Princeton Un.
79 Cassville Rd.
Cantonville
21030

ADVANCED CALCULUS

by

H. K. NICKERSON
D. C. SPENCER
AND
N. E. STEENROD
Department of Mathematics
Princeton University

D. VAN NOSTRAND COMPANY, INC.
PRINCETON, NEW JERSEY
TORONTO LONDON
NEW YORK

D. VAN NOSTRAND COMPANY, INC.
120 Alexander St., Princeton, New Jersey (*Principal office*)
24 West 40th Street, New York 18, New York

D. VAN NOSTRAND COMPANY, LTD.
358, Kensington High Street, London, W.14, England

D. VAN NOSTRAND COMPANY (Canada), LTD.
25 Hollinger Road, Toronto 16, Canada

Published simultaneously in Canada by
D. VAN NOSTRAND COMPANY (Canada), LTD.

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

These notes were prepared for the honors course in Advanced Calculus, Mathematics 303-304, Princeton University.

The standard treatises on this subject, at any rate those available in English, tend to be omnibus collections of seemingly unrelated topics. The presentation of vector analysis often degenerates into a list of formulas and manipulative exercises, and the student is not brought to grips with the underlying mathematical ideas.

In these notes a unity is achieved by beginning with an abstract treatment of vector spaces and linear transformations. This enables us to introduce a single basic derivative (the Fréchet derivative) in an invariant form. All other derivatives (gradient, divergence, curl and exterior derivative) are obtained from it by specialization. The corresponding theory of integration is likewise unified, and the various multiple integral theorems of advanced calculus appear as special cases of a general Stokes' formula concerning the integration of exterior forms. In a final chapter these concepts are applied to analytic functions of complex variables.

The authors thank J. C. Moore for advice and suggestions concerning the writing of the manuscript, and J. J. Kohn and D. Scott for their assistance during the Summer, 1958. Acknowledgment is made to the D. Van Nostrand Company for their care in reproducing the notes, and to Mrs. E. Anthony and Miss M. Murray for the typing.

TABLE OF CONTENTS
and
2013 Supplementary Material

TABLE OF CONTENTS

I. THE ALGEBRA OF VECTOR SPACES

§1. Axioms

1.1. Definition. A vector space V is a set, whose elements are called vectors, together with two operations. The first operation, called addition, assigns to each pair of vectors A, B a vector, denoted by $A + B$, called their sum. The second operation, called multiplication by a scalar, assigns to each vector A and each real number x a vector denoted by xA. The two operations are required to have the following eight properties:

Axiom 1. $A + B = B + A$ for each pair of vectors A, B. (I.e. addition is commutative.)

Axiom 2. $(A + B) + C = A + (B + C)$ for each triple of vectors A, B, C. (I.e. addition is associative.)

Axiom 3. There is a unique vector $\tilde{\tilde{0}}$, called the zero vector, such that $\tilde{\tilde{0}} + A = A$ for each vector A.

Axiom 4. To each vector A there corresponds a unique vector, denoted by $-A$, such that $A + (-A) = \tilde{\tilde{0}}$.

Axiom 5. $x(A + B) = xA + xB$ for each real number x and each pair of vectors A, B. (I.e. multiplication is distributive with respect to vector addition.)

Axiom 6. $(x + y)A = xA + yA$ for each pair x, y of real numbers and each vector A. (I. e. multiplication is distributive with respect to scalar addition.)

Axiom 7. $(xy)A = x(yA)$ for each pair x, y of real numbers and each vector A.

Axiom 8. For each vector A,

$$\text{(i)} \qquad 0A = \tilde{0} \quad ,$$

$$\text{(ii)} \qquad 1A = A \quad ,$$

$$\text{(iii)} \qquad (-1)A = -A \quad .$$

1.2. Definition. The difference $A - B$ of two vectors is defined to be the sum $A + (-B)$.

The subsequent development of the theory of vector spaces will be based on the above axioms as our starting point. There are other approaches to the subject in which the vector spaces are constructed. For example, starting with a euclidean space, we could define a vector to be an oriented line segment. Or, again, we could define a vector to be a sequence $(x_1, \ldots, x_n)$ of n real numbers. These approaches give particular vector spaces having properties not possessed by all vector spaces. The advantages of the axiomatic approach are that the results which will be obtained apply to all vector spaces, and the axioms supply a firm starting point for a logical development.

§2. Redundancy

The axioms stated above are redundant. For example the word "unique" in Axiom 3 can be omitted. For suppose $\tilde{0}$ and $\tilde{0}'$ are two vectors satisfying $\tilde{0} + A = A$ and $\tilde{0}' + A = A$ for every A. In the first identity, take $A = \tilde{0}'$; and in the second, take $A = \tilde{0}$. Using Axiom 1, we obtain

$$\tilde{0}' = \tilde{0} + \tilde{0}' = \tilde{0}' + \tilde{0} = \tilde{0} \quad .$$

This proves the uniqueness.

The word "unique" can likewise be omitted from Axiom 4.

For suppose A, B, C are three vectors such that

$$A + B = \tilde{0}, \quad \text{and} \quad A + C = \tilde{0} \quad .$$

Using these relations and Axioms 1, 2 and 3, we obtain

$$B = \tilde{0} + B = (A + C) + B = A + (C + B) = A + (B + C) = (A + B) + C$$
$$= \tilde{0} + C = C \ .$$

Therefore $B = C$, and so there can be at most one candidate for $-A$.

The Axiom 8(i) is a consequence of the preceding axioms:

$$\tilde{0} = 0A + (-0A) = (0 + 0)A + (-0A)$$
$$= (0A + 0A) + (-0A) = 0A + (0A + (-0A))$$
$$= 0A + \tilde{0} = \tilde{0} + 0A = 0A.$$

§3. Cartesian spaces

3.1. Definition. The cartesian k-dimensional space, denoted by R^k, is the set of all sequences $(a_1, a_2, \ldots, a_k)$ of k real numbers together with the operations

$$(a_1, a_2, \ldots, a_k) + (b_1, b_2, \ldots, b_k) = (a_1+b_1, a_2+b_2, \ldots, a_k+b_k)$$

and

$$x(a_1, a_2, \ldots, a_k) = (xa_1, xa_2, \ldots, xa_k) \quad .$$

In particular, $R^1 = R$ is the set of real numbers with the usual addition and multiplication. The number a_i is called the ith component of $(a_1, a_2, \ldots, a_k)$, $i = 1, \ldots, k$.

3.2. Theorem. For each integer $k > 0$, R^k is a vector space.

Proof. The proofs of Axioms 1 through 8 are based on

the axioms for the real numbers R.

Let $A = (a_1, a_2, \ldots, a_k)$, $B = (b_1, b_2, \ldots, b_k)$, etc. For each $i = 1, \ldots, k$, the ith component of $A + B$ is $a_i + b_i$, and that of $B + A$ is $b_i + a_i$. Since the addition of real numbers is commutative, $a_i + b_i = b_i + a_i$. This implies $A + B = B + A$; hence Axiom 1 is true.

The ith component of $(A + B) + C$ is $(a_i + b_i) + c_i$, and that of $A + (B + C)$ is $a_i + (b_i + c_i)$. Thus the associative law for real numbers implies Axiom 2.

Let $\tilde{0} = (0, 0, \ldots, 0)$ be the sequence each of whose components is zero. Since $0 + a_i = a_i$, it follows that $\tilde{0} + A = A$. This proves Axiom 3 since the uniqueness part of the axiom is redundant (see §2).

If $A = (a_1, a_2, \ldots, a_k)$, define $-A$ to be $(-a_1, -a_2, \ldots, -a_k)$. Then $A + (-A) = \tilde{0}$. This proves Axiom 4 (uniqueness is again redundant).

If x is a real number, the ith component of $x(A + B)$ is, by definition, $x(a_i + b_i)$; and that of $xA + xB$ is, by definition, $xa_i + xb_i$. Thus the distributive law for real numbers implies Axiom 5.

The verifications of Axioms 6, 7 and 8 are similar and are left to the reader.

§4. Exercises

1. Verify that R^k satisfies Axioms 6, 7 and 8.

2. Prove that Axiom 8(iii) is redundant. Show also that $(-x)A = -(xA)$ for each x and each A.

3. Show that Axiom 8(ii) is not a consequence of the

preceding axioms by constructing a set with two operations which satisfy the preceding axioms but not 8(ii). (Hint: Consider the real numbers with multiplication redefined by $xy = 0$ for all x and y.) Can such an example satisfy Axiom 8(iii)?

4. Show that $A + A = 2A$ for each A.

5. Show that $x\tilde{0} = \tilde{0}$ for each x.

6. If $x \neq 0$ and $xA = \tilde{0}$, show that $A = \tilde{0}$.

7. If x and A are such that $xA = \tilde{0}$, show that either $x = 0$ or $A = \tilde{0}$.

8. Show that the set consisting of a single vector $\tilde{0}$ is a vector space.

9. If a vector A is such that $A = -A$, then $A = \tilde{0}$.

10. If a vector space contains some vector other than $\tilde{0}$, show that it contains infinitely many distinct vectors. (Hint: Consider A, $2A$, $3A$, etc.)

11. Let D be any non-empty set, and define R^D to be the set of all functions having domain D and values in R. If f and g are two such functions, their sum $f + g$ is the element of R^D defined by

$$(f + g)(d) = f(d) + g(d) \quad \text{for each } d \text{ in } D.$$

If f is in R^D and x is a real number, let xf be the element of R^D defined by

$$(xf)(d) = xf(d) \quad \text{for each } d \text{ in } D.$$

Show that R^D is a vector space with respect to these operations.

12. Let V be a vector space and let D be a non-empty set. Let V^D be the set of all functions with domain D

and values in V. Define sum and product as in Exercise 11, and show that V^D is a vector space.

13. A sum of four vectors $A + B + C + D$ may be associated (parenthesized) in five ways, e.g. $(A + (B + C)) + D$. Show that all five sums are equal, and therefore $A + B + C + D$ makes sense without parentheses.

14. Show that $A + B + C + D = B + D + C + A$.

§5. Associativity and commutativity

5.1. Proposition. If k is an integer ≥ 3, then any two ways of associating a sum $A_1 + \cdots + A_k$ of k vectors give the same sum. Consequently parentheses may be dropped in such sums.

Proof. The proof proceeds by induction on the number of vectors. Axiom 2 gives the case of 3 vectors. Suppose now that $k > 3$, and that the theorem is true for sums involving fewer than k vectors. We shall show that the sum of k vectors obtained from any method M of association equals the sum obtained from the standard association M_0 obtained by adding each term in order, thus:

$$(\ldots(((A_1 + A_2) + A_3) + A_4) \ldots) + A_k .$$

A method M must have a last addition in which, for some integer i with $1 \leq i < k$, a sum of $A_1 + \cdots + A_i$ is added to a sum of $A_{i+1} + \cdots + A_k$. If i is $k - 1$, the last addition has the form

$$(A_1 + \cdots + A_{k-1}) + A_k .$$

The part in parentheses has fewer than k terms and, by the inductive hypothesis, is equal to the sum obtained by the standard association on $k - 1$ terms. This converts the full sum to the

standard association on k terms. If $i = k - 2$, it has the form

$$(A_1 + \cdots + A_{k-2}) + (A_{k-1} + A_k)$$

which equals

$$((A_1 + \cdots + A_{k-2}) + A_{k-1}) + A_k$$

by Axiom 2 (treating $A_1 + \cdots + A_{k-2}$ as a single vector). By the inductive hypothesis, the sum of the first $k - 1$ terms is equal to the sum obtained from the standard association. This converts the full sum to the standard association on k terms. Finally, suppose $i < k - 2$. Since $A_{i+1} + \cdots + A_k$ has fewer than k terms, the inductive hypothesis asserts that its sum is equal to a sum of the form $(A_{i+1} + \cdots + A_{k-1}) + A_k$. The full sum has the form

$$\begin{aligned}(A_1 + \cdots + A_i) &+ ((A_{i+1} + \cdots + A_{k-1}) + A_k) \\ &= ((A_1 + \cdots + A_i) + (A_{i+1} + \cdots + A_{k-1})) + A_k\end{aligned}$$

by Axiom 2 applied to the three vectors $A_1 + \cdots + A_i$, $A_{i+1} + \cdots + A_{k-1}$ and A_k. The inductive hypothesis permits us to reassociate the sum of the first $k - 1$ terms into the standard association. This gives the standard association on k terms.

The theorem just proved is called the general associative law; it says in effect that parentheses may be omitted in the writing of sums. There is a general commutative law as follows.

5.2. <u>Proposition.</u> The sum of any number of terms is independent of the ordering of the terms.

The proof is left to the student. The idea of the proof is to show that one can pass from any order to any other by a succession of steps each of which is an interchange of two adjacent terms.

§6. Notations

The symbols U, V, W will usually denote vector spaces. Vectors will usually be denoted by A, B, C, X, Y, Z. The symbol R stands for the real number system, and a, b, c, x, y, z will usually represent real numbers (= scalars). R^k is the vector space defined in 3.1. The symbols i, j, k, ℓ, m, n will usually denote integers.

We shall use the symbol $\in$ as an abbreviation for "is an element of". Thus $p \in Q$ should be read: p is an element of the set Q. For example, $x \in R$ means that x is a real number, and $A \in V$ means that A is a vector in the vector space V.

The symbol $\subset$ is an abbreviation for "is a subset of", or, equally well, " is contained in". Thus $P \subset Q$ means that each element of the set P is also an element of Q ($p \in P$ implies $p \in Q$). It is always true that $Q \subset Q$.

If P and Q are sets, the set obtained by uniting the two sets is denoted by $P \cup Q$ and is called the <u>union</u> of P and Q. Thus $r \in P \cup Q$ is equivalent to: $r \in P$ or $r \in Q$ or both. For example, if P is the interval $[1, 3]$ of real numbers and Q is the interval $[2, 5]$, then $P \cup Q$ is the interval $[1, 5]$. In case $P \subset Q$, then $P \cup Q = Q$.

It is convenient to speak of an "empty set". It is denoted by $\emptyset$ and is distinguished by the property of having no elements. If we write $P \cap Q = \emptyset$, we mean that P and Q have no element in common. Obvious tautologies are

$$\emptyset \subset P, \quad \emptyset \cup Q = Q, \quad \emptyset \cap P = \emptyset .$$

§7. Linear subspaces

7.1. Definition. A non-empty subset U of a vector space V is called a linear subspace of V if it satisfies the conditions:

(i) if $A \in U$ and $B \in U$, then $A + B \in U$,

(ii) if $A \in U$ and $x \in R$, then $xA \in U$.

These conditions assert that the two operations of the vector space V give operations in U.

7.2. Proposition. U is itself a vector space with respect to these operations.

Proof. The properties expressed by Axioms 1, 2, 5, 6, 7, 8 are automatically inherited by U. As for Axiom 3, $A \in U$ implies $0A \in U$ by (ii). Since $0A = \tilde{0}$ (Axiom 8), it follows that $\tilde{0} \in U$; hence Axiom 3 holds in U. Similarly, if $A \in U$, then $(-1)A \in U$ by (ii). Since $(-1)A = -A$ (Axiom 8), it follows that $-A \in U$; hence Axiom 4 holds in U.

The addition and multiplication in a linear subspace will always be assumed to be the ones it inherits from the whole space.

It is obvious that the subset of V consisting of the single element $\tilde{0}$ is a linear subspace. It is also trivially true that V is a linear subspace of V. Again, if U is a linear subspace of V, and if U' is a linear subspace of U, then U' is a linear subspace of V.

7.3. Proposition. If V is a vector space and $\{U\}$ is any family of linear subspaces of V, then the vectors common to all the subspaces in $\{U\}$ form a linear subspace of V denoted by $\cap \{U\}$.

Proof. Let $A \in \cap \{U\}$, and $B \in \cap \{U\}$, and $x \in R$.

Then, for each $U \in \{U\}$, we have $A \in U$ and $B \in U$. Since U is a linear subspace, it follows that $A + B \in U$ and $xA \in U$. Since these relations hold for each $U \in \{U\}$, it follows that $A + B \in \cap \{U\}$ and $xA \in \cap \{U\}$. Therefore $\cap \{U\}$ is a linear subspace.

7.4. Definition. If V is a vector space and D is a non-empty subset of V, then any vector obtained as a sum

$$x_1A_1 + x_2A_2 + \cdots + x_kA_k$$

(abbreviated $\Sigma_{i=1}^{k} x_iA_i$), where $A_1, \ldots, A_k$ are all in D, and $x_1, \ldots, x_k$ are any elements of R, is called a finite linear combination of the elements of D. Let $L(D)$ denote the set of all finite linear combinations of the elements of D. It is clearly a linear subspace of V, and it is called the linear subspace spanned by D. We make the convention $L(\emptyset) = \tilde{0}$.

7.5. Proposition. $D \subset L(D)$.

For, if $A \in D$, then $A = 1A$ is a finite linear combination of elements of D (with $k = 1$).

7.6. Proposition. If U is a linear subspace of V, and if D is a subset of U, then $L(D) \subset U$. In particular $L(U) = U$.

The proof is obvious.

Remark. A second method of constructing $L(D)$ is the following: Define $L'(D)$ to be the common part of all linear subspaces of V which contain D. By Proposition 7.3, $L'(D)$ is a linear subspace. Since $L'(D)$ contains D, Proposition 7.6 gives $L(D) \subset L'(D)$. But $L(D)$ is one of the family of

linear subspaces whose common part is $L'(D)$. Therefore $L'(D) \subset L(D)$. The two inclusions $L(D) \subset L'(D)$ and $L'(D) \subset L(D)$ imply $L(D) = L'(D)$. To summarize, $L(D)$ is the <u>smallest</u> linear subspace of V containing D.

