Wiley Classics Library

ARTIN

Geometric Algebra

This page intentionally left blank

GEOMETRIC ALGEBRA

E. ARTIN

Princeton University, Princeton, New Jersey

Wiley Classics Library Edition Published 1988

INTERSCIENCE PUBLISHERS, INC., NEW YORK

a division of John Wiley & Sons, Inc., New York · London · Sydney

15 14 13 12 11

LIBRARY OF CONGRESS CATALOG CARD NUMBER 57-6109

ISBN 0 470 03432 7

ISBN 0-471-60839-4 (pbk.).

TO NATASCHA

PREFACE

Many parts of classical geometry have developed into great independent theories. Linear algebra, topology, differential and algebraic geometry are the indispensable tools of the mathematician of our time. It is frequently desirable to devise a course of geometric nature which is distinct from these great lines of thought and which can be presented to beginning graduate students or even to advanced undergraduates. The present book has grown out of lecture notes for a course of this nature given at New York University in 1955. This course centered around the foundations of affine geometry, the geometry of quadratic forms and the structure of the general linear group. I felt it necessary to enlarge the content of these notes by including projective and symplectic geometry and also the structure of the symplectic and orthogonal groups. Lack of space forced me to exclude unitary geometry and the quadratic forms of characteristic 2.

I have to thank in the first place my wife who helped greatly with the preparation of the manuscript and with the proofs. My thanks go also to George Bachman who with the help of Bernard Sohmer wrote the notes for the original course, to Larkin Joyner who drew the figures, and to Susan Hahn for helping with the proofreading.

E. ARTIN

SUGGESTIONS FOR THE USE OF THIS BOOK

The most important point to keep in mind is the fact that Chapter I should be used mainly as a reference chapter for the proofs of certain isolated algebraic theorems. These proofs have been collected so as not to interrupt the main line of thought in later chapters.

An inexperienced reader should start right away with Chapter II. He will be able to understand for quite a while, provided he knows the definition of a field and the rudiments of group theory. More knowledge will be required from §8 on and he may then work his way through the first three paragraphs and the beginning of §9 of Chapter I. This will enable him to read the rest of Chapter II except for a few harder algebraic theorems which he should skip in a first reading.

This skipping is another important point. It should be done whenever a proof seems too hard or whenever a theorem or a whole paragraph does not appeal to the reader. In most cases he will be able to go on and later on he may return to the parts which were skipped.

The rest of the book definitely presupposes a good knowledge of §4 of Chapter I[1]. The content of this paragraph is of such a fundamental importance for most of modern mathematics that every effort should be devoted to its mastery. In order to facilitate this, the content of §4 is illustrated in §5 by the theory of linear equations and §6 suggests an exercise on which the reader can test his understanding of the preceding paragraphs. If he can do this exercise then he should be well equipped for the remainder of the book.

Chapter III gives the theory of quadratic and of skew symmetric bilinear forms in a geometric language. For a first reading the symplectic geometry may be disregarded.

Chapter IV is almost independent of the preceding chapters. If the reader does not find the going too heavy he may start the book with Chapter IV. But §4 of Chapter I will be needed.

Chapter V connects, so to speak, the ideas of Chapters III and

---

[1]It is sufficient to know it for finite dimensional spaces only.

and IV. The problems of Chapter IV are investigated for the groups introduced in Chapter III.

Any one of these chapters contains too much material for an advanced undergraduate course or seminar. I could make the following suggestions for the content of such courses.

1) The easier parts of Chapter II.

2) The linear algebra of the first five paragraphs of Chapter I followed by selected topics from Chapter III, either on orthogonal or on symplectic geometry.

3) The fundamental theorem of projective geometry, followed by some parts of Chapter IV.

4) Chapter III, but with the following modification:

All that is needed from §4 of Chapter I is the statement:

If W^* is the space orthogonal to a subspace W of a non-singular space V then $\dim W + \dim W^* = \dim V$. This statement could be obtained from the naive theory of linear equations and the instructor could supply a proof of it. Our statement implies then $W^{**} = W$ and no further reference to §4 of Chapter I is needed.

