiable manifolds and Lie groups*, Graduate texts in mathematics, Springer Verlag, Berlin, New York, 1983.

Index

Department of Mathematics, KTH, S–100 44 Stockholm, Sweden
E-mail address, Mats Boij: `boij@@math.kth.se`

Department of Mathematics, KTH, S–100 44 Stockholm, Sweden
E-mail address, Dan Laksov: `laksov@@math.kth.se`