§8. Exercises

1. Show that U is a linear subspace of V in each of the following cases:

(a) $V = R^3$ and U = set of triples (x_1, x_2, x_3) such that

$$x_1 + x_2 + x_3 = 0 .$$

(b) $V = R^3$ and U = set of triples (x_1, x_2, x_3) such that $x_3 = 0$.

(c) (See Exercise 4.11), $V = R^D$ and $U = R^{D'}$ where $D' \subset D$.

(d) $V = R^R$, i.e. V = set of all real-valued functions of a real variable, and U = the subset of continuous functions.

2. Find $L(D)$ in the following cases:

(a) $D = \emptyset$, V arbitrary.

(b) $D = \tilde{0}$, V arbitrary.

(c) $V \neq \tilde{0}$ and D consists of a single vector $A \neq \tilde{0}$.

(d) $V = R^3$ and D consists of the two vectors $(3, 0, 0)$ and $(1, 1, 0)$.

(e) $V = R^R$ and D = the set of all polynomial functions.

3. If V is a vector space, and $D' \subset D \subset V$, show that $L(D') \subset L(D)$.

4. If D, D' are subsets of the vector space V, show that $L(D \cap D') \subset L(D) \cap L(D')$. Show by an example that $L(D \cap D')$ may actually be smaller than $L(D) \cap L(D')$.

5. Show that $D \subset D' \subset L(D)$ implies $L(D') = L(D)$.

§9. Independent sets of vectors

9.1. Definition. A set of vectors D is called dependent if there are distinct vectors $A_1, \ldots, A_k$ in D, and real numbers $x_1, \ldots, x_k$ not all zero, such that $\Sigma_{i=1}^{k} x_i A_i = \tilde{0}$. A set of vectors is called independent if it is not dependent. A vector A is said to be dependent on a set D of vectors if $A \in L(D)$.

Examples of dependent sets are many. If $\tilde{0} \in D$, then D is dependent (take $k = 1$, $A_1 = \tilde{0}$ and $x_1 = 1$). If D consists of two vectors A and $B = bA$ where $b \in R$, then D is dependent (take $k = 2$, $A_1 = A$, $A_2 = B$, $x_1 = -b$, $x_2 = 1$). If D consists of three vectors A, B, C where $C = A + B$, then D is dependent because $A + B + (-1)C = \tilde{0}$.

A set consisting of a single vector different from $\tilde{0}$ is an example of an independent set. A set consisting of two non-zero vectors is independent if neither vector is a scalar multiple of the other. The empty set $\emptyset$ is also independent.

9.2. Proposition. A set D is dependent if and only if there is a vector $A \in D$ which is dependent on the other vectors of D.

Proof. If D is dependent, there is a relation $\Sigma x_i A_i = \tilde{0}$ where the A's are distinct in D and some coefficient, say x_1, is not zero. Solving for A_1 gives $A_1 = \Sigma_{i=2}^{k}(-x_i/x_1)A_i$. Therefore A_1 is dependent on the vectors of D other than A_1. Conversely, if $A \in D$ is dependent on the remaining vectors, then $A = \Sigma_{i=1}^{k} x_i B_i$ where the B's are in D and none is A. We may

suppose the B's are distinct; for, if not, we could combine the coefficients of equal B's and obtain a sum involving distinct vectors. Then $A + \Sigma_{i=1}^{k}(-x_i)B_i = \tilde{0}$ is a relation on distinct elements of D, and at least one coefficient (that of A) is not zero. Therefore D is dependent.

9.3. Theorem. If D is any finite set of vectors, then D contains a subset D' such that D' is independent, and $L(D') = L(D)$.

Proof. The proof is by induction on the number of elements of D. If D has but one element A, there are two cases. Either $A = \tilde{0}$ and we take $D' = \emptyset$; or $A \neq \tilde{0}$ and we take $D' = D$. Assume now that the theorem is true for any set of fewer than k elements where $k > 1$. Let D be a set of k distinct elements, say $A_1, \ldots, A_k$. If D is independent, take $D' = D$, and the assertion is true of such a D. Suppose D is dependent. By the above proposition, some A_i is dependent on the remaining. By relabelling if necessary, we may suppose that A_k is dependent on the set D" consisting of $A_1, \ldots, A_{k-1}$. The hypothesis of the induction asserts that D" contains an independent set D' such that $L(D') = L(D")$. Since $A_k \in L(D")$, it follows that $D" \subset D \subset L(D")$, and therefore $L(D) = L(D")$, (see Exercise 8.5). Combining this with $L(D') = L(D")$ gives $L(D') = L(D)$. This completes the induction.

9.4. Proposition. If $A_1, A_2, \ldots, A_k$ is a sequence of vectors which forms a dependent set, then there is an integer i in the range from 1 to k such that A_i is dependent on the set $A_1, A_2, \ldots, A_{i-1}$.

Proof. By hypothesis there is a relation $\Sigma_{j=1}^{k} x_j A_j = \tilde{0}$ where at least one coefficient is not zero. Let i be the largest index of the non-zero coefficients. Then solving for A_i expresses it in terms of $A_1, \ldots, A_{i-1}$.

9.5. Theorem. If $A_1, \ldots, A_k$ is any finite set of vectors in V, and if D is an independent set of vectors in $L(A_1, \ldots, A_k)$, then D is a finite set and the number of its vectors is $\leq k$.

Proof. Suppose the conclusion were false; then D would contain at least $k + 1$ distinct vectors, say $B_1, B_2, \ldots, B_{k+1}$. Abbreviate $L(A_1, \ldots, A_k)$ by L. Since $B_1 \in L$, we have

$$L = L(B_1, A_1, A_2, \ldots, A_k), \tag{1}$$

and the sequence $B_1, A_1, A_2, \ldots, A_k$ is dependent. By Proposition 9.4, some term of the sequence depends on the preceding terms. This term is not B_1 since B_1 by itself is independent (it constitutes a subset of the independent set D). So some A_i depends on $B_1, A_1, A_2, \ldots, A_{i-1}$. We may relabel the A's if necessary, and obtain that A_k belongs to $L(B_1, A_1, A_2, \ldots, A_{k-1})$. Then (1) implies

$$L = L(B_1, A_1, A_2, \ldots, A_{k-1}). \tag{1'}$$

Since $B_2 \in L$, we have

$$L = L(B_2, B_1, A_1, A_2, \ldots, A_{k-1}), \tag{2}$$

and the sequence $B_2, B_1, A_1, A_2, \ldots, A_{k-1}$ is dependent. By Proposition 9.4, some term depends on the preceding terms. This term is not B_2 or B_1 because D is independent. Hence it must

be an A-term. Relabelling the A's if necessary, we can suppose it is the term A_{k-1}. Then (2) gives

(2') $$L = L(B_2, B_1, A_1, A_2, \ldots, A_{k-2}) .$$

Since $B_3 \in L$, we have

(3) $$L = L(B_3, B_2, B_1, A_1, A_2, \ldots, A_{k-2}) ,$$

and the sequence $B_3, B_2, B_1, A_1, A_2, \ldots, A_{k-2}$ is dependent. By Proposition 9.4, some term depends on the preceding terms. It is not a B-term because D is independent. So it must be an A-term which we may discard. Then

(3') $$L = L(B_3, B_2, B_1, A_1, \ldots, A_{k-3}) .$$

We continue in this fashion, first adjoining a B-term, and then discarding an A-term. After k steps we obtain

(k') $$L = L(B_k, B_{k-1}, \ldots, B_2, B_1) ,$$

and there are no more A-terms. Since $B_{k+1} \in L$, it follows that $B_{k+1}, B_k, \ldots, B_1$ is a dependent set. This contradicts the independence of D. Thus the assumption that D has more than k elements leads to a contradiction. This proves the theorem.

§10. Bases and dimension

10.1. Definition. A subset D of a vector space V is called a basis (or base) for V if D is independent and $L(D) = V$. A vector space V is said to be finite dimensional if there is a finite subset D' of V such that $L(D') = V$.

10.2. Theorem. If V is a finite dimensional vector space, then V has a basis. Any basis for V is a finite set, and any two bases have the same number of vectors. (This number is

called the <u>dimension</u> of V, and denoted by $\dim V$.)

<u>Proof</u>. By hypothesis there is a finite set D' of vectors such that $L(D') = V$. By Theorem 9.3, D' contains an independent set D such that $L(D) = L(D')$. Then $L(D) = V$, and D is a basis for V. Let $A_1, \ldots, A_k$ be the distinct elements of D. Let D'' be also a basis for V. By Theorem 9.5 (with D'' in place of D), it follows that D'' is a finite set and the number h of its elements is $\leq k$. But also D is an independent set in $L(D'')$, and Theorem 9.5 asserts that $k \leq h$. Therefore $k = h$, and the proof is complete.

10.3. <u>Theorem</u>. If $A_1, \ldots, A_k$ is a basis for V, and $A \in V$, then there is one and only one sequence of coefficients $x_1, \ldots, x_k$ such that $A = \Sigma_{i=1}^{k} x_i A_i$.

<u>Proof</u>. Since $A \in L(A_1, \ldots, A_k)$, there is at least one such representation. Suppose $A = \Sigma_{i=1}^{k} x_i A_i$ and $A = \Sigma_{i=1}^{k} y_i A_i$ are two representations. Then $\Sigma_{i=1}^{k} (x_i - y_i) A_i = \tilde{0}$. Since $A_1, \ldots, A_k$ are independent, each coefficient must be zero. Therefore $x_i = y_i$ for each i, and the representation is unique.

§11. <u>Exercises</u>

1. If $D' \subset D \subset V$, and if D' is dependent, show that D is also dependent.

2. If $a \in R$, $b \in R$, show that the three vectors $(1, 0)$, $(0, 1)$, (a, b) in R^2 form a dependent set.

3. Let V be a finite dimensional vector space, and let U be a linear subspace:

(a) show that U is finite dimensional,

(b) show that $\dim U \leq \dim V$,

(c) show that $\dim U = \dim V$ implies $U = V$,

(d) show that any basis D' for U can be enlarged to a basis for V,

(e) show by an example that a basis for V need not contain a basis for U.

4. Find a basis for R^k, thus proving $k = \dim R^k$.

5. Find the dimensions of the following vector spaces:

(a) the set of vectors in R^3 satisfying $x_1 + x_2 + x_3 = 0$,

(b) the set of vectors in R^3 of the form $(x, 2x, 3x)$ for all $x \in R$,

(c) R^N where N is a set of n elements (see Exercise 4.11).

6. If $\dim V = k$, and if D is a set of k vectors such that $L(D) = V$, show that D is independent.

7. For $n = 0, 1, \ldots,$ let f_n be the function in R^R defined by $f_n(x) = x^n$. Show that the set of all these functions is independent. Show that the linear subspace of polynomials of degrees $\leq n$ has dimension $n + 1$. Show that the linear subspace of continuous functions in R^R is not finite dimensional.

§12. Parallels and affine subspaces

12.1. Definition. Let U be a linear subspace of V, and A a vector of V. Denote by $A + U$ the set of all vectors of V of the form $A + X$ for some $X \in U$. The set $A + U$ is called a parallel of U in V. It is said to result from parallel translation of U by the vector A. A parallel of some linear subspace of V is called an affine subspace.

12.2. <u>Proposition</u>. If U is a linear subspace of V, then

(i) $A \in A + U$ for each $A \in V$.

(ii) If $B \in A + U$, then $B + U = A + U$.

(iii) Two parallels of U either coincide or have no vector in common.

(iv) Two vectors A, B in V are in the same parallel if and only if $A - B \in U$.

<u>Proof</u>. Since $A = A + \tilde{0}$ and $\tilde{0} \in U$, it follows that $A \in A + U$. To prove (ii), note that $B \in A + U$ implies

(1) $$B = A + C \quad \text{for some} \quad C \in U .$$

An element of $B + U$ has the form $B + X$ for some $X \in U$; so by (1) it has the form $A + (C + X)$ and, since $C + X$ is in U, is an element of $A + U$. Thus (1) implies $B + U \subset A + U$. But (1) can be rewritten $A = B + (-C)$ and $-C \in U$. Then the same argument shows that $A + U \subset B + U$. Since each contains the other, they must coincide. Statement (iii) is an immediate consequence of (ii). To prove (iv), suppose first that A and B are in the same parallel. This must be $B + U$ by (i) and (iii); hence $A = B + C$ for some $C \in U$, and $C = A - B$. Conversely, if $A - B \in U$, then $A = B + (A - B)$ implies $A \in B + U$.

12.3. <u>Proposition</u>. Let $A_1, \ldots, A_h$ be vectors of V. Denote by $E(A_1, \ldots, A_h)$ the set of vectors of the form $\Sigma_{i=1}^{h} x_i A_i$ where $x_1, \ldots, x_h$ are real numbers satisfying the condition

(2) $$\Sigma_{i=1}^{h} x_i = 1 .$$

Then $E(A_1, \ldots, A_h)$ is an affine subspace of V which contains $A_1, \ldots, A_h$ and is the smallest affine subspace containing $A_1, \ldots, A_h$.

Proof. Let U be the linear subspace spanned by the vectors $A_i - A_1$ for $i = 2, 3, \ldots, h$. We assert that

$$E(A_1, \ldots, A_h) = A_1 + U \quad . \tag{3}$$

A vector on the left has the form $\Sigma x_i A_i$ where the x's satisfy (2). But (2) implies $x_1 = 1 - \Sigma_{i=2}^h x_i$, so

$$\Sigma_{i=1}^h x_i A_i = A_1 + \Sigma_{i=2}^h x_i (A_i - A_1)$$

and thus is a vector of $A_1 + U$. Conversely a vector of $A_1 + U$ has the form

$$A_1 + \Sigma_{i=2}^h x_i (A_i - A_1) = (1 - \Sigma_{i=2}^h x_i) A_1 + \Sigma_{i=2}^h x_i A_i \ .$$

The sum of the coefficients on the right is 1; hence the vector is in $E(A_1, \ldots, A_h)$. This proves (3), and the first conclusion of the proposition.

Since $A_i = A_1 + (A_i - A_1)$, and $A_i - A_1$ is in U, we have $A_i \in A_1 + U$, so $A_i \in E(A_1, \ldots, A_h)$ for each i.

Any affine subspace containing $A_1, \ldots, A_h$ must have the form $A_1 + U'$ for some linear subspace U'. Now $A_i \in A_1 + U'$ implies $A_i - A_1 \in U'$. Since this holds for $i = 2, \ldots, h$, it follows that $U' \supset U$, and therefore $A_1 + U' \supset A_1 + U$. Thus $A_1 + U$ is the smallest.

12.4. Definition. A binary relation $\equiv$ defined on a set S is called an equivalence relation if it satisfies the following axioms:

Axiom 1. Reflexivity: $A \equiv A$ for each $A \in S$.

Axiom 2. Symmetry: if $A \equiv B$, then $B \equiv A$.

Axiom 3. Transitivity: if $A \equiv B$ and $B \equiv C$, then $A \equiv C$.

12.5. Proposition. An equivalence relation on a set S defines a partition of S into non-empty, mutually disjoint subsets called equivalence classes, such that two elements of S lie in the same equivalence class if and only if they are equivalent.

The proof of this proposition is left as an exercise.

12.6. Definition. If U is a linear subspace of a vector space V, let $A \equiv B \pmod{U}$, read "modulo U", if and only if $A - B \in U$, $A, B \in V$.

A review of Definition 12.1, and Proposition 12.2 shows that the equivalence classes mod U of V can be identified with the parallels of U in V and that Definition 12.6 does indeed define an equivalence relation.

§13. Exercises

1. Let U, U' be linear subspaces of V, and Let A, A' be vectors in V. Show that the intersection

$$(A + U) \cap (A' + U')$$

is either empty or is a parallel of $U \cap U'$.

2. If $U = L(B_1, \ldots, B_h)$, show that $A + U = E(A, A + B_1, \ldots, A + B_h)$.

3. Prove Proposition 12.5.

4. Verify that the following definitions give equivalence relations:

(a) S = set of all triangles in the euclidean plane, with $A \equiv B$ if and only if A is congruent to B;

(b) S as in (a), with "congruent" replaced by "similar";

(c) S = set of integers, with $a \equiv b \pmod{m}$ if and only if $a - b$ is divisible by m, where m is a fixed integer.

II. LINEAR TRANSFORMATIONS OF VECTOR SPACES

§1. Introduction

Our ultimate objective is the study of functions whose domains and ranges are vector spaces. This chapter treats the simplest class of such functions, the linear transformations. The only linear transformations having the real numbers R as domain and range are those of the form $f(x) = mx$ where m is a real number. This is too simple a case to indicate the importance of the general concept. A linear transformation from R^2 to R^3 is given by a system of linear equations

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \\ y_3 &= a_{31}x_1 + a_{32}x_2 \end{aligned}$$

where (x_1, x_2) are the components of a vector X of R^2, the a's are fixed real numbers, and the resulting (y_1, y_2, y_3) are the components of the vector in R^3 which is the transform of X. Briefly, linear transformations are those representable (in terms of components) by systems of linear equations. There are two points to emphasize. When the dimensions of the domain and range exceed 1, there are a great variety of linear transformations (rotations, dilations, projections, etc.), and they have and deserve an extensive theory. The second point is that it is cumbersome to define a linear transformation as something given by linear equations. Instead we shall define it as a function having two simple properties as follows.

1.1. Definition. If V, W are vector spaces, a linear transformation T of V into W, written $T : V \longrightarrow W$, is a function having V as its domain and W as its range, and such that

(1) $T(A + B) = T(A) + T(B)$ for all $A, B \in V$,

(2) $T(xA) = xT(A)$ for all $A \in V$, $x \in R$.

These properties can be paraphrased by saying that T preserves addition and multiplication by a scalar. It is easily checked that the linear functions mentioned above from R to R, and from R^2 to R^3 do have these properties.