CONTENTS

Chapter IV

The General Linear Group

Chapter V

The Structure of Symplectic and Orthogonal Groups

CHAPTER 1

Preliminary Notions

1. *Notions of set theory*

We begin with a list of the customary symbols:

$a \in S$	means a is an element of the set S.
$S \subset T$	means S is a subset of T.
$S \cap T$	means the intersection of the sets S and T; should it be empty we call the sets disjoint.
$S \cup T$	stands for the union of S and T.

$\bigcap_i S_i$ and $\bigcup_i S_i$ stand for intersection and union of a family of indexed sets. Should S_i and S_j be disjoint for $i \neq j$ we call $\bigcup_i S_i$ a disjoint union of sets. Sets are sometimes defined by a symbol $\{\cdots\}$ where the elements are enumerated between the parenthesis or by a symbol $\{x|A\}$ where A is a property required of x; this symbol is read: "the set of all x with the property A". Thus, for example:

$$S \cap T = \{x|x \in S, x \in T\}.$$

If f is a map of a non-empty set S into a set T, i.e., a function $f(s)$ defined for all elements $s \in S$ with values in T, then we write either

$$f : S \to T \quad \text{or} \quad S \xrightarrow{f} T.$$

If $S \xrightarrow{f} T$ and $T \xrightarrow{g} U$ we also write $S \xrightarrow{f} T \xrightarrow{g} U$. If $s \in S$ then we can form $g(f(s)) \in U$ and thus obtain a map from S to U denoted by $S \xrightarrow{gf} U$. Notice that the associative law holds trivially for these "products" of maps. The order of the two factors gf comes from the notation $f(s)$ for the image of the elements. Had we written $(s)f$ instead of $f(s)$, it would have been natural to write fg instead of gf. Although we will stick (with rare exceptions) to the notation $f(s)$ the reader should be able to do everything in the reversed notation. Sometimes it is even convenient to write s^f instead of $f(s)$ and we should notice that in this notation $(s^f)^g = s^{gf}$.

If $S \xrightarrow{f} T$ and $S_0 \subset S$ then the set of all images of elements of S_0 is denoted by $f(S_0)$; it is called the image of S_0 . This can be done particularly for S itself. Then $f(S) \subset T$; should $f(S) = T$ we call the map *onto* and say that f maps S onto T.

Let T_0 be a subset of T. The set of all $s \in S$ for which $f(s) \in T_0$ is called the inverse image of T_0 and is denoted by $f^{-1}(T_0)$. Notice that $f^{-1}(T_0)$ may very well be empty, even if T_0 is not empty. Remember also that f^{-1} is *not* a map. By $f^{-1}(t)$ for a certain $t \in T$ we mean the inverse image of the set $\{t\}$ with the one element t. It may happen that $f^{-1}(t)$ never contains more than one element. Then we say that f is a one-to-one *into* map. If f is onto and one-to-one into, then we say that f is one-to-one onto, or a "*one-to-one correspondence.*" In this case only can f^{-1} be interpreted as a map $T \xrightarrow{f^{-1}} S$ and is also one-to-one onto. Notice that $f^{-1}f : S \to S$ and $ff^{-1} : T \to T$ and that both maps are identity maps on S respectively T.

If $t_1 \neq t_2$ are elements of T, then the sets $f^{-1}(t_1)$ and $f^{-1}(t_2)$ are disjoint. If s is a given element of S and $f(s) = t$, then s will be in $f^{-1}(t)$, which shows that S is the disjoint union of all the sets $f^{-1}(t)$:

$$S = \bigcup_{t \in T} f^{-1}(t).$$

Some of the sets $f^{-1}(t)$ may be empty. Keep only the non-empty ones and call S_f the set whose elements are these non-empty sets $f^{-1}(t)$. Notice that the elements of S_f are *sets* and not elements of S. S_f is called a quotient set and its elements are also called equivalence classes. Thus, s_1 and s_2 are in the same equivalence class if and only if $f(s_1) = f(s_2)$. Any given element s lies in precisely one equivalence class; if $f(s) = t$, then the equivalence class of s is $f^{-1}(t)$.