1.2. Definition. Let V and W be sets and let T be a function with domain V and range W. For any $A \in V$, the value of T at A, denoted by $T(A)$, is called the image of A under T; $T(A)$ is also referred to as an image point. If D is any subset of the domain of T, the image of D under T, denoted by $T(D)$, consists of the set of images of elements of D. The set $T(V)$ is also denoted by $\operatorname{im} T$. If E is any subset of W, then the inverse image of E under T, denoted by $T^{-1}(E)$, is the set of those elements of V whose images are in E. For example, $T^{-1}(W) = V$. If no element of V has an image in E, then $T^{-1}(E) = \emptyset$.

It should be noted that, if E consists of a single element of W, this need not be true of $T^{-1}(E)$, i.e. T^{-1} need not define a function from $\operatorname{im} T \subset W$ to V; T^{-1} is a function from the subsets of W to the subsets of V.

1.3. Definition. Let V and W be sets. A function $T : V \longrightarrow W$ is called

(i) <u>injective</u> ("one-one into") if $A \neq B$ implies $T(A) \neq T(B)$, for all $A, B \in V$;

(ii) <u>surjective</u> ("onto") if $T(V) = W$, or $\operatorname{im} T = W$;

(iii) <u>bijective</u> ("one-one") if it is both injective and surjective.

<u>Remark</u>. If $T : V \longrightarrow W$ is injective, then $T^{-1}(E)$, when E consists of a single element of W, is a single element of V or is $\emptyset$. If T is bijective, then the bijective function $T^{-1} : W \longrightarrow V$ is also defined.

1.4. <u>Proposition</u>. If $S : U \longrightarrow V$ and $T : V \longrightarrow W$, then the composition $TS : U \longrightarrow W$, defined by $(TS)(X) = T(S(X))$ for $X \in U$, is

(i) injective if S and T are injective;

(ii) surjective if S and T are surjective;

(iii) bijective if S and T are bijective.

The proof of this proposition is left as an exercise.

1.5. <u>Proposition</u>. If $S : U \longrightarrow V$, $T : V \longrightarrow W$, and $Y : W \longrightarrow Z$, then

$$Y(TS) = (YT)S .$$

<u>Proof</u>. By the definition of composition,

$$(Y(TS))(X) = Y((TS)X) = Y(T(S(X)))$$

and

$$((YT)S)(X) = (YT)(S(X)) = Y(T(S(X)))$$

for each $X \in U$.

1.6. <u>Proposition</u>. For any set V, the function $I_V : V \longrightarrow V$, defined by $I_V(X) = X$ for each $X \in V$, is

bijective and satisfies $TI_V = T$ for any function with domain V and $I_V S = S$ for any function with range V.

1.7. Proposition. If $T : V \longrightarrow W$ is bijective, then

$$T^{-1}T = I_V, \quad TT^{-1} = I_W .$$

If $S : W \longrightarrow V$ satisfies $ST = I_V$, then $S = T^{-1}$.

Proof. For each $X \in V$,

$$(T^{-1}T)(X) = T^{-1}(T(X)) = X ,$$

and for each $Y \in W$,

$$(TT^{-1})(Y) = T(T^{-1}(Y)) = Y .$$

If $ST = I_V$, then

$$S = SI_W = S(TT^{-1}) = (ST)T^{-1} = I_V T^{-1} = T^{-1} .$$

1.8. Definition. Let V and W be sets having an algebraic structure (i.e. certain laws of composition, together with a set of axioms satisfied by these operations) of the same type. A function $T : V \longrightarrow W$ which preserves the given operations is called a homomorphism. A bijective homomorphism is called an isomorphism. A homomorphism $T : V \longrightarrow V$ is called an endomorphism of V, and a bijective endomorphism is called an automorphism of V.

Remark. When the algebraic structure is that of a vector space, a homomorphism is usually called a linear transformation, as in Definition 1.1, but no special terms are introduced for isomorphisms, etc.

§2. Properties of linear transformations

2.1. Theorem. If U, V and W are vector spaces, and

if $S : U \longrightarrow V$ and $T : V \longrightarrow W$ are linear, then the composition $TS : U \longrightarrow W$ is a linear transformation.

<u>Proof</u>. If $A, B, \in U$, and $x \in R$, then

$$\begin{aligned}(TS)(A + B) &= T(S(A + B)) = T(S(A) + S(B)) \\ &= T(S(A)) + T(S(B)) = (TS)(A) + (TS)(B) , \\ (TS)(xA) &= T(S(xA)) = T(xS(A)) \\ &= xT(S(A)) = x(TS)(A) .\end{aligned}$$

2.2. <u>Theorem</u>. If $T : V \longrightarrow W$ is a linear transformation, then:

(i) $T(\tilde{0}_V) = \tilde{0}_W$, i.e. T transforms zero into zero.

(ii) If U is a linear subspace of V, then its image $T(U)$ is a linear subspace of W.

(iii) If $A_1, \ldots, A_k$ are in V, and $x_1, \ldots, x_k$ are in R, then

$$T(\Sigma_{i=1}^{k} x_i A_i) = \Sigma_{i=1}^{k} x_i T(A_i) .$$

(iv) If $D \subset V$, and $L(D)$ is the linear subspace spanned by D, then $T(L(D)) = L(T(D))$.

<u>Proof</u>. Since $0\tilde{0}_V = \tilde{0}_V$, the linearity property (2) of Definition 1.1 gives

$$T(\tilde{0}_V) = T(0\tilde{0}_V) = 0T(\tilde{0}_V) = \tilde{0}_W .$$

This proves (i).

To prove (ii), let A', B' be vectors in $T(U)$, and $x \in R$. Then A' and B' are the images of vectors A, B of U. Using the linearity property (1), we have

$$A' + B' = T(A) + T(B) = T(A + B) .$$

Since U is a linear subspace, $A + B \in U$; hence $A' + B' \in T(U)$. Using the linearity property (2), we have

$$xA' = xT(A) = T(xA) .$$

Since U is a linear subspace, $xA \in U$; hence $xA' \in T(U)$. Thus $T(U)$ is a linear subspace of W.

The proof of (iii) is by induction on the number k of vectors. The case of one vector reduces to $T(x_1A_1) = x_1T(A_1)$ which is property (2) of linearity. Suppose (iii) is true if there are fewer than k vectors. Then

$$\begin{aligned} T(\Sigma_{i=1}^{k} x_iA_i) &= T(\Sigma_{i=1}^{k-1} x_iA_i + x_kA_k) \\ &= T(\Sigma_{i=1}^{k-1} x_iA_i) + T(x_kA_k) \\ &= \Sigma_{i=1}^{k-1} x_iT(A_i) + x_kT(A_k) = \Sigma_{i=1}^{k} x_iT(A_i) . \end{aligned}$$

The second step above follows from property (1) of linearity, and the third step applies the hypothesis of the induction, and property (2) of linearity. This completes the proof of the inductive step, and hence of (iii).

To prove (iv), we must show that each vector in $T(L(D))$ is a vector of $L(T(D))$ and conversely. A vector of $T(L(D))$ is an image $T(A)$ for some $A \in L(D)$. By Definition I, 7.4, $A = \Sigma_{i=1}^{k} x_iA_i$ where $A_1, \ldots, A_k$ are in D. Applying part (iii) proved above, we obtain $T(A) = \Sigma_{i=1}^{k} x_iT(A_i)$. Since each $T(A_i)$ is in $T(D)$ it follows that $T(A) \in L(T(D))$. This proves the first half. Now let B be a vector in $L(T(D))$. Then $B = \Sigma_{j=1}^{m} y_jB_j$ where $B_1, \ldots, B_m$ are vectors in $T(D)$. These are the images of vectors $A_1, \ldots, A_m$, respectively, belonging to D. Applying

part (iii) again, we have

$$B = \Sigma_{j=1}^{m} y_j T(A_j) = T(\Sigma_{j=1}^{m} y_j A_j) .$$

Since $\Sigma y_j A_j \in L(D)$, it follows that $B \in T(L(D))$. This completes the proof of the theorem.

2.3. Theorem. If $T : V \longrightarrow W$ is linear, and U is a linear subspace of W, then $T^{-1}(U)$ is a linear subspace of V.

Proof. Suppose A and B are in $T^{-1}(U)$, and $x \in R$. By definition $T(A)$ and $T(B)$ are in U. Since U is a linear subspace, $T(A) + T(B)$ and $xT(A)$ are in U. Since T is linear, it follows that $T(A + B)$ and $T(xA)$ are in U. Therefore $A + B$ and xA are in $T^{-1}(U)$. Hence $T^{-1}(U)$ is a linear subspace of V.

2.4. Definition. If $T : V \longrightarrow W$ is linear, the linear subspace $T^{-1}(\tilde{0}_W)$ of V is called the kernel of T, and denoted by $\ker T$.

For example, if $T : V \longrightarrow W$ is the "constant" linear transformation which transforms each vector of V into $\tilde{0}_W$, then $\ker T = V$.

2.5. Theorem. If $T : V \longrightarrow W$ is linear and V is finite dimensional, then $\ker T$ and $\operatorname{im} T$ are finite dimensional, and

$$\dim V = \dim (\ker T) + \dim (\operatorname{im} T) .$$

Proof. By Exercise I, 11.3(a) and Theorem I, 10.2, the kernel is finite dimensional, and has a basis. Let $A_1, \ldots, A_j$ be a basis for $\ker T$. (In case the kernel is $\tilde{0}_V$, then $j = 0$ and the basis is empty.) By Exercise I, 11.3(d), we may extend the

basis for $\ker T$ to a basis $A_1, \ldots, A_j, A_{j+1}, \ldots, A_k$ for V. By definition (I, 10.2),

$$k = \dim V, \qquad j = \dim (\ker T) .$$

So if we can prove that $T(A_{j+1}), \ldots, T(A_k)$ is a basis for $\operatorname{im} T$, it will follow that

$$\dim (\operatorname{im} T) = k - j ,$$

and the theorem will be proved. A vector of $\operatorname{im} T = T(V)$ has the form $T(A)$ for some $A \in V$. Since $A_1, \ldots, A_k$ is a basis for V, we have $A = \Sigma_{i=1}^{k} x_i A_i$ for some $x_1, \ldots, x_k$. Then, by Theorem 2.2 (iii),

$$T(A) = \Sigma_{i=1}^{k} x_i T(A_i) = \Sigma_{i=j+1}^{k} x_i T(A_i)$$

because $T(A_1), \ldots, T(A_j)$ are all zero. This proves that $L(T(A_{j+1}), \ldots, T(A_k))$ is $T(V)$. It remains to show that $T(A_{j+1}), \ldots, T(A_k)$ are independent. Suppose $\Sigma_{i=j+1}^{k} y_i T(A_i) = \tilde{0}_W$. This implies $T(\Sigma_{i=j+1}^{k} y_i A_i) = \tilde{0}_W$, and therefore $\Sigma_{i=j+1}^{k} y_i A_i \in \ker T$. This element of $\ker T$ must be expressible in terms of the basis $A_1, \ldots, A_j$ of $\ker T$:

$$\Sigma_{i=j+1}^{k} y_i A_i = \Sigma_{h=1}^{j} z_h A_h .$$

The independence of all the A's implies that each coefficient in this relation is zero. In particular, all the y's are zero; thus we have shown that $T(A_{j+1}), \ldots, T(A_k)$ are independent. This completes the proof.

2.6. <u>Definition</u>. An <u>affine transformation</u> $S : V \longrightarrow W$ is one which is obtained from a linear transformation $T : V \longrightarrow W$

by adding a constant vector $C \in W$. Thus

$$S(X) = C + T(X) \qquad \text{for } X \in V .$$

The affine transformations $R \longrightarrow R$ are those of the form $mx + c$ where m and c are constants.

Since affine transformations differ from linear transformations by constants, they have essentially the same properties. These will be developed in the exercises.

§3. Exercises

1. Let V and W be sets, and let T be a function with domain V and range W. Let D and E be subsets of V. Show that

(a) $$T(D \cup E) = T(D) \cup T(E) ,$$

(b) $$T(D \cap E) \subset T(D) \cap T(E) .$$

Let D' and E' be subsets of W. Show that

(c) $$T^{-1}(D' \cup E') = T^{-1}(D') \cup T^{-1}(E') ,$$

(d) $$T^{-1}(D' \cap E') = T^{-1}(D') \cap T^{-1}(E') .$$

2. Let V and W be sets, and let T be a function from V to W. Suppose there exists a function S from W to V such that (i) $TS = I_W$ or (ii) $ST = I_V$. Show that (i) implies that T is surjective, that (ii) implies that T is injective, and that (i) and (ii) together imply that T is bijective, so $S = T^{-1}$.

3. If V, W are vector spaces, and T is a function from V to W having property (iii) of Theorem 2.2, show that T is linear.

4. Let $T : R^3 \longrightarrow R$ be defined by

$$T(x_1, x_2, x_3) = x_1 - 3x_2 + 2x_3 .$$

Show that T is linear. Describe $\ker T$. What is its dimension? Describe the inverse image of $1 \in R$.

5. Let $T : R^3 \longrightarrow R^3$ be defined by

$$T(x_1, x_2, x_3) = (x_1, 0, x_2).$$

Show that T is linear. Describe $\ker T$ and $\operatorname{im} T$. What are their dimensions? Let U be the linear subspace of vectors satisfying $x_3 = 0$. Describe $T(U)$ and $T^{-1}(U)$.

6. Let T and T' be linear transformations $V \longrightarrow W$, let D be a basis for V, and suppose $T(A) = T'(A)$ for each $A \in D$. Show that $T(A) = T'(A)$ for each $A \in V$.

7. Let W be the vector space of continuous functions $R \longrightarrow R$ (see Exercise I, 4.11), and let V be the subset of functions having continuous derivatives. Show that V is a linear subspace of W. Let $T : V \longrightarrow W$ transform each function in V into its derivative. Show that T is linear. What is $\ker T$, and what is $\operatorname{im} T$? If U_n is the subspace of polynomials of degrees $\leq n$, what is $T(U_n)$?

8. Let $T : V \longrightarrow W$ be linear, and let U be a linear subspace of V. Show that T transforms each parallel of U in V (see Definition I, 12.1) into a parallel of $T(U)$ in $T(V)$. (This proves that a linear transformation carries affine subspaces into affine subspaces.)

9. Let $T : V \longrightarrow W$ be linear, and let $B \in \operatorname{im} T$. Show that $T^{-1}(B)$ is a parallel of $\ker T$.

10. Let $T : V \longrightarrow W$ be linear, and let E be an affine subspace of W. Show that $T^{-1}(E)$ is either empty or an affine subspace of V.

11. Show that a function $S : V \longrightarrow W$ is affine if and only if the function $T : V \longrightarrow W$ defined by $T(X) = S(X) - S(\tilde{0})$ is linear.

12. If $S : V \longrightarrow W$ is affine, and U is a linear subspace of V, then S transforms the family of parallels of U into a family of parallels in $S(V)$.

13. Show that an affine transformation $R^2 \longrightarrow R^2$ carries each parallelogram into a parallelogram.

14. In R^3, let $A = (1, 2, 0)$, $B = (-2, 1, 2)$ and $C = (0, 3, -1)$. Find equations for an affine transformation $S : R^2 \longrightarrow R^3$ such that $S(R^2)$ is the plane passing through A, B, C.

15. Find an affine transformation $S : R^3 \longrightarrow R$ such that $S^{-1}(0)$ is the plane described in Exercise 14.

16. Show that the composition of two affine transformations is an affine transformation.

§4. Existence of linear transformations

4.1. Theorem. Let V be a vector space of finite dimension, and let $A_1, \ldots, A_k$ be a basis for V. Let W be a vector space, and let $B_1, \ldots, B_k$ be a set of k vectors in W. Then there exists one and only one linear transformation $T : V \longrightarrow W$ such that $T(A_i) = B_i$ for $i = 1, \ldots, k$.

Proof. For any $X \in V$, construct $T(X)$ as follows. By Theorem I, 10.3, X has a unique representation

$$X = \Sigma_{i=1}^{k} x_i A_i \ . \tag{1}$$

Define $T(X)$ by

(2) $$T(X) = \Sigma_{i=1}^{k} x_i B_i \ .$$

To prove the linearity of T, let $X, Y \in V$. The representations of X, Y and $X + Y$ are $X = \Sigma x_i A_i$, $Y = \Sigma y_i A_i$, $X + Y = \Sigma(x_i + y_i)A_i$. By (2) we must have $T(X) = \Sigma x_i B_i$, $T(Y) = \Sigma y_i B_i$, $T(X + Y) = \Sigma(x_i + y_i)B_i$. Hence $T(X + Y) = T(X) + T(Y)$. If $a \in R$, the representation of aX is $\Sigma a x_i A_i$; hence

$$T(aX) = \Sigma a x_i B_i = a\Sigma x_i B_i = aT(X) \ .$$

This proves the linearity of T.

If $X = A_i$ for some i, then in (1) all coefficients $x_j = 0$ except $x_i = 1$. Then (2) gives $T(X) = B_i$. Therefore $T(A_i) = B_i$.

To prove there is only one such T, suppose $T' : V \longrightarrow W$ is linear and $T'(A_i) = B_i$. By Theorem 2.2 (iii),

$$T'(X) = \Sigma x_i T'(A_i) = \Sigma x_i B_i = T(X) \ .$$

So $T' = T$, and the proof is complete.

The above theorem enables us to construct a great variety of linear transformations, and it also gives a rough idea of the quantity of linear transformations. Consider, for example, the linear transformations $R^2 \longrightarrow R^2$. A basis for R^2 consists of two vectors, e.g. $A_1 = (1, 0)$ and $A_2 = (0, 1)$. Then any ordered pair of vectors B_1, B_2 determines a unique T carrying A_1 into B_1, and A_2 into B_2. The totality of linear transformations is the same as the totality of pairs of vectors.

§5. Matrices

Let V and W be vector spaces having bases $A_1, \ldots, A_k$

and $B_1, \ldots, B_n$ respectively. Let $T : V \longrightarrow W$ be linear. For each $i = 1, \ldots, k$, the image $T(A_i)$ is uniquely expressible in terms of the B's, say

$$(1) \qquad T(A_i) = \Sigma_{j=1}^{n} \alpha_{ji} B_j \ .$$

That is to say, the number α_{ji} is the coefficient of B_j in the expansion of $T(A_i)$.

We have seen in §4 that T is completely determined by $T(A_i)$, and hence also by the system α_{ji} of coefficients. In fact, let $X \in V$, and let its representation in terms of the basis be

$$(2) \qquad X = \Sigma_{i=1}^{k} x_i A_i .$$

Let $Y = T(X)$, and let its representation be

$$(3) \qquad Y = \Sigma_{j=1}^{n} y_j B_j \ .$$

If we apply T to both sides of (2), and substitute for $T(A_i)$ from (1), we obtain

$$(4) \qquad \begin{aligned} Y = T(X) &= \Sigma_{i=1}^{k} x_i (\Sigma_{j=1}^{n} \alpha_{ji} B_j) \\ &= \Sigma_{j=1}^{n} (\Sigma_{i=1}^{k} \alpha_{ji} x_i) B_j \ . \end{aligned}$$

Since the B's form a basis, the representations (3) and (4) of Y must coincide. Therefore

$$(5) \qquad y_j = \Sigma_{i=1}^{k} \alpha_{ji} x_i \qquad \text{for each } j = 1, \ldots, n \ .$$

This justifies the statement of §1 that a linear transformation is a function given by a system of linear equations.

The system α_{ji} of coefficients is usually regarded as

a rectangular array of numbers, having n rows and k columns, called a <u>matrix</u>, thus

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nk} \end{pmatrix} .$$

The numbers in the ith column are the components of $T(A_i)$ with respect to the basis $B_1, \ldots, B_n$. (See (1).)

5.1. <u>Definition</u>. Let V and W be vector spaces having bases $A_1, \ldots, A_k$ and $B_1, \ldots, B_n$ respectively, and $T : V \longrightarrow W$ be linear. The matrix (α_{ji}) of coefficients in (1) (or in (5)) defines the <u>matrix representation</u> of T relative to the given choice of bases for V and W.

If T, T' are distinct linear transformations $V \longrightarrow W$ they will differ on some A_i, and hence will have different matrices. Any $n \times k$ rectangular array of numbers is the matrix of a corresponding T defined by (1). Thus we have a bijective correspondence between the set of all linear transformations $V \longrightarrow W$ and the set of all $n \times k$ matrices. The numbers in the matrix can be thought of as coordinates of the corresponding T.

The advantage of the matrix notation is that it provides an easy method of designating a particular T. For example, a linear $T : R^3 \longrightarrow R^3$ is given by

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

Referring to (5), this means that the coordinates (y_1, y_2, y_3) of

the image of (x_1, x_2, x_3) are

$$\begin{aligned} y_1 &= 2x_1 + 0x_2 - 3x_3 \ , \\ y_2 &= x_1 - x_2 + 2x_3 \ , \\ y_3 &= 0x_1 + 3x_2 + x_3 \ . \end{aligned}$$

In the literature of mathematics, matrices are accorded a much greater prominence than their usefulness deserves. The reason for this lies in the history of the development of the ideas of vector space and linear transformation. Before the axiomatic approach was found, a vector space was an R^k for some k, and a linear transformation was a system of linear equations (abbreviated as a matrix). Linear transformations were treated through their matrices, and their properties were described as properties of matrices. In this way an extensive theory of matrices arose. It is a cumbersome theory both in notation and conception. When the axiomatic approach was developed, it became clear that the matrix theory tends to obscure the geometric insight. As a result matrices are no longer used in the development of the theory. They are still used and are useful in applications of the theory to special cases.

It should be emphasized that the matrix representation of T depends on the choices of the bases in V and W.

5.2. <u>Theorem</u>. If V and W are finite dimensional, and $T : V \longrightarrow W$ is linear, then there exist bases $A_1, \ldots, A_k$ and $B_1, \ldots, B_n$ for V and W, respectively, and an integer h in the range $0 \leq h \leq k$ such that

(6) $$T(A_i) = B_i \qquad \text{for } i = 1, \ldots, h \ ,$$

(7) $$T(A_i) = \tilde{0}_W \qquad \text{for } i = h+1, \ldots, k .$$

The integer h is the dimension of $T(V) = \operatorname{im} T$.

Proof. Choose a basis $B_1, \ldots, B_h$ for $T(V)$. By Exercise I, 11.3 (d), we may adjoin vectors $B_{h+1}, \ldots, B_n$ so that $B_1, \ldots, B_n$ is a basis for W. For each $i = 1, \ldots, h$, choose a vector A_i such that $T(A_i) = B_i$. This is possible since $B_i \in T(V)$. By Theorem 2.5, a basis for $\ker T$ must have $k - h$ vectors. Choose such a basis and designate its elements by $A_{h+1}, \ldots, A_k$. Then formulas (6) and (7) hold by construction. It remains to show that $A_1, \ldots, A_k$ is a basis for V. Suppose $\Sigma_{i=1}^{k} x_i A_i = \tilde{0}_V$. Applying T to both sides and using (6) and (7), we get $\Sigma_{i=1}^{h} x_i B_i = \tilde{0}_W$. Since the B's form a basis, we must have $x_i = 0$ for $i = 1, \ldots, h$. So the original relation reduces to $\Sigma_{i=h+1}^{k} x_i A_i = \tilde{0}_V$. Since $A_{h+1}, \ldots, A_k$ is a basis for $\ker T$, we must have $x_i = 0$ for $i = h+1, \ldots, k$. Thus all the x's are zero, and the proof is complete.

With bases chosen as in the theorem, the resulting matrix representation of T has a specially simple form, namely, all α_{ji} are zero save $\alpha_{ii} = 1$ for $i = 1, \ldots, h$. Thus the rectangular array has zeros everywhere except along the diagonal starting at the upper left, and along this diagonal there is a string of h ones followed by zeros. Because of the appearance of this matrix, the transformation is said to be <u>in diagonal form</u> with respect to the two bases.

§6. Exercises

1. Define a linear transformation $T : R^2 \longrightarrow R^3$ by

$$T(1, 0) = (1, 1, 0), \quad T(0, 1) = (0, 1, 1) .$$

Show that $\ker T$ is zero. What is the matrix for T in terms of the usual bases? Write an equation for the plane $T(R^2)$.

2. Define a linear transformation $T : R^3 \longrightarrow R^2$ by

$$T(1,0,0) = (1,1), \quad T(0,1,0) = (0,1), \quad T(0,0,1) = (-1,1).$$

Find a basis for $\ker T$. Find a vector $A \in R^3$ such that $T(A) = (1, 0)$. What is the matrix for T in terms of the usual bases?

3. If $T : R^2 \longrightarrow R^3$ is as defined in Exercise 1, find the components of basis vectors A_1, A_2 in R^2 and B_1, B_2, B_3 in R^3 such that the matrix of T with respect to these bases is in diagonal form.

4. Do the same as in Exercise 3 for $T : R^3 \longrightarrow R^2$ as defined in Exercise 2.

5. If we change the hypotheses of Theorem 4.1 by replacing "let $A_1, \ldots, A_k$ be a basis for V" by "let $A_1, \ldots, A_k$ span V", then the first paragraph of the proof need be altered only by the deletion of "unique". That is, we <u>choose</u> a representation of X of the form (1) and define $T(X)$ by (2). Is the remainder of the proof valid? If not, is the conclusion of the altered theorem still valid?

§7. Isomorphism of vector spaces

7.1. <u>Proposition</u>. A linear transformation $T : V \longrightarrow W$ is injective if and only if $\ker T = \tilde{0}_V$.

<u>Proof</u>. Since T is linear, $\tilde{0}_V \in \ker T$. If T is injective, only one vector of V can have $\tilde{0}_W$ as an image, so $\tilde{0}_V = \ker T$. Conversely, let T be linear and such that

$\ker T = \tilde{0}_V$. Suppose A, A' are vectors of V such that $T(A) = T(A')$. Then $T(A - A') = \tilde{0}_W$, $A - A' \in \ker T = \tilde{0}_V$; hence $A = A'$. That is, T is injective.

7.2. <u>Proposition</u>. Let $T : V \longrightarrow W$ be linear, where V and W are finite dimensional.

(i) If T is injective, then $\dim V \leq \dim W$.

(ii) If T is surjective, then $\dim V \geq \dim W$.

(iii) If T is bijective, then $\dim V = \dim W$.

<u>Proof</u>. By Theorem 2.5,

$$\dim V = \dim (\ker T) + \dim (\operatorname{im} T) .$$

If T is injective, then $\ker T = \tilde{0}_V$, and

$$\dim V = \dim (\operatorname{im} T) \leq \dim W. \tag{1}$$

If T is surjective, then $\operatorname{im} T = W$, and

$$\dim V = \dim (\ker T) + \dim W \geq \dim W.$$

7.3. <u>Theorem</u>. If V and W are finite dimensional, and $T : V \longrightarrow W$ is linear, then T is an isomorphism (i.e. bijective) if and only if $\dim V = \dim W$, and $\ker T = \tilde{0}_V$.

<u>Proof</u>. It remains only to show that $\dim V = \dim W$ implies $\operatorname{im} T = W$ if $\ker T = \tilde{0}_V$. This follows from (1) and Exercise I, 11.3 (c).

7.4. <u>Proposition</u>. If $T : V \longrightarrow W$ is an isomorphism, then the inverse function T^{-1} is a linear transformation $T^{-1} : W \longrightarrow V$, and therefore an isomorphism.

<u>Proof</u>. By Proposition 1.7, we have $T(T^{-1}(X)) = X$ for $X \in W$. Suppose A, B are in W, and $x \in R$. Using the linearity of T, we have

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

Also $T(T^{-1}(A + B)) = A + B$. Since $A + B$ is the image of only one vector, it follows that

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

Similarly, applying T to $T^{-1}(xA)$ and to $xT^{-1}(A)$, the image of both is found to be xA. Since only one vector has xA as image, it follows that

$$T^{-1}(xA) = xT^{-1}(A) .$$

7.5. Theorem. If V and W are vector spaces having the same finite dimension n, then there exists an isomorphism $T : V \longrightarrow W$.

Proof. By definition of dimension n, there exist bases in V and W having n elements, say $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$. By Theorem 4.1, a linear transformation $T : V \longrightarrow W$ is defined by setting $T(A_i) = B_i$ for $i = 1, \ldots, n$. The proof that T is an isomorphism is left to the reader.

7.6. Corollary. Each n-dimensional vector space is isomorphic to R^n.

§8. The space of linear transformations

Let $L(V, W)$ denote the set of all linear transformations of V into W. For S and T in $L(V, W)$, define a function $S + T$ from V to W by

(1) $$(S + T)(X) = S(X) + T(X) \qquad \text{for } X \in V .$$

It is easily verified that $S + T$ satisfies the two conditions (Definition 1.1) for linearity; hence $S + T \in L(V, W)$. For

$T \in L(V, W)$, and $a \in R$, define a function aT from V to W by

(2) $$(aT)(X) = aT(X) \qquad \text{for } X \in V .$$

Again we have $aT \in L(V, W)$.

8.1. Theorem. With respect to the addition and scalar multiplication defined above, the set $L(V, W)$ is a vector space. If V and W have dimensions k and n, then $L(V, W)$ has dimension kn.

Proof. The proof that $L(V, W)$ is a vector space is the same as the proof (Exercise I, 4.12) that the set W^V of all functions from V to W is a vector space with respect to the addition and multiplication defined by (1) and (2). In fact $L(V, W)$ is a linear subspace of W^V. The zero vector $\tilde{0}_{L(V,W)}$ is the constant linear transformation of V into W.

To prove the second statement, choose bases $\{A_i\}$, $\{B_j\}$ for V, W as in §5. Then each $T \in L(V, W)$ has a matrix $M(T) = (\alpha_{ji})$ where

(3) $$T(A_i) = \Sigma_{j=1}^{n} \alpha_{ji} B_j .$$

If the kn numbers α_{ji} are regarded as components of a vector in R^{kn}, then M is a function from $L(V, W)$ to R^{kn}. The conclusion follows from Theorem 7.3 and

8.2. Proposition. The matrix representation

$$M : L(V, W) \longrightarrow R^{kn}$$

is an isomorphism.

Proof. If T is given by (3) above, and $x \in R$,

then $(xT)(A_i) = xT(A_i) = \Sigma x\alpha_{ji}B_j$. Therefore $M(xT) = (x\alpha_{ji}) = xM(T)$. If $M(S) = (\beta_{ji})$, then

$$(S + T)(A_i) = S(A_i) + T(A_i) = \Sigma\alpha_{ji}B_j + \Sigma\beta_{ji}B_j$$
$$= \Sigma(\alpha_{ji} + \beta_{ji})B_j \ .$$

Therefore $M(S + T) = M(S) + M(T)$. This proves that M is linear. We have seen already in §5 that M is bijective, and hence is an isomorphism.

8.3. <u>Proposition</u>. Let U, V, W be vector spaces, and let $S : U \longrightarrow V$ be linear. The induced function

$$S^* : L(V, W) \longrightarrow L(U, W) \quad ,$$

defined by $S^*(T) = TS$ for each $T \in L(V, W)$, is linear. Similarly, if $T : V \longrightarrow W$ is linear, the induced function

$$T_* : L(U, V) \longrightarrow L(U, W) \ ,$$

defined by $T_*(S) = TS$ for each $S \in L(U, V)$, is linear.

The proof of this proposition is left as an exercise.

<u>Remark</u>. There is a transformation S^* for each choice of W; that is, $S^* = S^*_W$. In general, the notation S^* is used for any transformation, induced by S, which "reverses" the direction of S. The precise transformation denoted by S^* in any particular case is either clear from the context or given explicitly, as in the statement of Proposition 8.3. Analogously, the notation S_* is used for any transformation, induced by S, which goes in the "same" direction as S.

8.4. <u>Definition</u>. For any vector space V, the vector space $L(V, R)$ is called the <u>dual space</u> of V and is denoted by V^*.

Dual spaces will be studied in Chapter IX. However, we state the following corollaries of Propositions 8.2 and 8.3, obtained by taking $W = R$.

8.5. Corollary. If V is a vector space of dimension k, then V^* has dimension k.

8.6. Corollary. Any linear transformation $S : U \longrightarrow V$ induces a linear transformation $S^* : V^* \longrightarrow U^*$.

§9. Endomorphisms

By Definition 1.8, an endomorphism of a vector space V is an element of the vector space $L(V, V)$, which will be denoted by $E(V)$. The operations of addition and scalar multiplication in $E(V)$ are defined by (1) and (2) of §8 and satisfy the axioms of Definition I, 1.1. The zero vector of $E(V)$ will be denoted by $\tilde{0}_E$.

In addition, the composition TS of any two endomorphisms T, S of V is defined and again an endomorphism of V, by Theorem 2.1. Thus, a multiplication is defined in $E(V)$.

9.1. Theorem. With multiplication in $E(V)$ defined by composition of endomorphisms, the vector space $E(V)$ is an algebra (with unit) over the real numbers; that is, the following additional axioms are satisfied.

Axiom 9. $T_1(T_2T_3) = (T_1T_2)T_3$ for each triple T_1, T_2, T_3 of elements of $E(V)$. (I.e. multiplication is associative.)

Axiom 10. $S(T_1 + T_2) = ST_1 + ST_2$ and $(T_1 + T_2)S = T_1S + T_2S$ for each triple S, T_1, T_2 of elements of $E(V)$. (I.e. multiplication is distributive with respect to (vector) addition.)

Axiom 11. $a(ST) = (aS)T = S(aT)$ for each pair S, T of elements of $E(V)$, for each real number a.

Axiom 12. There is a (unique) element I in $E(V)$ such that

$$IT = TI = T$$

for each $T \in E(V)$.

Proof. Axiom 9 is a consequence of Proposition 1.5. For Axiom 12, note that the function $I = I_V$ of Proposition 1.6 is linear and therefore in $E(V)$. The remaining axioms are verified by computation, using the definitions of operations in $E(V)$:

Axiom 10:

$$\begin{aligned}(S(T_1 + T_2))(X) &= S((T_1 + T_2)(X)) = S(T_1(X) + T_2(X)) \\ &= S(T_1(X)) + S(T_2(X)) = (ST_1)(X) + (ST_2)(X) \\ &= (ST_1 + ST_2)(X).\end{aligned}$$

$$\begin{aligned}((T_1 + T_2)S)(X) &= (T_1 + T_2)(S(X)) = T_1(S(X)) + T_2(S(X)) \\ &= (T_1S)(X) + (T_2S)(X) = (T_1S + T_2S)(X) .\end{aligned}$$

Axiom 11:

$$\begin{aligned}(a(ST))(X) &= a((ST))(X) = a(S(T(X))) \\ &= \begin{cases}(aS)(T(X)) = ((aS)T)(X) \\ S(a(T(X))) = S((aT)(X)) = (S(aT))(X).\end{cases}\end{aligned}$$

Remarks. $E(V)$ has many of the standard algebraic properties one is accustomed to using. But several important properties are missing. For example, in R^2 define endomorphisms S, T by

$$S(A_1) = A_2,\ S(A_2) = A_1; \quad T(A_1) = -A_1,\ T(A_2) = A_2 .$$

Then $ST(A_1) = -A_2$ and $TS(A_1) = A_2$; hence ST and TS are

different. Thus the commutative law for multiplication does not hold (except when V is of dimension 1). More important is the fact that $ST = \tilde{0}_E$ does not necessarily imply that one of the factors is $\tilde{0}_E$. As an example, let A_1, A_2 be a basis in R^2, and define an endomorphism T by $T(A_1) = A_2$ and $T(A_2) = \tilde{0}_V$. Clearly T is not $\tilde{0}_E$; however the composition of T with T does give $\tilde{0}_E$, i.e. $T^2 = TT = \tilde{0}_E$. This means that the algebra $E(V)$ does does not admit an operation of "division" having the usual properties, e.g. if $ST = \tilde{0}_E$ and $S \neq \tilde{0}_E$ we cannot divide by S and conclude that $T = \tilde{0}_E$ [unless S is an automorphism (see below)]. However, it is still true that $T\tilde{0}_E = \tilde{0}_E T = \tilde{0}_E$ for any $T \in E(V)$.