We construct now a map $f_1 : S \to S_f$ by mapping each $s \in S$ onto its equivalence class. Thus, if $f(s) = t$, then $f_1(s) = f^{-1}(t)$. This map is an onto map.

Next we construct a map $f_2 : S_f \to f(S)$ by mapping the non-empty equivalence class $f^{-1}(t)$ onto the element $t \in f(S)$. If $t \in f(S)$, hence $t = f(s)$, then t is the image of the equivalence class $f^{-1}(t)$ and of no other. This map f_2 is therefore one-to-one and onto. If $s \in S$ and $f(s) = t$, then $f_1(s) = f^{-1}(t)$ and the image of $f^{-1}(t)$ under the map f_2 is t. Therefore, $f_2 f_1(s) = t$.

Finally we construct a very trivial map $f_3 : f(S) \to T$ by setting $f_3(t) = t$ for $t \in f(S)$. This map should not be called identity since

it is a map of a subset into a possibly bigger set T. A map of this kind is called an injection and is of course one-to-one into. For $f(s) = t$ we had $f_2 f_1(s) = t$ and thus $f_3 f_2 f_1(s) = t$. We have $S \xrightarrow{f_1} S_f \xrightarrow{f_2} f(S) \xrightarrow{f_3} T$, so that $f_3 f_2 f_1 : S \to T$. We see that our original map f is factored into three maps

$$f = f_3 f_2 f_1 .$$

To repeat: f_1 is onto, f_2 is a one-to-one correspondence and f_3 is one-to-one into. We will call this the canonical factoring of the map f. The word "*canonical*," or also "*natural*," is applied in a rather loose sense to any mathematical construction which is unique in as much as no free choices of objects are used in it.

As an example, let G and H be groups, and $f : G \to H$ a homomorphism of G into H, i.e., a map for which $f(xy) = f(x)f(y)$ holds for all $x, y \in G$. Setting $x = y = 1$ (unit of G) we obtain $f(1) = 1$ (unit in H). Putting $y = x^{-1}$, we obtain next $f(x^{-1}) = (f(x))^{-1}$. We will now describe the canonical factoring of f and must to this effect first find the quotient set G_f. The elements x and y are in the same equivalence class if and only if $f(x) = f(y)$ or $f(xy^{-1}) = 1$ or also $f(y^{-1}x) = 1$; denoting by K the inverse image of 1 this means that both $xy^{-1} \in K$ and $y^{-1}x \in K$ (or $x \in Ky$ and $x \in yK$). The two cosets yK and Ky are therefore the same and the elements x which are equivalent to y form the coset yK. If we take y already in K, hence y in the equivalence class of 1 we obtain $yK = K$, so that K is a group. The equality of left and right cosets implies that K is an invariant subgroup and our quotient set merely the factor group G/K. The map f_1 associates with each $x \in G$ the coset xK as image: $f_1(x) = xK$. The point now is that f_1 is a homomorphism (onto). Indeed $f_1(xy) = xyK = xyK \cdot K = x \cdot Ky \cdot K = xK \cdot yK = f_1(x)f_1(y)$.

This map is called the *canonical* homomorphism of a group onto its factor group.

The map f_2 maps xK onto $f(x) : f_2(xK) = f(x)$. Since $f_2(xK \cdot yK) = f_2(xy \cdot K) = f(xy) = f(x)f(y) = f_2(xK)f_2(yK)$ it is a homomorphism. Since it is a one-to-one correspondence it is an isomorphism and yields the statement that the factor group G/K is isomorphic to the image group $f(G)$. The invariant subgroup K of G is called the kernel of the map f.

The map f_3 is just an injection and therefore an isomorphism into H.