By Definition 1.6, an endomorphism which is an isomorphism is called an automorphism of V. The set of automorphisms of V will be denoted by $A(V)$.

The set $A(V)$ of bijective endomorphisms of V is a subset of $E(V)$, but is <u>not</u> a linear subsapce of $E(V)$. For example, $\tilde{0}_E$ is not in $A(V)$ (unless V consists of a single vector $\tilde{0}$). Also if S and T are automorphisms, $S + T$ need not be an automorphism (e.g. if $T = -S$). However, the composition TS of two automorphisms T, S is an automorphism, that is, $TS \in A(V)$. This follows from Proposition 1.5 (iii) and Theorem 2.1. With respect to this multiplication, $A(V)$ has the structure of a group and is called the multiplicative group of automorphisms of V or the <u>general linear group</u> of V.

9.2. <u>Theorem</u>. With multiplication in $A(V)$ defined by composition of automorphisms, the set $A(V)$ forms a multiplicative group; that is, the following axioms are satisfied:

Axiom G1. $T_1(T_2T_3) = (T_1T_2)T_3$ for each triple T_1, T_2, T_3 of elements of $A(V)$.

Axiom G2. There is a (unique) element $I \in A(V)$ such that

$$IT = TI = T$$

for each $T \in A(V)$.

Axiom G3. To each $T \in A(V)$ there corresponds a (unique) element, denoted by T^{-1}, such that

$$T^{-1}T = TT^{-1} = I.$$

Proof. The associativity of multiplication in $A(V)$ (Axiom G1) follows from the associativity of multiplication in $E(V)$. For Axiom G2, note that the identity $I \in E(V)$ is bijective and therefore an element of $A(V)$. For Axiom G3, we use Theorem 7.4 to conclude that the inverse function T^{-1} is an isomorphism and therefore an element of $A(V)$.

Remark. We can "divide" in $E(V)$ in the following sense. Let T be an automorphism, and suppose S_1, $S_2 \in E(V)$ are such that

$$TS_1 = S_2 .$$

Multiply both sides by T^{-1} on the left. Using Theorems 9.1 and 9.2, we obtain

$$T^{-1}(TS_1) = (T^{-1}T)S_1 = IS_1 = S_1 ;$$

that is, we have solved for $S_1 = T^{-1}S_2$. Similarly $S_1T = S_2$ can solved for $S_1 = S_2T^{-1}$. Thus the usual rule "one can divide by anything different from zero", must be replaced by "one can divide, on the left or on the right, by any automorphism". This means to

multiply by T^{-1} on the left or on the right.

9.3. Proposition. If $S, T \in A(V)$ and $a \in R$, $a \neq 0$, then $(ST)^{-1} = T^{-1}S^{-1}$ and $aT \in A(V)$ with $(aT)^{-1} = a^{-1}T^{-1}$.

Proof. The first result follows from

$$(T^{-1}S^{-1})(ST) = T^{-1}(S^{-1}S)T = T^{-1}IT = T^{-1}T = I ,$$

and Proposition 1.7. For the second, we compute in $E(V)$, since we do not yet have $aT \in A(V)$, to get

$$(a^{-1}T^{-1})(aT) = (a(a^{-1}T^{-1}))T = (aa^{-1})(T^{-1}T) = I$$

and

$$(aT)(a^{-1}T^{-1}) = (a^{-1}(aT))T^{-1} = (a^{-1}a)(TT^{-1}) = I ,$$

and then apply Exercise 3.2 to conclude that aT is bijective, therefore in $A(V)$, and that $(aT)^{-1} = a^{-1}T^{-1}$.

§10. Exercises

1. If $T \in L(V, W)$ and $\operatorname{im} T$ is finite dimensional, show that there is a linear subspace U of V which T transforms isomorphically onto $T(V)$. Show that each parallel (see Definition I, 12.1) of U intersects each parallel of $\ker T$ in precisely one vector.

2. Show that $T : R^2 \longrightarrow R^2$ is an isomorphism if and only if its matrix representation (with respect to the usual basis) has a non-zero determinant, i.e.

$$\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} \neq 0 .$$

3. An endomorphism $P : V \longrightarrow V$ is called a projection if $PP = P$. Show that $\ker P \cap \operatorname{im} P = \tilde{0}_V$. Show that each vector A of V is uniquely representable as the sum of a vector of

ker P and a vector of im P. Show that $Q = I - P$ is also a projection, and that ker Q = im P, im Q = ker P. Show that $PQ = QP = \tilde{0}_E$.

4. Let dim V be finite. Show that any endomorphism T of V can be expressed as a composition SP where P is a projection and S is an automorphism.

5. An endomorphism $J : V \longrightarrow V$ is called an <u>involution</u> if $JJ = I$. Show that, if P is a projection, then $J = I - 2P$ is an involution. What involution corresponds to the projection $Q = I - P$?

6. If S and T are affine transformations $V \longrightarrow W$, show that $S + T$ and aT, for $a \in R$, are affine. In this way the set $A(V, W)$ of all affine transformations is a vector space containing $L(V, W)$ as a linear subspace. If V and W have dimensions k and n, what is the dimension of $A(V, W)$? If we define

$$P : A(V, W) \longrightarrow L(V, W)$$

to be the function which assigns to each affine transformation S the linear transformation T defined by $T(X) = S(X) - S(\tilde{0})$, show that P is linear, and is a projection if considered as an endomorphism of $A(V, W)$. Find the dimension of ker P.

7. Show that the matrix representation of Proposition 8.2, in the case $W = V$, $B_i = A_i$, satisfies

$$M(TS) = M(T)M(S)$$

where the product on the right is the usual matrix multiplication.

8. Show that a product is defined in $E(V)$ by setting

$[S, T] = ST - TS$, for any pair of elements S, T in $E(V)$. Show that the "bracket product" satisfies (for any $S, T, Y \in A(V)$, $a \in R$):

$$[S + T, Y] = [S, Y] + [T, Y] ,$$

$$a[S, T] = [aS, T] = [S, aT]$$

$$[S, S] = \tilde{0}_E ,$$

$$[S, T] = - [T, S] ,$$

$$[[S, T], Y] + [[T, Y], S] + [[Y, S], T] = \tilde{0}_E .$$

The last result is called the "Jacobi identity" and shows that this product is not associative. The vector space $E(V)$, with the bracket product, forms a "Lie algebra", the Lie algebra of endomorphisms of V.

§11. Quotient; direct sum

11.1. Definition. If U is a linear subspace of the vector space V, the quotient of V by U, denoted by V/U, is the vector space whose elements are the equivalence classes of V modulo U (Definition I, 12.6) and for which the operations of addition and scalar multiplication are defined by the condition that the function

$$j : V \longrightarrow V/U ,$$

which assigns to each $A \in V$ the equivalence class in which it lies, be a linear transformation.

Proof. The properties of the equivalence relation show that j is surjective and that $j(A) = j(B)$ if and only if $B = A + X$ for some $X \in U$. If the set V/U can be made into a vector space such that j is linear, we must have

(1) $\quad j(A) + j(B) = j(A + B) \qquad$ for all $A, B \in V$,

(2) $\quad xj(A) = j(xA) \qquad$ for all $A \in V$, $x \in R$,

by Definition II, 1.1. It remains to verify that these conditions do indeed determine an addition, and a multiplication by scalars, in V/U and that operations so defined satisfy the axioms given in I, 1.1. The possibility of ambiguity in the definition of operations in V/U arises from the fact that, although $A \in V$ determines a unique equivalence class $j(A) \in V/U$, an equivalence class in V/U does not determine a unique antecedent in V. Let $E = j(A)$, $F = j(B)$ be elements of V/U. In order that the element $E + F$ be well-defined by (1), it is necessary that the equivalence class $j((A + X) + (B + Y))$ coincide with the equivalence class $j(A + B)$ for any $X, Y \in U$. This follows from

$$j((A + X) + (B + Y)) = j(A + B + X + Y) = j(A + B)$$

since <u>addition</u> in V <u>is commutative</u> and $X + Y \in U$ (U is a linear subspace). Similarly, xE, for $x \in R$, is well-defined by (2) since

$$j(x(A + X)) = j(xA + xX) = j(xA)$$

for any $X \in U$. Clearly, $\tilde{0}_{V/U}$ is the equivalence class $j(\tilde{0}_V) = j(X)$, $X \in U$. The verification of the axioms is left as an exercise.

11.2. <u>Proposition</u>. The kernel of j is U and, if V is finite dimensional,

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

If $U = \tilde{0}_V$, then $j : V \longrightarrow V/U$ is an isomorphism; if $U = V$, then $V/U = \tilde{0}$.

The proof of this proposition is left as an exercise.

11.3. Proposition. Let V, W be vector spaces, and let U be a linear subspace of V, $j: V \longrightarrow V/U$. Let $T \in L(V, W)$. Then there exists a $T' \in L(V/U, W)$, such that $T = T'j$, if and only if $U \subset \ker T$. Further, T' is surjective if T is.

Proof. If $T = T'j$ then, for $X \in U$, $T(X) = T'j(X) = \tilde{0}_W$ since $j(X) = \tilde{0}_{V/U}$; i.e. $U \subset \ker T$. Conversely, if $U \subset \ker T$, let

$$T'(E) = T(A), \quad E \in V/U, \quad E = j(A), \quad A \in V.$$

Then T' is well-defined since $T(A + X) = T(A)$, $X \in U$. It is easily verified that T' is linear and it is obvious that $T = T'j$, and that T' is surjective if T is.

11.4. Definition. If U is a linear subspace of the vector space V, then the inclusion $\iota = \iota_U : U \longrightarrow V$ is the injective linear transformation which assigns to each element of U the same element considered as an element of V.

For given V and linear subspace U, the pair ι, U is "dual" to the pair j, V/U. For example, we have the following two propositions, the proofs of which are left as an exercise.

11.5. Proposition. Let V, W be vector spaces, and let U be a linear subspace of V, $\iota: U \longrightarrow V$. Let $T \in L(W, V)$. Then there exists a $T' \in L(W, U)$, such that $T = \iota T'$, if and only if $\operatorname{im} T \subset U$. Further, T' is injective if T is.

11.6. Proposition. Let U be a linear subspace of the vector space V, $\iota : U \longrightarrow V$, $j : V \longrightarrow V/U$.

(i) If $T \in L(V, U)$ satisfies $T\iota = I_U$, then T is surjective, i.e. $\operatorname{im} T = U$, and $P = \iota T \in E(V)$ is a projection (see Exercise 10.5);

(ii) if $S \in L(V/U, V)$ satisfies $jS = I_{V/U}$, then S is injective (and therefore gives an isomorphism of V/U with the linear subspace $\operatorname{im} S \subset V$), and $P = Sj \in E(V)$ is a projection.

11.7. Definition. Let U and W be linear subspaces of a vector space V. Then U and W give a direct sum decomposition of V, or $V = U \oplus W$, if and only if

(i) $U \cap W = \tilde{0}_V$,

(ii) for each $B \in V$, we have $B = A + C$, for some $A \in U$, $C \in W$.

Remarks. By (i) the decomposition in (ii) is unique. The direct sum decomposition is trivial if $U = \tilde{0}_V$ (so $W = V$) or if $W = \tilde{0}_V$ (so $U = V$).

11.8. Definition. If U and W are arbitrary vector spaces, the direct sum of U and W is the vector space V constructed as follows. The elements B of V are ordered pairs (A, C) where $A \in U$, $C \in W$. Addition and scalar multiplication in V are defined by

$$B + B' = (A, C) + (A', C') = (A + A', C + C')$$
$$xB = x(A, C) = (xA, xC) , \qquad x \in R .$$

Then $\tilde{0}_V = (\tilde{0}_U, \tilde{0}_W)$.

The transformation $\iota_U \in L(U, V)$ defined by

$$\iota_U(A) = (A, \tilde{0}_W) , \qquad A \in U ,$$

is injective and gives an isomorphism of U with the linear subspace $\iota_U(U) \subset V$ by which U may be identified with $\iota_U(U)$. Similarly, if $\iota_W \in L(W, V)$ is defined by

$$\iota_W(C) = (\tilde{0}_U, C) , \qquad C \in W ,$$

the vector space W may be identified with the linear subspace $\iota_W(W) \subset V$. After the identification we have $V = U \oplus W$ with $B = (A, C) = (A, \tilde{0}_W) + (\tilde{0}_V, C)$.

The next two propositions are essentially Exercise 10.5.

11.9. <u>Proposition</u>. If $V = U \oplus W$, let $P_U \in E(V)$ be defined by

$$P_U(B) = A\ , \qquad \text{if } B = A + C,\ A \in U,\ C \in W.$$

Then P_U is a projection: $P_U P_U = P_U$, and $\operatorname{im} P_U = U$, $\ker P_U = W$. Similarly, if P_W is defined by

$$P_W(B) = C\ , \qquad \text{if } B = A + C,\ A \in U,\ C \in W,$$

then P_W is a projection, and $\operatorname{im} P_W = W$, $\ker P_W = U$. We have

$$P_U + P_W = I, \quad P_U P_W = P_W P_U = \tilde{0}_E\ .$$

11.10. <u>Proposition</u>. If $P \in E(V)$ is a projection, then

$$V = \operatorname{im} P \oplus \ker P$$

is a direct sum decomposition of V.

Next we suppose that a linear subspace U of the vector space V is given. What can be said about direct sum compositions of V with U as a direct summand?

(1) Note that giving U determines $\iota_U : U \longrightarrow V$ but does <u>not</u> determine P_U. Each choice of $P \in E(V)$ with $PP = P$ and $\operatorname{im} P = U$ (or of $T \in L(V, U)$ with $T\iota_U = I_U$, $P = \iota_U T$) determines a direct sum decomposition $V = U \oplus W$ by taking $W = \ker P$, and for this decomposition $P_U = P$.

(2) Each choice of $P \in E(V)$ with $PP = P$ and $\ker P = U$ (or of $S \in L(V/U, V)$ with $jS = I_{V/U}$, $P = Sj$, where

$j : V \longrightarrow V/U$) determines a direct sum decomposition $V = U \oplus W$ by taking $W = \text{im } P$, and for this decomposition $P_W = P$.

11.11. <u>Theorem</u>. If U is a linear subspace of a vector space V, then a direct sum decomposition of the form $V = U \oplus W$ always exists.

The proof of existence will be omitted in the case that V/U is infinite dimensional. If V/U is finite dimensional, the proof of existence (of S) is essentially the same as Exercise 10.1. From any method of construction used to prove existence, it is clear that different choices of W depend on the non-uniqueness of the process of completing a basis for U to a basis for V (although two different completions may determine the same W), if V is finite dimensional, or of choosing A such that $j(A) = E \in V/U$ for given E, in the more general case. However, we always have W isomorphic to V/U, e.g. $P_W S : V/U \longrightarrow W$ can be shown to be an isomorphism if we start the construction from S (see also Corollary 12.6).

11.12. <u>Definition</u>. Let U_i, $i = 1, \ldots, s$, be linear subspaces of a vector space V. Then the U_i give a <u>direct sum decomposition</u> of V, or $V = U_1 \oplus U_2 \oplus \ldots \oplus U_s$, if and only if

(i) $U_i \cap U_j = \tilde{0}_V$, $i \neq j$,

(ii) for each $B \in V$, we have $B = A_1 + A_2 + \ldots + A_s$ for some choice of $A_i \in U_i$, $i = 1, \ldots, s$.

<u>Remarks</u>. For $s = 2$, the above definition is the same as Definition 11.7. It is again true that the condition (i) implies that the decomposition in (ii) is unique.

<u>Example</u>. Let V be a finite dimensional vector space,

dim $V = n$, and let $C_1, \ldots, C_n$ be a basis for V. Let U_i be the one-dimensional linear subspace consisting of all vectors of the form xC_i, $x \in R$. Then the U_i, $i = 1, \ldots, n$, give a direct sum decomposition of V.

The next two propositions are the generalizations of Propositions 11.9 and 11.10.

11.13. <u>Proposition</u>. Let $V = U_1 \oplus U_2 \oplus \ldots \oplus U_s$ and define $P_i \in E(V)$ by

$$P_i(B) = A_i , \qquad \text{if } B = A_1 + A_2 + \ldots + A_s,\ A_j \in U_j .$$

Then

$$(3) \qquad P_iP_i = P_i , \qquad i = 1, \ldots, s ,$$

that is, P_i is a projection, with $\operatorname{im} P_i = U_i$, and

$$(4) \qquad P_iP_j = \tilde{0}_E , \qquad i \neq j ,$$

$$(5) \qquad P_1 + P_2 + \ldots + P_s = I .$$

11.14. <u>Proposition</u>. If $P_i \in E(V)$, $i = 1, \ldots, s$, satisfy the conditions (3), (4), (5) of Proposition 11.13, then a direct sum decomposition of V is given by $U_i = \operatorname{im} P_i$, $i = 1, \ldots, s$.