2. *Theorems on vector spaces*

We shall assume that the reader is familiar with the notion and the most elementary properties of a vector space but shall repeat its definition and discuss some aspects with which he may not have come into contact.

DEFINITION 1.1. A right vector space V over a field k (k need not be a commutative field) is an additive group together with a composition Aa of an element $A \in V$ and an element $a \in k$ such that $Aa \in V$ and such that the following rules hold:

1) $(A + B)a = Aa + Ba$, 2) $A(a + b) = Aa + Ab$,

3) $(Aa)b = A(ab)$, 4) $A \cdot 1 = A$,

where $A, B \in V$, $a, b \in k$ and where 1 is the unit element of k.

In case of a left vector space the composition is written aA and similar laws are supposed to hold.

Let V be a right vector space over k and S an arbitrary subset of V. By a linear combination of elements of S one means a finite sum $A_1a_1 + A_2a_2 + \cdots + A_ra_r$ of elements A_i of S. It is easy to see that the set $\langle S \rangle$ of all linear combinations of elements of S forms a subspace of V and that $\langle S \rangle$ is the smallest subspace of V which contains S. If S is the empty set we mean by $\langle S \rangle$ the smallest subspace of V which contains S and, since 0 is in any subspace, the space $\langle S \rangle$ consists of the zero vector alone. This subspace is also denoted by 0.

We call $\langle S \rangle$ the space generated (or spanned) by S and say that S is a system of generators of $\langle S \rangle$.

A subset S is called independent if a linear combination $A_1a_1 + A_2a_2 + \cdots + A_ra_r$ of distinct elements of S is the zero vector only in the case when all $a_i = 0$. The empty set is therefore independent.

If S is independent and $\langle S \rangle = V$ then S is called a basis of V. This means that every vector of V is a linear combination of distinct elements of S and that such an expression is unique up to trivial terms $A \cdot 0$.

If T is independent and L is any system of generators of V then T can be "completed" to a basis of V by elements of L. This means that there exists a subset L_0 of L which is disjoint from T such that the set $T \cup L_0$ is a basis of V. The reader certainly knows this

statement, at least when V is finite dimensional. The proof for the infinite dimensional case necessitates a transfinite axiom such as Zorn's lemma but a reader who is not familiar with it may restrict all the following considerations to the finite dimensional case.

If V has as basis a finite set S, then the number n of elements of S ($n = 0$ if S is empty) depends only on V and is called the dimension of V. We write $n = \dim V$. This number n is then the maximal number of independent elements of V and any independent set T with n elements is a basis of V. If U is a subspace of V, then $\dim U \leq \dim V$ and the equal sign holds only for $U = V$.

The fact that V does not have such a finite basis is denoted by writing $\dim V = \infty$. A proper subspace U of V may then still have the dimension ∞. (One could introduce a more refined definition of $\dim V$, namely the cardinality of a basis. We shall not use it, however, and warn the reader that certain statements we are going to make would not be true with this refined definition of dimension.)

The simplest example of an n-dimensional space is the set of all n-tuples of elements of k with the following definitions for sum and product:

$$(x_1, x_2, \cdots, x_n) + (y_1, y_2, \cdots, y_n) = (x_1 + y_1, \cdots, x_n + y_n),$$

$$(x_1, x_2, \cdots, x_n)a = (x_1a, x_2a, \cdots, x_na).$$

If U and W are subspaces of V (an arbitrary space), then the space spanned by $U \cup W$ is denoted by $U + W$. Since a linear combination of elements of U is again an element of U we see that $U + W$ consists of all vectors of the form $A + B$ where $A \in U$ and $B \in W$. The two spaces U and W may be of such a nature that an element $U + W$ is *uniquely* expressed in the form $A + B$ with $A \in U$, $B \in W$. One sees that this is the case if and only if $U \cap W = 0$. We say then that the sum $U + W$ is *direct* and use the symbol $U \oplus W$. Thus one can write $U \oplus W$ for $U + W$ if and only if $U \cap W = 0$.