11.15. <u>Proposition</u>. Let V, U_i and P_i, $i = 1, \ldots, s$, be as in Proposition 11.13. Define $Q_i \in E(V)$, $i = 1, \ldots, s$, by $Q_i = I - P_i$ and let $W_i = \operatorname{im} Q_i = \ker P_i$. Then

$$V = U_i \oplus W_i \qquad \text{for each choice of } i = 1, \ldots, s.$$

Let $Q_{ij} = Q_iQ_j$, $P_{ij} = P_i + P_j$, $i \neq j$. Then Q_{ij} and P_{ij} are projections and $Q_{ij} = I - P_{ij}$. Thus, if we set $W_{ij} = \operatorname{im} Q_{ij}$, $U_{ij} = \operatorname{im} P_{ij}$, we have

$$V = U_{ij} \oplus W_{ij} .$$

Moreover, $U_{ij} = U_i \oplus U_j$. Analogous constructions may be carried out for arbitrary groupings of the direct summands of V, and each grouping is a direct sum of its components. In particular, W_i is the direct sum of the U_j, $j \neq i$; W_{ij} is the direct sum of the U_k, $k \neq i$, $k \neq j$, etc.

Proof. The properties of the endomorphisms Q_{ij}, etc., are trivial consequences of the formulas (3), (4), (5) of Proposition 11.13. To show that $U_{ij} = U_i \oplus U_j$, for example, we note first that $U_i \subset U_{ij}$ and $U_j \subset U_{ij}$. In fact, $U_{ij} = \operatorname{im} P_{ij} = \ker Q_{ij} = \ker (Q_i Q_j) \supset \ker Q_j = \operatorname{im} P_j = U_j$. The result for U_i follows similarly, using $Q_i Q_j = Q_j Q_i$. We already have $U_i \cap U_j = \tilde{0}_V$ by (i) of Definition 11.12. Finally, if $B \in U_{ij} = \operatorname{im} P_{ij}$, there is a $C \in V$ such that $B = P_{ij}(C) = P_i(C) + P_j(C)$ where $P_i(C) \in \operatorname{im} P_i = U_i$, $P_j(C) \in \operatorname{im} P_j = U_j$.

11.16. Proposition. If $V = U_1 \oplus U_2 \oplus \dots \oplus U_s$, where V is finite dimensional, then

$$\dim V = \dim U_1 + \dim U_2 + \dots + \dim U_s .$$

Proof. For $s = 2$, this is a consequence of Proposition 11.2 and the fact that U_2 is isomorphic to V/U_1. For $s > 2$, the result is proved by induction, using the grouping properties described in Proposition 11.15.

§12. Exact sequences

In this section we shall write $V \xrightarrow{T} W$ in place of $T : V \longrightarrow W$. The explicit mention of the linear transformation T is omitted if either the domain or the range of T is the

vector space consisting of the single element $\tilde{0}$. In fact, $\tilde{0} \longrightarrow V$ is unambiguous since $L(\tilde{0}, V)$ consists of a single (injective) linear transformation. Analogously, $V \longrightarrow \tilde{0}$ is unambiguous, since $L(V, \tilde{0})$ contains only one (surjective) transformation; the kernel of $V \longrightarrow \tilde{0}$ is V.

12.1. <u>Definition</u>. Let U, V, W be vector spaces and let $S \in L(U, V)$, $T \in L(V, W)$. The sequence

$$U \xrightarrow{S} V \xrightarrow{T} W$$

is called <u>exact</u> (at V) if $\operatorname{im} S = \ker T$.

<u>Remark</u>. Then TS must be $\tilde{0}_{L(U,W)}$. The condition $TS = \tilde{0}$ is necessary, but not sufficient, for exactness; this condition is equivalent to $\operatorname{im} S \subset \ker T$.

The next two propositions are restatements of Proposition 7.1, Definition 1.3 (ii), and Proposition 11.2. (Similarly, the propositions stated below without proof are restatements of earlier propositions.)

12.2. <u>Proposition</u>. Let $T \in L(V, W)$. Then

(i) T is injective if and only if the sequence

$$\tilde{0} \longrightarrow V \xrightarrow{T} W$$

is exact at V;

(ii) T is surjective if and only if the sequence

$$V \xrightarrow{T} W \longrightarrow \tilde{0}$$

is exact at W;

(iii) T is an isomorphism (bijective) if and only if the sequence

$$\tilde{0} \longrightarrow V \xrightarrow{T} W \longrightarrow \tilde{0}$$

is exact, i.e. exact at V and at W.

12.3. <u>Proposition</u>. If U is a linear subspace of the vector space V, then the sequence

$$\tilde{0} \longrightarrow U \xrightarrow{\iota} V \xrightarrow{j} V/U \longrightarrow \tilde{0}$$

is exact.

12.4. <u>Proposition</u>. If

$$\tilde{0} \longrightarrow U \xrightarrow{S} V \xrightarrow{T} W \longrightarrow \tilde{0}$$

is exact, then U is isomorphic to $\ker T$, and W is isomorphic to $V/\text{im } S$.

<u>Proof</u>. Exactness at U implies that $S : U \longrightarrow \text{im } S$ is an isomorphism and exactness at V implies $\text{im } S = \ker T$. Thus S is an isomorphism of U with $\ker T$. Consider next the transformation

$$V \xrightarrow{j} V/\text{im } S.$$

We have $\ker j = \text{im } S = \ker T$, so by Proposition 11.3, there is a $T' \in L(V/U, W)$ such that $T = T'j$ and T' is surjective. Suppose $E \in \ker T'$, i.e. $T'(E) = \tilde{0}$, and let $E = j(A)$, $A \in V$. Then $T(A) = T'j(A) = \tilde{0}$, so $A \in \ker T = \text{im } S$ which implies $E = j(A) = \tilde{0}$. Thus $\ker T' = \tilde{0}$ and T' is an isomorphism.

12.5. <u>Proposition</u>. If $V = U \oplus W$ and if P_U(Proposition 11.9) is considered as an element of $L(V, U)$ (that is, if P_U now denotes P_U' of Proposition 11.5), and if P_W is considered as an element of $L(V, W)$, then the sequences

$$\tilde{0} \longrightarrow U \xrightarrow{\iota_U} V \xrightarrow{P_W} W \longrightarrow \tilde{0}$$

and

$$\tilde{0} \longrightarrow W \xrightarrow{\iota_W} V \xrightarrow{P_U} U \longrightarrow \tilde{0}$$

are exact. Further, $P_U\iota_U = I_U$ and $P_W\iota_W = I_W$.

12.6. <u>Corollary</u>. If $V = U \oplus W$, then W is isomorphic to V/U.

12.7. <u>Definition</u>. A <u>splitting</u> of the exact sequence

$$\tilde{0} \longrightarrow U \xrightarrow{\iota} V \xrightarrow{j} V/U \longrightarrow \tilde{0} \quad ,$$

is an exact sequence

$$\tilde{0} \longrightarrow V/U \xrightarrow{S} V \xrightarrow{T} U \longrightarrow \tilde{0}$$

such that $jS = I_{V/U}$ and $T\iota = I_U$.

<u>Remark</u>. It is not necessary to give both T and S since each determines the other.

12.8. <u>Proposition</u>. If U is a linear subspace of a vector space V, then there is a bijective correspondence between the direct sum decompositions of the form $V = U \oplus W$ and the splittings of the exact sequence

$$\tilde{0} \longrightarrow U \xrightarrow{\iota} V \xrightarrow{j} V/U \longrightarrow \tilde{0} \ .$$

12.9. <u>Corollary</u> (to Theorem 11.11). An exact sequence of linear transformations of vector spaces always splits.

<u>Remark</u>. The earlier definitions and propositions about exact sequences retain their meaning if linear transformations of vector spaces are replaced by homomorphisms of sets with a given type of algebraic structure, provided there is enough algebraic structure to define the kernel of a homomorphism and to define a quotient structure. However, in these other cases, splittings and direct sum decompositions need not exist in general.

12.10. <u>Lemma</u>. Let U, V, W be vector spaces and let $S \in L(U, V)$. Then S induces a linear transformation

$$S^* : L(V, W) \longrightarrow L(U, W)$$

where $T \in L(V, W) \longrightarrow TS \in L(U, W)$ (Proposition 8.3). Then

(i) $T \in L(V, W)$ is an element of $\ker S^*$ if and only if $\operatorname{im} S \subset \ker T \subset V$, i.e.

$$\ker S^* = \{T \mid T \in L(V, W) \text{ and } \operatorname{im} S \subset \ker T \subset V\} ;$$

(ii) $\tilde{T} \in L(U, W)$ is an element of $\operatorname{im} S^*$ if and only if $\ker S \subset \ker \tilde{T} \subset U$, i.e.

$$\operatorname{im} S^* = \{\tilde{T} \mid \tilde{T} \in L(U, W) \text{ and } \ker S \subset \ker \tilde{T} \subset U\} .$$

<u>Proof</u>. (i) If $T \in L(V, W)$ satisfies $\ker T \supset \operatorname{im} S$, it is obvious that $S^*(T) = TS$ is the zero transformation of $L(U, W)$; that is, $T \in \ker S^*$. If $\ker T \not\supset \operatorname{im} S$, then there is a vector $B \in \operatorname{im} S$ for which $T(B) \neq \tilde{0}_W$. Choose any $A \in U$ such that $S(A) = B \in \operatorname{im} S$. Then $S^*(T)(A) = TS(A) = T(B) \neq \tilde{0}_W$, so $S^*(T)$ is not the zero transformation of $L(U, W)$; that is $T \notin \ker S$.

(ii) If $\tilde{T} \in L(U, W)$ is such that $\tilde{T} \in \operatorname{im} S^*$, then there is a $T \in L(V, W)$ such that $\tilde{T} = TS$. Then $\tilde{T}(A) = TS(A) = \tilde{0}_W$ if $S(A) = \tilde{0}_V$, or $A \in \ker S$; that is, $\ker S \subset \ker \tilde{T}$. To show that $\tilde{T} \in \operatorname{im} S^*$, given that $\tilde{T} \in L(U, W)$ and $\ker S \subset \ker \tilde{T}$, we must construct a $T \in L(V, W)$ such that $\tilde{T} = TS$. The construction depends on the existence of a direct sum decomposition

$$V = \operatorname{im} S \oplus V' ,$$

that is, on Theorem 11.11, unless $\operatorname{im} S = V$ or $\operatorname{im} S = \tilde{0}_V$. Let $\iota_1 = \iota_{\operatorname{im} S}$ and let $P = P_{\operatorname{im} S}$; then $P\iota_1 = I_{\operatorname{im} S}$. Let S_1 denote the surjective transformation obtained by considering $S \in L(U, V)$ as an element of $L(U, \operatorname{im} S)$; that is, $S = \iota_1 S_1$. Then $\ker S_1 = \ker S$. Note also that $\operatorname{im} S$ is isomorphic to $U/\ker S_1$, by Proposition 12.3. The condition $\ker S_1 = \ker S \subset \ker \tilde{T}$

is used with Proposition 11.3 to show that there is a $T' \in L(\text{im } S, W)$ such that $\tilde{T} = T'S_1$. Define $T \in L(V, W)$ by $T = T'P$. Then

$$TS = T'P\iota_1 S_1 = T'S_1 = \tilde{T} \quad .$$

12.11. Corollary. If S of Lemma 12.10 is surjective, then S^* is injective; if S is injective, then S^* is surjective; if S is an isomorphism, then S^* is an isomorphism.

Proof. If S is surjective, then $\text{im } S = V$. Then (i) of Lemma 12.10 shows that $\ker T = V$ for every $T \in \ker S^*$, or T is the zero transformation in $L(V, W)$; that is, $\ker S^* = \tilde{0}$. If S is injective, then $\ker S = \tilde{0}_U$ and the condition in (ii) of Lemma 12.10 is satisfied by every $T \in L(U, W)$; that is, $\text{im } S^* = L(U, W)$.

12.12. Theorem. Let U, V, W, Z be vector spaces and let

$$\tilde{0} \longrightarrow U \xrightarrow{S} V \xrightarrow{T} W \longrightarrow \tilde{0}$$

be an exact sequence. Then the induced sequence

$$\tilde{0} \longrightarrow L(W, Z) \xrightarrow{T^*} L(V, Z) \xrightarrow{S^*} L(U, Z) \longrightarrow \tilde{0}$$

is also exact.

Proof. The only part not covered by Corollary 12.11 is exactness at $L(V, Z)$, i.e. $\text{im } T^* = \ker S^*$. By Lemma 12.10 we have

$$\text{im } T^* = \{\tilde{T} | \tilde{T} \in L(V, Z) \quad \text{and} \quad \ker T \subset \ker \tilde{T} \subset V\}$$

and

$$\ker S^* = \{\tilde{T} | \tilde{T} \in L(V, Z) \quad \text{and} \quad \text{im } S \subset \ker \tilde{T} \subset V\} \ ,$$

and these two sets are the same since $\operatorname{im} S = \ker T$ by hypothesis.

12.13. <u>Corollary</u>. If U, V, W are vector spaces and the sequence

$$\tilde{0} \longrightarrow U \xrightarrow{S} V \xrightarrow{T} W \longrightarrow \tilde{0}$$

is exact, then the dual sequence (see Definition 8.4)

$$\tilde{0} \longrightarrow W^* \xrightarrow{T^*} V^* \xrightarrow{S^*} U^* \longrightarrow \tilde{0}$$

is also exact.

<u>Proof</u>. Take $Z = R$ in Corollary 12.12.

12.14. <u>Corollary</u>. If U is a linear subspace of the vector space V, then $(V/U)^*$ is isomorphic to V^*/U^*.

<u>Proof</u>. Apply Corollary 12.13, and then Corollary 12.9 and Proposition 12.4 to the exact sequence

$$\tilde{0} \longrightarrow U \xrightarrow{\iota} V \xrightarrow{j} V/U \longrightarrow \tilde{0} .$$

12.15. <u>Corollary</u>. If $V = U \oplus W$, then $V^* = W^* \oplus U^*$.

<u>Proof</u>. Applying Corollary 12.13 to the exact sequence

$$\tilde{0} \longrightarrow U \xrightarrow{\iota_U} V \xrightarrow{P_W} W \longrightarrow \tilde{0}$$

of Proposition 12.5, which splits, we obtain the exact sequence

$$\tilde{0} \longrightarrow W^* \xrightarrow{P_W^*} V^* \xrightarrow{\iota_U^*} U^* \longrightarrow \tilde{0}$$

which also splits. For example, verify that $P_U^* : U^* \longrightarrow V^*$ satisfies $\iota_U^* P_U^* = I_{U^*}$.

Arthur L. Besse

Einstein Manifolds

Reprint of the 1987 Edition

With 22 Figures

Springer

K. Tables

Table 1. The Eight Candidates to a Holonomy Fellowship

$G \subset SO(n)$		$\dim G$	$\mathrm{Norm}_{O(n)} G$	Signification of $\mathrm{Hol}(g) \subset G$	Ricci curvature	Algebra of invariant exterior differential forms	Existence			Existence of non-trivial compact deformations
							Local	Complete	Compact	
(I) $SO(n)$		$\frac{n(n-1)}{2}$	$O(n)$			None	Generic Riemannian manifold			Yes
(II) $U(n)$	$2m$	m^2	$U(m)$	Equivalent to be Kähler		The Kähler form ω	Generic Kähler manifold			Yes
(III) $SU(m)$	$2m$	$m^2 - 1$	$U(m)$	Equivalent to Ricci flat Kähler	Ricci flat	The Kähler form ω, the complex volume form and its dual	Yes	Yes	Yes	Yes
(IV) $Sp(1) \cdot Sp(m)$	$4m$	$m^2 + 2m + 3$	$Sp(1) \cdot Sp(m)$	Equivalent to quaternion-Kähler	Einstein	The 4-form θ	Yes	Some homogeneous by Aleksevskii	Yes but only standards known	?
(V) $Sp(m)$	$4m$	$m^2 + 2m$	$Sp(1) \cdot Sp(m)$	Equivalent to hyperkälherian	Ricci flat	Probably three 2-forms	Yes	Yes	Yes	Yes
(VI) Spin(9)	16	36	Spin(9)	?	Einstein	One 8-form	Necessarily isometric (locally) to $(\mathrm{Ca}P^2, \mathrm{can})$ or to its non-compact dual			No
(VII) Spin(7)	8	21	Spin(7)	?	Ricci flat	One 4-form	Yes	?	?	?
(VIII) G_2	7	14	G_2	?	Ricci flat	One 3-form and its dual	Yes	?	?	?