If U_1, U_2, U_3 are subspaces and if we can write $(U_1 \oplus U_2) \oplus U_3$, then an expression $A_1 + A_2 + A_3$ with $A_i \in U_i$ is unique and thus one can also write $U_1 \oplus (U_2 \oplus U_3)$. We may therefore leave out the parenthesis: $U_1 \oplus U_2 \oplus U_3$. An intersection of subspaces is always a subspace.

Let U now be a subspace of V. We remember that V was an additive group. This allows us to consider the additive factor group V/U

whose elements are the cosets $A + U$. ($A + U$ for an arbitrary but *fixed* $A \in V$ means the set of all vectors of the form $A + B$, $B \in U$.) Equality $A_1 + U = A_2 + U$ of two cosets means $A_1 - A_2 \in U$, addition is explained by $(A_1 + U) + (A_2 + U) = (A_1 + A_2) + U$. We also have the canonical map

$$\varphi : V \to V/U$$

which maps $A \in V$ onto the coset $A + U$ containing A. The map φ is an additive homomorphism of V onto V/U. We make V/U into a vector space by defining the composition of an element $A + U$ of V/U and an element $a \in k$ by:

$$(A + U)\cdot a = Aa + U.$$

One has first to show that this composition is *well* defined, i.e., does not depend on the particular element A of the coset $A + U$. But if $A + U = B + U$, then $A - B \in U$, hence $(A - B)a \in U$ which shows $Aa + U = Ba + U$. That the formal laws of Definition 1.1 are satisfied is pretty obvious. For the canonical map φ we have

$$\varphi(Aa) = Aa + U = (A + U)\cdot a = \varphi(A)\cdot a$$

in addition to the fact that φ is an additive homomorphism. This suggests

DEFINITION 1.2. Let V and W be two right vector spaces (W not necessarily a subspace of V) over k. A map $f : V \to W$ is called a homomorphism of V into W if

1) $f(A + B) = f(A) + f(B)$, $\quad A \in V$ and $B \in V$,

2) $f(Aa) = f(A)\cdot a$, $\quad A \in V$ and $a \in k$.

Should f be a one-to-one correspondence, we call f an isomorphism of V onto W and we denote the mere existence of such an isomorphism by $V \simeq W$ (read: "V isomorphic to W").

Notice that such a homomorphism is certainly a homomorphism of the additive group. The notion of kernel U of f is therefore already defined, $U = f^{-1}(0)$, the set of all $A \in V$ for which $f(A) = 0$. If $A \in U$ then $f(Aa) = f(A)\cdot a = 0$ so that $Aa \in U$. This shows that U is not only a subgroup but even a subspace of V.

Let U be an arbitrary subspace of V and $\varphi : V \to V/U$ the canonical map. Then it is clear that φ is a homomorphism of V onto V/U.

The zero element of V/U is the image of 0, hence U itself. The kernel consists of all $A \in V$ for which

$$\varphi(A) = A + U = U.$$

It is therefore the given subspace U. One should mention the special case $U = 0$. Each coset $A + U$ is now the set with the single element A and may be identified with A. Strictly speaking we have only a canonical isomorphism $V/0 \simeq V$ but we shall write $V/0 = V$.

Let us return to any homomorphism $f : V \to W$ and let U be the kernel of V. Since f is a homomorphism of the additive groups we have already the canonical splitting

$$V \xrightarrow{f_1} V/U \xrightarrow{f_2} f(V) \xrightarrow{f_3} W$$

where $f_1(A) = A + U$ is the canonical map $V \to V/U$, where $f_2(A + U) = f(A)$ and, therefore,

$$f_2((A + U)a) = f_2(Aa + U) = f(Aa) = f(A)a = f_2(A + U)a$$

and where f_3 is the injection. All three maps are consequently homomorphisms between the vector spaces, and f_2 is an isomorphism onto. We have, therefore,

THEOREM 1.1. *To a given homomorphism $f : V \to W$ with kernel U we can construct a canonical isomorphism f_2 mapping V/U onto the image space $f(V)$.*