Table 2. Irreducible Symmetric Spaces of Type I

Helgason's type	G	H	$\dim G/H$	rank	Isotropy representation[1])	Kähler or not	Geometric realization
A I	$SU(n)$	$SO(n)$	$\frac{(n-1)(n+2)}{2}$	$n-1$	$\wedge^p SO(n) \underline{\otimes} \wedge^p SO(n)$ $p = 2$ if n odd $p = 3$ if n even	No	Set of the $\mathbb{R}P^{n-1}$'s in $\mathbb{C}P^{n-1}$ or set of the real structures of $\mathbb{C}^n$ (which leave invariant the complex determinant)
A II	$SU(2n)$	$Sp(n)$	$(n-1)(2n+1)$	$n-1$	$\underline{\wedge}^2 Sp(n)$	No	Set of quaternionic structures of $\mathbb{C}^{2n}$ compatible with its Hermitian structure or set of the metric compatible fibrations $S^3 \to \mathbb{C}P^{2n-1} \to \mathbb{H}P^{n-1}$
A III	$SU(p+q)$	$S(U(p) \times U(q))$ $p \leqslant q$	$2pq$	$\min(p,q)$	$S(U(p) \otimes U(q))$	Yes	Complex p-Grassman manifold of $\mathbb{C}^{p+q}$ (in particular $\mathbb{C}P^q$ if $p = 1$) or set of the $\mathbb{C}P^{p-1}$'s in $\mathbb{C}P^{p+q-1}$
BD I	$SO(p+q)$	$SO(p) \times SO(q)$	pq	$\min(p,q)$	$SO(p) \otimes SO(q)$	Yes if and only if $p = 2$	real p-Grassman manifold of $\mathbb{R}^{p+q}$ (in particular $\mathbb{R}P^q$ if $p = 1$) or set of the $\mathbb{R}P^{p-1}$'s in $\mathbb{R}P^{p+q-1}$
D III	$SO(2n)$	$U(n)$	$n(n-1)$	$\left[\frac{n}{2}\right]$	$\wedge^2 U(n)$	Yes	Set of complex structures of $\mathbb{R}^{2n}$ compatible with its Euclidean structures or set of the metric-compatible fibrations $S^1 \to \mathbb{R}P^{2n-1} \to \mathbb{C}P^{n-1}$

C I	$Sp(n)$	$U(n)$	$n(n+1)$	n	$U(n) \otimes U(n)$	Yes	Set of the $\mathbb{C}P^{n-1}$'s in $\mathbb{H}P^{n-1}$ or set of the complex structures of $\mathbb{H}^n$
C II	$Sp(p,q)$	$Sp(p) \times Sp(q)$ $p \leqslant q$	$4pq$	$\min(p,q)$	$Sp(p) \otimes Sp(q)$	No	Quaternionic p-Grassman manifold of $\mathbb{H}^{p+q}$ (in particular $\mathbb{H}P^q$ if $p=1$) or set of the $\mathbb{H}P^{p-1}$'s of $\mathbb{H}P^{p+q-1}$
E I	E_6	$Sp(4)$	42	6	$\Lambda^4 Sp(4)$	No	Antichains of $(\mathbb{C} \otimes \mathbb{O})P^2$
E II	E_6	$SU(6) \times SU(2)$	40	4	$\Lambda^3 SU(6) \otimes SU(2)$	No	Set of the $(\mathbb{C} \otimes \mathbb{H})P^2$'s in $(\mathbb{C} \otimes \mathbb{O})P^2$
E III	E_6	$SO(10) \times SO(2)$	32	2	$\mathrm{Spin}(10) \otimes SO(2)$	Yes	Rosenfeld's elliptic projective plane $(\mathbb{C} \otimes \mathbb{O})P^2$
E IV	E_6	F_4	26	2	F_4	No	Set of the $\mathbb{O}P^2$'s in $(\mathbb{C} \otimes \mathbb{O})P^2$
E V	E_7	$SU(8)$	70	7	$\Lambda^4 SU(8)$	No	Antichains of $(\mathbb{H} \otimes \mathbb{O})P^2$
E VI	E_7	$SO(12) \times SU(2)$	64	4	$\mathrm{Spin}(12) \otimes SU(2)$	No	Rosenfeld's elliptic projective plane $(\mathbb{H} \otimes \mathbb{O})P^2$
E VII	E_7	$E_6 \times SO(2)$	54	3	$E_6 \otimes SO(2)$	Yes	Set of the $(\mathbb{C} \otimes \mathbb{O})P^2$'s in $(\mathbb{H} \otimes \mathbb{O})P^2$
E VIII	E_8	$SO(16)$	128	8	$\mathrm{Spin}(16)$	No	Rosenfeld's elliptic projective plane $(\mathbb{O} \otimes \mathbb{O})P^2$ [illegible]
E IX	E_8	$E_7 \times SU(2)$	112	4	$\Lambda^2 E_7 \otimes SU(2)$	No	Set of the $(\mathbb{H} \otimes \mathbb{O})P^2$ in "$(\mathbb{O} \otimes \mathbb{O})P^2$" [illegible]

Table 2 (continued)

Helgason's type	G	H	$\dim G/H$	rank	Isotropy representation[1])	Kähler or not	Geometric realization
F I	F_4	$SP(3) \times SU(2)$	28	4	$\underline{\wedge}^3 Sp(3) \otimes SU(2)$	No	Set of the $\mathbb{H}P^2$'s in $\mathbb{C}aP^2$
F II	F_4	$SO(9)$	16	1	Spin(9)	No	Cayley elliptic projective plane $\mathbb{C}aP^2$
G I	G_2	$SU(2) \times SU(2)$	8	2	$\underline{\otimes}^3 SU(2) \otimes SU(2)$	No	Set of the quaternionic sub-algebras of $\mathbb{C}a$

[1]) here $\wedge$ (resp. $\otimes$) denotes the exterior (resp. tensor) product representation and $\underline{\wedge}$ (resp. $\underline{\otimes}$) denotes the natural irreducible representation deduced from it

[2]) up to this day an algebraic definition of this projective plane over $\mathbb{C}a \otimes \mathbb{C}a$ seems pending, see [Fre] and [Ros]

Table 3. Irreducible Symmetric Spaces of Type III

Helgason's type	G	H	$\dim G/H$	rank	Isotropy representation[1])	Kähler or not	Geometric interpretation
A I	$SL(n,\mathbb{R})$	$SO(n)$	$\frac{(n-1)(n+2)}{2}$	$n-1$	$\wedge^p SO(n) \underline{\otimes} \wedge^p SO(n)$ $p=2$ if n odd $p=3$ if n even	No	Set of Euclidean structures on $\mathbb{R}^n$ or set of the $\mathbb{R}P^n_{hyp}$'s in $\mathbb{C}P^n_{hyp}$
A II	$SU^*(2n) = SL(n,\mathbb{H})$	$Sp(n)$	$(n-1)(2n+1)$	$n-1$	$\underline{\wedge}^2 Sp(n)$	No	Set of the $\mathbb{H}P^{n-1}_{hyp}$'s in $\mathbb{C}P^{2n-1}_{hyp}$
A III	$SU(p,q)$	$S(U(p)\times U(q))$ $p \leqslant q$	$2pq$	$\min(p,q)$	$S(U(p)\otimes U(q))$	Yes	Grassman manifold of positive definite $\mathbb{C}^p$'s in $\mathbb{C}^{p,q}$, or set of the $\mathbb{C}P^{p-1}_{hyp}$'s in $\mathbb{C}P^{p+q+1}_{hyp}$ (in particular, complex hyperbolic space $\mathbb{C}P^q_{hyp}$ if $p=1$)
BD I	$SO_0(p,q)$	$SO(p)\times SO(q)$ $p \leqslant q$	pq	$\min(p,q)$	$SO(p)\otimes SO(q)$	Yes if and only if $p=2$	Grassman manifold of positive definite $\mathbb{R}^p$'s in $\mathbb{R}^{p,q}$, or set of the $\mathbb{R}P^{p-1}_{hyp}$'s in $\mathbb{R}P^{p+q+1}_{hyp}$ (in particular, real hyperbolic space $\mathbb{R}P^q_{hyp}$—denoted by H^q in 1.37—if $p=1$)
D III	$SO^*(2n) = SO(n,\mathbb{H})$	$U(n)$	$n(n-1)$	$[n/2]$	$\wedge^2 U(n)$	Yes	Set of quaternionic quadratic forms on $\mathbb{R}^{2n}$, or set of the $\mathbb{C}P^{n-1}_{hyp}$'s in $\mathbb{R}P^{2n-1}_{hyp}$
C I	$Sp(n,\mathbb{R})$	$U(n)$	$n(n+1)$	n	$U(n)\underline{\otimes}U(n)$	Yes	Set of Lagrangian subspaces of $\mathbb{R}^{2n}$ or set of the $\mathbb{C}P^n_{hyp}$'s in $\mathbb{H}P^n_{hyp}$
C II	$Sp(p,q)$	$Sp(p)\times Sp(q)$ $p \leqslant q$	$4pq$	$\min(p,q)$	$Sp(p)\otimes Sp(q)$	No	Grassman manifold of positive definite $\mathbb{H}^p$'s in $\mathbb{H}^{p,q}$, or set of the $\mathbb{H}P^{p-1}_{hyp}$'s in $\mathbb{H}P^{p+q+1}_{hyp}$ (in particular, quaternionic hyperbolic space $\mathbb{H}P^q_{hyp}$ if $p=1$)

Table 3 (continued)

Helgason's type	G	H	dim G/H	rank	Isotropy representation[1])	Kähler or not	Geometric interpretation
E I	E_6^6	$Sp(4)$	42	6	$\underline{\wedge}^4 Sp(4)$	No	Anti-chains of $(\mathbb{C} \otimes \mathbb{C}a)P^2_{hyp}$
E II	E_6^2	$SU(6) \times SU(2)$	40	4	$\wedge^3 SU(6) \otimes SU(2)$	No	Set of the $(\mathbb{C} \otimes \mathbb{H})P^2_{hyp}$'s in $(\mathbb{C} \otimes \mathbb{C}a)P^2_{hyp}$
E III	E_6^{-14}	$SO(10) \times SO(2)$	32	2	$\mathrm{Spin}(10) \cdot SO(2)$	Yes	Rosenfeld's hyperbolic projective plane $(\mathbb{C} \otimes \mathbb{C}a)P^2_{hyp}$
E IV	E_6^{-26}	F_4	26	2	F_4	No	Set of the $\mathbb{C}aP^2_{hyp}$'s in $(\mathbb{C} \otimes \mathbb{C}a)P^2_{hyp}$
E V	E_7^7	$SU(8)$	70	7	$\wedge^4 SU(8)$	No	Anti-chains of $(\mathbb{H} \otimes \mathbb{C}a)P^2_{hyp}$
E VI	E_7^{-5}	$SO(12) \times SU(2)$	64	4	$\mathrm{Spin}(12) \otimes SU(2)$	No	Rosenfeld's hyperbolic projective plane $(\mathbb{H} \otimes \mathbb{C}a)P^2_{hyp}$
E VII	E_7^{-25}	$E_6 \times SO(2)$	54	3	$E_6 \otimes SO(2)$	Yes	Set of the $(\mathbb{C} \otimes \mathbb{C}a)P^2_{hyp}$'s in $(\mathbb{H} \otimes \mathbb{C}a)P^2_{hyp}$
E VIII	E_8^8	$SO(16)$	128	8	$\mathrm{Spin}(16)$	No	Rosenfeld's hyperbolic projective plane "$(\mathbb{C}a \otimes \mathbb{C}a)P^2_{hyp}$" [2])
E IX	E_8^{-24}	$E_7 \times SU(2)$	112	4	$\wedge^2 E_7 \otimes SU(2)$	No	Set of the $(\mathbb{H} \otimes \mathbb{C}a)P^2_{hyp}$'s in "$(\mathbb{C}a \otimes \mathbb{C}a)P^2_{hyp}$" [2])
F I	F_4^4	$Sp(3) \times SU(2)$	28	4	$\underline{\wedge}^3 Sp(3) \otimes SU(2)$	No	Set of the $\mathbb{H}P^2_{hyp}$'s in $\mathbb{C}aP^2_{hyp}$
F II	F_4^{-20}	$SO(9)$	16	1	$\mathrm{Spin}(9)$	No	Hyperbolic Cayley projective plane $\mathbb{C}aP^2_{hyp}$
G I	G_2^2	$SU(2) \times SU(2)$	8	2	$\underline{\otimes}^3 SU(2) \otimes SU(2)$	No	Set of the non-division quaternionic sub-algebras of the non-division Cayley algebra

[1]) here $\wedge$ (resp. $\otimes$) denotes the exterior (resp. tensor) product representation and $\underline{\wedge}$ (resp. $\underline{\otimes}$) denotes the natural irreducible representation deduced from it

[2]) up to this day an algebraic definition of this hyperbolic plane over $\mathbb{C}a \otimes \mathbb{C}a$ seems pending, see [Fre] and [Ros]

Table 4. Irreducible Symmetric Spaces of Type II and IV

Type II: $M = G$	*Type* IV: G	*dim* G	*rank* G
$SU(n+1)$	$SL(n+1, \mathbb{C})$	$n(n+2)$	n
$SO(2n+1)$	$SO(2n+1, \mathbb{C})$	$n(2n+1)$	n
$Sp(n)$	$Sp(n, \mathbb{C})$	$n(2n+1)$	n
$SO(2n)$	$SO(2n, \mathbb{C})$	$n(2n-1)$	n
E_6	$E_6^{\mathbb{C}}$	78	6
E_7	$E_7^{\mathbb{C}}$	133	7
E_8	$E_8^{\mathbb{C}}$	248	8
F_4	$F_4^{\mathbb{C}}$	52	4
G_2	$G_2^{\mathbb{C}}$	14	2

III. THE SCALAR PRODUCT

§1. Introduction

In the geometric approach, the scalar product of two vectors A and B, denoted by $A \cdot B$, is defined by

(1) $$A \cdot B = |A|\ |B| \cos \theta$$

where $|A|$ denotes the length of A, and θ is the angle between the two vectors. The assumption behind such a definition is that the notions of length and angle have already been defined in some way. This is a reasonable approach for the vector space R^3, because length and angle may be assigned their usual meanings. But in the vector space of real-valued functions on a non-empty set (Exercise I, 4.11), there are no obvious meanings to be assigned to "length of a function", and "angle between two functions".

To avoid these difficulties and to achieve greater generality, we shall use the axiomatic method to define what shall be called a scalar product.

1.1. Definition. A scalar product in a vector space V is a function which assigns to each pair of vectors A, B in V a real number, denoted by $A \cdot B$, having the following properties:

Axiom S1. For all A and B in V, $A \cdot B = B \cdot A$.

Axiom S2. For all A and B in V, and $x \in R$,

(i) $(xA) \cdot B = x(A \cdot B)$,

(ii) $A \cdot (xB) = x(A \cdot B)$.

Axiom S3. For all A, B, C in V,

(i) $(A + B) \cdot C = A \cdot C + B \cdot C$,

(ii) $A \cdot (B + C) = A \cdot B + A \cdot C$.

Axiom S4. For all A in V, $A \cdot A \geq 0$.

Axiom S5. $A \cdot A = 0$ if and only if $A = \tilde{0}$.

If we were to start with the usual length and angle in R^3, and define $A \cdot B$ by (1) above, then the properties S1 to S5 are easily derived. But we shall reverse the procedure and show how a scalar product, in the sense of the definition, may be used to define length and angle, and in such a way that formula (1) holds. To justify this procedure, we must show that many vector spaces have scalar products, and that the derived lengths and and angles possess the expected properties.

Remarks. The scalar product is so called because it assigns to each pair of vectors in V a real number, that is, a scalar. The scalar product does not define a multiplication in V in the sense of §9 of Chapter II.

The five properties of $A \cdot B$ can be paraphrased as follows. Axiom S1 asserts that $A \cdot B$ is symmetric. Axioms S2 and S3 assert that $A \cdot B$ is a linear function of each variable (e.g. for each $A \in V$ the function $V \longrightarrow R$ sending each $X \in V$ into $A \cdot X$ is a linear transformation). Axioms S4 and S5 assert that $A \cdot A$ is positive definite.

§2. Existence of scalar products

2.1. Theorem. Let V be a finite dimensional vector space, and let $A_1, \ldots, A_n$ be a basis for V (the existence of which is shown in Theorem I, 10.2). For each A, B in V, define $A \cdot B$ by

(1) $$A \cdot B = \Sigma_{i=1}^{n} a_i b_i$$

where $A = \Sigma_{i=1}^{n} a_i A_i$ and $B = \Sigma_{i=1}^{n} b_i A_i$ are the unique expressions of A, B in terms of the basis (see Theorem I, 10.3). Then the resulting function of A and B is a scalar product in V.

Proof. Axiom S1 follows from $a_i b_i = b_i a_i$. Axiom S2 follows from $xA = \Sigma\, xa_i A_i$ and $\Sigma\, xa_i b_i = x\, \Sigma a_i b_i$. Axiom S3 follows from $(a_i + b_i)c_i = a_i c_i + b_i c_i$. Since $A \cdot A = \Sigma a_i^2$ and $a_i^2 \geq 0$, Axiom S4 holds. Because $\Sigma a_i^2 = 0$ implies each $a_i = 0$, Axiom S5 holds.

2.2. Corollary. Each finite dimensional vector space has at least one scalar product.

Since a basis can be chosen in many ways it is reasonable to expect that a finite dimensional vector space will have many scalar products. This is the case, even though different bases can lead to the same scalar product.

2.3. Definition. Henceforth the symbol R^n will denote the vector space of n-tuples of real numbers (Definition I, 3.1) together with the standard scalar product

$$(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = \Sigma_{i=1}^{n} a_i b_i .$$

This is the scalar product associated with the standard basis $A_1, \ldots, A_n$ where A_i has all components zero except the ith component, which is 1. The vector space R^n is called the n-dimensional euclidean space.

§3. Length and angle

Throughout this section, V is a vector space and $A \cdot B$ is a scalar product in V. It is not assumed that $\dim V$ is finite.

3.1. Definition. If $A \in V$, Axiom S4 asserts that $A \cdot A$ has a unique square root which is ≥ 0; this root is denoted by $|A|$ and is called the length of A. Thus

$$|A| = \sqrt{A \cdot A} \quad .$$

The distance between two vectors A, B in V is defined to be $|A - B|$.

The length is sometimes called the norm, or the absolute value; it has nearly all the properties enjoyed by the absolute value of a real or complex number.

3.2. Theorem. The length function has the following properties:

(i) For each $A \in V$, $|A| \geq 0$.

(ii) If $A \in V$, then $|A| = 0$ if and only if $A = \tilde{0}$.

(iii) For each $A \in V$ and each $x \in R$, $|xA| = |x||A|$.

(iv) For each pair A, B in V, $|A \cdot B| \leq |A||B|$ (Schwarz inequality).

(v) For each pair A, B in V, $|A + B| \leq |A| + |B|$.