Suppose now that U and W are given subspaces of V. Let φ be the canonical map $V \xrightarrow{\varphi} V/U$. The restriction ψ of φ to the given subspace W is a canonically constructed homomorphism $W \xrightarrow{\psi} V/U$. What is $\psi(W)$? It consists of all cosets $A + U$ with $A \in W$. The union of these cosets forms the space $W + U$, the cosets $A + U$ are, therefore, the stratification of $W + U$ by cosets of the subspace U of $W + U$. This shows $\psi(W) = (U + W)/U$. What is the kernel of ψ? For all elements $A \in W$ we have $\psi(A) = \varphi(A)$. But φ has, in V, the kernel U so that ψ has $U \cap W$ as kernel. To ψ we can construct the canonical map ψ_2 which exhibits the isomorphism of $W/(U \cap W)$ with the image $(U + W)/U$. Since everything was canonical we have

THEOREM 1.2. *If U and W are subspaces of V then $(U + W)/U$ and $W/(U \cap W)$ are canonically isomorphic.*

In the special case $V = U \oplus W$ we find that V/U and $W/(U \cap W) = W/0 = W$ are canonically isomorphic. Suppose now that

only the subspace U of V is given. Does there exist a subspace W such that $V = U \oplus W$? Such a subspace shall be called supplementary to U. Let S be a basis of U and complete S to a basis $S \cup T$ of V where S and T are disjoint. Put $W = \langle T \rangle$, then $U + W = V$ and obviously $V = U \oplus W$. This construction involves choices and is far from being canonical.

THEOREM 1.3. *To every subspace U of V one can find (in a non-canonical way) supplementary spaces W for which $V = U \oplus W$. Each of these supplementary subspaces W is, however, canonically isomorphic to the space V/U. If $V = U \oplus W_1 = U \oplus W_2$ then W_1 is canonically isomorphic to W_2.*

If $f : V \to W$ is an isomorphism *into* then the image $f(S)$ of a basis of V will at least be independent. One concludes the inequality $\dim V \leq \dim W$. Should f be also onto then equality holds.

In our construction of W we also saw that $\dim V = \dim U + \dim W$ and since $W \simeq V/U$ one obtains

$$\dim V = \dim U + \dim V/U$$

hence also, whenever $V = U \oplus W$, that

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

Let now $U_1 \subset U_2 \subset U_3$ be subspaces of V. Find subspaces W_2 and W_3 such that

$$U_2 = U_1 \oplus W_2, \qquad U_3 = U_2 \oplus W_3$$

and, therefore,

$$U_3 = U_1 \oplus (W_2 \oplus W_3).$$

We have $\dim U_2/U_1 = \dim W_2$, $\dim U_3/U_2 = \dim W_3$ and $\dim U_3/U_1 = \dim(W_2 \oplus W_3) = \dim W_2 + \dim W_3$. Thus we have proved: if $U_1 \subset U_2 \subset U_3$, then

$$\dim U_3/U_1 = \dim U_2/U_1 + \dim U_3/U_2. \tag{1.1}$$

Let now U and W be two given subspaces of V. Use (1.1) for $U_1 = 0$, $U_2 = U$, $U_3 = U + W$. We obtain

$$\begin{aligned} \dim(U + W) &= \dim U + \dim(U + W)/U \\ &= \dim U + \dim W/(U \cap W). \end{aligned}$$

If we add on both sides $\dim(U \cap W)$ and use $\dim W/(U \cap W) + \dim(U \cap W) = \dim W$ we get

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

Next we use (1.1) for $U_1 = U \cap W$, $U_2 = W$, $U_3 = V$:

$$\begin{aligned}\dim V/(U \cap W) &= \dim W/(U \cap W) + \dim V/W \\ &= \dim (U + W)/U + \dim V/W.\end{aligned}$$

If we add $\dim V/(U + W)$ and use

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

we obtain

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

If the dimension of V is finite all subspaces of V have finite dimension. If, however, $\dim V = \infty$, then our interest will be concentrated on two types of subspaces U. Those whose dimension is finite and, on the other hand, those which are extremely large, namely those which have a finite dimensional supplement. For spaces of the second type $\dim U = \infty$ but $\dim V/U$ is finite; $\dim U$ tells us very little about U, but $\dim V/U$ gives us the amount by which U differs from the whole space V. We give, therefore, to $\dim V/U$ a formal status by

DEFINITION 1.3. The dimension of the space V/U is called the codimension of U:

$$\operatorname{codim} U = \dim V/U.$$

The various results we have obtained are expressed in

THEOREM 1.4. *The following rules hold between dimensions and codimensions of subspaces*:

$$\dim U + \operatorname{codim} U = \dim V, \tag{1.2}$$

$$\dim(U + W) + \dim(U \cap W) = \dim U + \dim W, \tag{1.3}$$

$$\operatorname{codim}(U + W) + \operatorname{codim}(U \cap W) = \operatorname{codim} U + \operatorname{codim} W. \tag{1.4}$$

These rules are of little value unless the terms on one side are finite (then those on the other side are also) since an ∞ could not be transposed to the other side by subtraction.

Spaces of dimension one are called lines, of dimension two planes and spaces of codimension one are called hyperplanes.

3. *More detailed structure of homomorphisms*

Let V and V' be right vector spaces over a field k and denote by $\mathrm{Hom}(V, V')$ the set of all homomorphisms of V into V'. We shall make $\mathrm{Hom}(V, V')$ into an abelian additive group by defining an addition:

If f and g are $\in \mathrm{Hom}(V, V')$, let $f + g$ be the map which sends the vector $X \in V$ onto the vector $f(X) + g(X)$ of V'; in other words,

$$(f + g)(X) = f(X) + g(X).$$

That $f + g$ is a homomorphism and that the addition is associative and commutative is easily checked. The map which sends every vector $X \in V$ onto the 0 vector of V' is obviously the 0 element of $\mathrm{Hom}(V, V')$ and shall also be denoted by 0. If $f \in \mathrm{Hom}(V, V')$, then the map $-f$ which sends X onto $-(f(X))$ is a homomorphism and indeed $f + (-f) = 0$. The group property is established.

In special situations it is possible to give more structure to $\mathrm{Hom}(V, V')$ and we are going to investigate some of the possibilities.

a) $V' = V$.

An element of $\mathrm{Hom}(V, V)$ maps V into V; one also calls it an endomorphism of V. If $f, g \in \mathrm{Hom}(V, V)$, then it is possible to combine them to a map $gf : V \xrightarrow{f} V \xrightarrow{g} V$ as we did in §1 : $gf(X) = g(f(X))$. One sees immediately that gf is also a homomorphism of $V \to V$.

Since

$$(g_1 + g_2)f(X) = g_1f(X) + g_2f(X) = (g_1f + g_2f)(X)$$

and

$$\begin{aligned} g(f_1 + f_2)(X) &= g(f_1(X) + f_2(X)) = gf_1(X) + gf_2(X) \\ &= (gf_1 + gf_2)X, \end{aligned}$$

we see that both distributive laws hold; $\mathrm{Hom}(V, V)$ now becomes a ring. This ring has a unit element, namely the identity map.

The maps f which are a one-to-one correspondence lead to an inverse map f^{-1} which is also in $\mathrm{Hom}(V, V)$. These maps f form

therefore a group under multiplication. All of Chapter IV is devoted to the study of this group if dim V is finite.

Let us now investigate some elementary properties of $\mathrm{Hom}(V, V)$ if dim $V = n$ is finite. Let $f \in \mathrm{Hom}(V, V)$ and let U be the kernel of f. Then $V/U \simeq f(V)$ so that the dimension of the image $f(V)$ is $n - \dim U$. This shows that f is an onto map if and only if dim $U = 0$, i.e., if and only if f is an isomorphism into.

Let $A_1, A_2, \cdots, A_n$ be a basis of V and set $f(A_i) = B_i$. If $X = A_1x_1 + A_2x_2 + \cdots + A_nx_n \in V$ then

$$f(X) = B_1x_1 + B_2x_2 + \cdots + B_nx_n . \tag{1.5}$$

Conversely choose any n vectors $B_i \in V$ and define a map f by (1.5). One sees easily that $f \in \mathrm{Hom}(V, V)$ and that $f(A_i) = B_i$. Consequently f is completely determined by the images B_i of the basis elements A_i and the B_i can be any system of n vectors of V. If we express each B_i by the basis A_ν,

$$f(A_j) = B_j = \sum_{\nu=1}^{n} A_\nu a_{\nu j}, \qquad j = 1, 2, \cdots, n,$$

then we see that f is described by an n-by-n matrix (a_{ij}) where i is the index of the rows and j the index of the columns.

Let $g \in \mathrm{Hom}(V, V)$ be given by the matrix (b_{ij}) which means that

$$g(A_j) = \sum_{\nu=1}^{n} A_\nu b_{\nu j} .$$

Then

$$(f + g)(A_j) = \sum_{\nu=1}^{n} A_\nu(a_{\nu j} + b_{\nu j})$$

and

$$\begin{aligned}(fg)(A_j) = f\Big(\sum_{\nu=1}^{n} A_\nu b_{\nu j}\Big) &= \sum_{\nu=1}^{n} f(A_\nu)b_{\nu j} \\ &= \sum_{\nu=1}^{n} \Big(\sum_{\mu=1}^{n} A_\mu a_{\mu\nu}\Big)b_{\nu j} \\ &= \sum_{\mu=1}^{n} A_\mu\Big(\sum_{\nu=1}^{n} a_{\mu\nu}b_{\nu j}\Big).\end{aligned}$$

We see that $f + g$ is described by the matrix $(a_{ij} + b_{ij})$ and fg by $(\sum_{r=1}^{n} a_{ir}b_{rj})$. This is the reason for defining addition and multiplication of matrices by

$$(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}),$$

$$(a_{ij})\cdot(b_{ij}) = \left(\sum_{r=1}^{n} a_{ir}b_{rj}\right).$$

Under this definition of addition and multiplication the correspondence $f \to (a_{ij})$ becomes an isomorphism between $\mathrm{Hom}(V, V)$ and the ring of all n-by-n matrices.

This isomorphism is far from canonical since it depends on the choice of the basis A_i for V.

Let g be another element of $\mathrm{Hom}(V, V)$, but suppose that g is one-to-one. Let (b_{ij}) be the matrix associated with the element gfg^{-1} of $\mathrm{Hom}(V, V)$. The meaning of the matrix (b_{ij}) is that

$$gfg^{-1}(A_j) = \sum_{r=1}^{n} A_r b_{rj} .$$

If we apply g^{-1} to this equation it becomes

$$f(g^{-1}(A_j)) = \sum_{r=1}^{n} g^{-1}(A_r)\cdot b_{rj} .$$

Since g^{-1} is any one-to-one onto map of V the vectors $g^{-1}(A_r)$ are another basis of V, and g can be chosen in such a way that $g^{-1}(A_r)$ is any given basis of V. Looking at the equation from this point of view we see that the matrix (b_{ij}) is the one which would describe f if we had chosen $g^{-1}(A_r)$ as basis of V. Therefore:

The matrix describing f in terms of the new basis is the same as the one describing gfg^{-1} in terms of the old basis A_r. In this statement g was the map which carries the "new basis" $g^{-1}(A_r)$ into the old one,

$$g(g^{-1}(A_r)) = A_r .$$

This g is, therefore, a fixed map once the new basis is given. Suppose now that $f \to A$, $g \to D$ are the descriptions of f and g in terms of the original basis. Then $gfg^{-1} \to DAD^{-1}$. The attitude should be that g is fixed, determined by the old and the new basis, and that f ranges over $\mathrm{Hom}(V, V)$. We can state