Proof. Parts (i), (ii) and (iii) follow quickly from the axioms of the scalar product, and are left as exercises.

To prove (iv), suppose first that $A = \tilde{0}$. Axiom S2(i) gives $\tilde{0} \cdot B = (0\tilde{0}) \cdot B = 0(\tilde{0} \cdot B) = 0$. Then both sides of (iv) are zero. The same is true if $B = \tilde{0}$. Suppose then that neither A nor B is zero. Axiom S4 applied to the vector $(B \cdot B)A - (A \cdot B)B$ gives

$$((B \cdot B)A - (A \cdot B)B) \cdot ((B \cdot B)A - (A \cdot B)B) \geq 0 \quad . \tag{1}$$

Using bilinearity, we can expand the left side and simplify to

(2) $$\{|A|^2|B|^2 - (A\cdot B)^2\}\ (B\cdot B) \geq 0 .$$

By Axiom S5, $B\cdot B > 0$, and therefore

$$|A|^2|B|^2 \geq (A\cdot B)^2 .$$

Taking positive square roots gives (iv).

To prove (v), note first that

$$|A + B|^2 = (A + B)\cdot(A + B) = A\cdot A + B\cdot B + 2A\cdot B = |A|^2 + |B|^2 + 2A\cdot B .$$

By (iv), the size of the last expression is not diminished if we replace $A\cdot B$ by $|A||B|$. Hence

$$|A + B|^2 \leq |A|^2 + |B|^2 + 2|A||B| = (|A| + |B|)^2 .$$

Taking the non-negative square roots leaves the order of the inequality unchanged. Hence (v) is proved.

3.3. <u>Corollary</u>. The distance function (or metric) has the properties:

(i) For each pair A, B in V, $|A - B| = |B - A|$.

(ii) For each pair A, B in V, $|A - B| \geq 0$, with $|A - B| = 0$ if and only if $A = B$.

(iii) For each triple A, B, C in V,

$$|A - C| \leq |A - B| + |B - C|$$

(triangle inequality).

<u>Remark</u>. There can exist distance functions, having the above properties, which are not derived from a scalar product.

3.4. <u>Definition</u>. If A and B are non-zero vectors in V, the Schwarz inequality gives

$$-1 \leq \frac{A\cdot B}{|A||B|} \leq 1 .$$

We may therefore define the <u>angle</u> θ <u>between</u> A <u>and</u> B by

$$\cos\theta = \frac{A\cdot B}{|A||B|}$$

where $0 \leq \theta \leq \pi$.

This definition gives the formula

$$A\cdot B = |A||B|\cos\theta$$

of §1. In the triangle whose sides are the vectors A, B and $A - B$, we have

$$\begin{aligned}|A - B|^2 &= (A - B)\cdot(A - B) = A\cdot A + B\cdot B - 2A\cdot B \\ &= |A|^2 + |B|^2 - 2|A||B|\cos\theta\ .\end{aligned}$$

This is the standard "cosine law", and thereby justifies the definition of θ.

3.5. <u>Lemma</u>. Two vectors A, B are dependent if and only if $A = \tilde{0}$, or $B = \tilde{0}$, or $\theta = 0$ or π.

<u>Proof</u>. Suppose A and B are dependent. If either is zero, there is nothing to prove. If neither is zero, then dependence implies $A = aB$ for some $a \neq 0$. This gives $A\cdot B = a|B|^2$ and $|A| = |a||B|$. Therefore $\cos\theta = a/|a| = \pm 1$, so $\theta = 0$ or π.

Conversely, if A or B is zero, then A and B are dependent (Definition I, 9.1). If neither A nor B is zero but $\theta = 0$ or π, then $\cos^2\theta = 1$, and Definition 3.4 gives $(A\cdot B)^2 = |A|^2|B|^2$. Therefore formula (2), in 3.2 above, is zero, which implies that (1) is zero. Hence Axiom S5 asserts that

$$(B\cdot B)A - (A\cdot B)B = \tilde{0}\ .$$

Since $B \neq \tilde{0}$, we have $B\cdot B \neq 0$, and it follows that A and B are dependent.

§4. Exercises

1. Show that Axiom S2 (ii) of scalar product (see Definition 1.1) is a consequence of Axioms S1 and S2 (i).

2. Show that Axiom S3 (ii) follows from Axioms S1 and S3 (i).

3. Show that $\tilde{0}\cdot A = 0$ for all A follows from Axiom S2 (i), and also from Axiom S3 (i).

4. Show that Axioms S1, S2, S3 and S5 of Definition 1.1 imply that either Axiom S4 holds or else that $A\cdot A \leq 0$ for every $A \in V$. (Hint: Assume the contrary that, for some vector A and some vector B, $A\cdot A > 0$ and $B\cdot B < 0$. Then show that $L(A, B)$ contains a vector $C \neq \tilde{0}$ such that $C\cdot C = 0$, thus contradicting Axiom S5.)

5. Let $A_1, \ldots, A_n$ be a basis for V. Let $A\cdot B$ and $A:B$ denote two scalar products in V. Show that $A\cdot B = A:B$ for all A and B if and only if $A_i\cdot A_j = A_i:A_j$ for all i and j in the range $1, \ldots, n$.

6. Describe the totality of distinct scalar products that a 1-dimensional vector space can have.

7. Prove parts (i), (ii) and (iii) of Theorem 3.2.

8. Show that $|A\cdot B| = |A||B|$ if and only if A and B are dependent.

9. Under what conditions is it true that $|A + B| = |A| + |B|$?

10. Prove Corollary 3.3.

11. Let V be the vector space of continuous real-valued functions defined over the interval $[a, b]$. For f and g in

V, define $f \cdot g \in R$ by

$$f \cdot g = \int_a^b f(x)g(x)dx \ .$$

Show that $f \cdot g$ is a scalar product in V. Interpreting integrals as areas, what is length $|f|$? What is the interpretation of perpendicularity $f \cdot g = 0$?

§5. Orthonormal bases

Throughout this section, V will be a vector space with a fixed scalar product.

5.1. Definition. Two vectors A and B in V are orthogonal if $A \cdot B = 0$. A basis in V is orthogonal if each two distinct basis vectors are orthogonal. A vector A in V is normal if its length is 1 or, equivalently, if $A \cdot A = 1$. A basis in V is orthonormal if it is orthogonal and each basis vector is normal.

Remarks. Since $\tilde{0} \cdot A = 0$, it follows that $\tilde{0}$ is orthogonal to every vector of V including itself. If A and B are orthogonal, and neither is $\tilde{0}$, then $\cos\theta = A \cdot B/|A||B| = 0$. Therefore the angle θ between A and B is $\pi/2$. Thus, for non-zero vectors, orthogonal is the same as perpendicular.

The standard basis $A_1, \ldots, A_n$ (Definition 2.3) is an orthonormal basis in R^n because

$$(1) \qquad A_i \cdot A_j = \begin{cases} 0 & \text{if } i \neq j \ , \\ 1 & \text{if } i = j \ . \end{cases}$$

The main objective of this section is to show that V has an orthonormal basis if its dimension is finite.

5.2. Proposition. If $A_1, \ldots, A_n$ is an orthogonal basis in V and $X \in V$, then

$$X = \Sigma_{i=1}^{n} \frac{X \cdot A_i}{A_i \cdot A_i} A_i \quad . \tag{2}$$

If the basis is also orthonormal, then

$$X = \Sigma_{i=1}^{n} (X \cdot A_i) A_i \quad . \tag{3}$$

Proof. By definition of a basis, there are numbers $a_1, \ldots, a_n$ (depending on X) such that $X = \Sigma_{i=1}^{n} a_i A_i$. Let j be an integer in the range $1, \ldots, n$. By the linearity of scalar product, we have

$$X \cdot A_j = (\Sigma\, a_i A_i) \cdot A_j = \Sigma_{i=1}^{n} a_i (A_i \cdot A_j) = a_j (A_j \cdot A_j) \ ,$$

since the orthogonality of the basis gives $A_i \cdot A_j = 0$ for each $i \neq j$. Since a basis vector is never zero, we may divide by $A_j \cdot A_j$, thereby solving for a_j. Hence

$$X = \Sigma_{j=1}^{n} a_j A_j = \Sigma_{j=1}^{n} \frac{X \cdot A_j}{A_j \cdot A_j} A_j \quad .$$

This proves (2). Since normality gives $A_i \cdot A_i = 1$, (3) follows from (2).

5.3. Proposition. Let U be a linear subspace of V. Then the set of those vectors of V which are orthogonal to every vector of U is a linear subspace of V called the orthogonal complement of U, and denoted by $U^{\perp}$.

Proof. Suppose A and B are orthogonal to every vector $X \in U$, that is, $A \cdot X = 0$ and $B \cdot X = 0$ for every $X \in U$. Then, by Axiom S3,

$$(A + B)\cdot X = A\cdot X + B\cdot X = 0 \qquad \text{for } X \in U.$$

Also, if $a \in R$, then by Axiom S2,

$$(aA)\cdot X = a(A\cdot X) = 0 \qquad \text{for } X \in U .$$

Thus $U^{\perp}$ is a linear subspace.

5.4. Proposition. Let U be a linear subspace of V such that U has an orthonormal basis $A_1, \ldots, A_k$. Then each vector X in V is uniquely representable as the sum of a vector in U and a vector in $U^{\perp}$, namely:

$$X = \{\Sigma_{i=1}^{k} (X\cdot A_i)A_i\} + \{X - \Sigma_{i=1}^{k} (X\cdot A_i)A_i\} . \tag{4}$$

Remark. Then $V = U \oplus U^{\perp}$ is a direct sum decomposition of V (Definition II, 11.7).

Proof. Let $X' = \Sigma_{i=1}^{k}(X\cdot A_i)A_i$. It is obvious that $X' \in U$, and that $X = X' + (X - X')$. We must show that $X - X' \in U^{\perp}$. For each $j = 1, \ldots, k$, we have

$$(X - X')\cdot A_j = X\cdot A_j - \Sigma_{i=1}^{k}(X\cdot A_i)(A_i\cdot A_j) = X\cdot A_j - X\cdot A_j = 0 .$$

If $A \in U$, then $A = \Sigma_{j=1}^{k} a_jA_j$ where $a_j = A\cdot A_j$. Therefore

$$(X - X')\cdot A = \Sigma_{j=1}^{k} a_j(X - X')\cdot A_j = 0 .$$

This proves that $X - X'$ is in $U^{\perp}$.

It remains to prove uniqueness. Suppose $X = A + C$ where $A \in U$ and $C \in U^{\perp}$. Then, for each $i = 1, \ldots, k$, $X\cdot A_i = A\cdot A_i$ since $C\cdot A_i = 0$. Therefore

$$X' = \Sigma_{i=1}^{k}(X\cdot A_i)A_i = \Sigma_{i=1}^{k} (A\cdot A_i)A_i .$$

By Proposition 5.2, for U rather than V, we get $A = \Sigma(A\cdot A_i)A_i$.

That is, $A = X'$. Hence also $C = X - A = X - X'$.

5.5. <u>Theorem</u>. If $B_1, \ldots, B_n$ are independent vectors in V, then the linear subspace $L(B_1, \ldots, B_n)$ has an orthonormal basis $A_1, \ldots, A_n$ such that

(5) $\quad L(B_1, \ldots, B_i) = L(A_1, \ldots, A_i)$ for each $i = i, \ldots, n$.

<u>Remark</u>. The construction given in the following proof is known as the Gram-Schmidt process of orthonormalization.

<u>Proof</u>. The proof proceeds by induction on n. If $n = 1$, we set $A_1 = B_1/|B_1|$. Then A_1 is a normal vector, and is an orthonormal basis for $L(B_1)$. Assume inductively that $A_1, \ldots, A_{k-1}$ have been found such that they form an orthonormal basis for $L(B_1, \ldots, B_{k-1})$, and that formula (5) holds for $i = 1, \ldots, k - 1$. Apply Proposition 5.4 with $U = L(A_1, \ldots, A_{k-1})$ and $X = B_k$. Then $B_k = B' + B''$, where $B' \in U$ and $B'' \in U^{\perp}$. Since B_k is not in U, B'' is not zero, so we can define $A_k = B''/|B''|$. Then A_k has length 1 and is orthogonal to U. Therefore $A_1, \ldots, A_{k-1}, A_k$ is an orthonormal set of vectors. Since

$$B_k = B' + |B''|A_k$$

and $B' \in U$, it follows that $B_k \in L(A_1, \ldots, A_k)$. But $A_k = (B_k - B')/|B''|$ is in $L(A_1, \ldots, A_{k-1}, B_k)$, so

$$L(B_1, \ldots, B_k) = L(A_1, \ldots, A_{k-1}, B_k) = L(A_1, \ldots, A_k) .$$

This completes the inductive step and the proof of the theorem.

5.6. <u>Corollary</u>. Each finite dimensional subspace of V has an orthonormal basis. If $\dim V$ is finite, then V has an orthonormal basis.

In fact, by Theorem I, 10.2, a finite dimensional subspace

has a basis $B_1, \ldots, B_n$ to which the theorem may be applied to give an orthonormal basis.

5.7. Corollary. If $\dim V = n$, and $A_1, \ldots, A_k$ is an orthonormal basis for a subspace U of V, then there are vectors $A_{k+1}, \ldots, A_n$ such that $A_1, \ldots, A_k, A_{k+1}, \ldots, A_n$ is an orthonormal basis for V.

Proof. The orthogonal complement $U^{\perp}$, being a linear subspace of V, has, by Corollary 5.6, an orthonormal basis, say $C_1, \ldots, C_h$. Then $A_1, \ldots, A_k, C_1, \ldots, C_h$ is an orthonormal set of vectors. By Proposition 5.4, it is a basis for V. By Theorem I, 10.2, any basis for V has n vectors. Therefore $h + k = n$.

5.8. Theorem. Let V be finite dimensional, and let $T : V \longrightarrow R$ be linear. Then there exists a unique vector $A \in V$ (depending on T) such that

$$T(X) = A \cdot X \qquad \text{for each } X \in V .$$

Proof. The only linear subspaces of R are 0 and R. If $\operatorname{im} T = 0$, set $A = \tilde{0}$. Otherwise $\operatorname{im} T = R$. Let $U = \ker T$. If $\dim V = n$, then $\dim U = n - 1$ by Theorem II, 2.5. Let $A_1, \ldots, A_{n-1}$ be an orthonormal basis in U. By Corollary 5.7, there is a vector A_n such that $A_1, \ldots, A_n$ is an orthonormal basis for V. For $X \in V$, Proposition 5.2 gives

$$X = \Sigma_{i=1}^{n} (X \cdot A_i) A_i ,$$

and therefore

$$T(X) = \Sigma_{i=1}^{n} (X \cdot A_i) T(A_i) .$$

But for $i < n$, $A_i \in U$ and so $T(A_i) = 0$. Hence

$$T(X) = (X\cdot A_n)T(A_n) = (T(A_n)A_n)\cdot X ,$$

by Axiom S2. Thus $A = T(A_n)A_n$ is the required vector.

To prove uniqueness, suppose A and B are such that $A\cdot X = B\cdot X$ for all X. Then $(A - B)\cdot X = 0$. Take X to be $A - B$; then Axiom S5 gives $A - B = \tilde{0}$.

5.9. Theorem. If $A_1, \ldots, A_n$ is an orthonormal basis in V, and if

$$A = \Sigma_{i=1}^n a_i A_i, \quad B = \Sigma_{i=1}^n b_i A_i ,$$

then $A\cdot B = \Sigma_{i=1}^n a_i b_i$.

Proof. Using the linearity of the scalar product,

$$(\Sigma_{i=1}^n a_i A_i)\cdot(\Sigma_{j=1}^n b_j A_j) = \Sigma_{i=1}^n \Sigma_{j=1}^n a_i b_j (A_i\cdot A_j)$$
$$= \Sigma_{i=1}^n a_i b_i .$$

Remark. Theorem 2.1 associates with each basis $A_1, \ldots, A_n$ a scalar product defined to be $\Sigma\, a_i b_i$. Thus Theorem 5.9 asserts that each scalar product can be obtained as one associated with a basis, namely, any orthonormal basis. It appears from this result that all scalar products are pretty much alike. A precise formulation of this fact requires the notion of isometry.

§6. Isometries

6.1. Definition. Let V and W denote vector spaces with scalar products. A linear transformation $T : V \longrightarrow W$ is called an isometry if $\operatorname{im} T = W$ and

$$T(A)\cdot T(B) = A\cdot B \qquad \text{for all } A \text{ and } B \text{ in } V .$$

If such a T exists, V and W are said to be isometric.

The justification for the use of "iso" in the above

definition lies in the following

6.2. Proposition. If $T : V \longrightarrow W$ is an isometry, then T is an isomorphism, and T^{-1} is an isometry.

Proof. If $A \in \ker T$, then $A \cdot A = T(A) \cdot T(A) = 0$, which implies $A = \tilde{0}_V$; that is, $\ker T$ is zero. Since $\operatorname{im} T = W$, it follows that T is an isomorphism. If A' and B' are in W, let $A = T^{-1}(A')$ and $B = T^{-1}(B')$. Then

$$T^{-1}(A') \cdot T^{-1}(B') = A \cdot B = T(A) \cdot T(B) = A' \cdot B' \quad ,$$

that is, T^{-1} is an isometry.

Stated briefly, an isometry is an isomorphism preserving the scalar product. It will therefore preserve anything derived from the scalar product, e.g. length, angle, and distance. An isometry is a rigid transformation: a configuration of points (= vectors), lines, and planes in V is carried by an isometry into a congruent configuration in W. Thus, from the point of view of euclidean geometry, isometric spaces are fully equivalent.

6.3. Theorem. If V and W are n-dimensional vector spaces with scalar products, then there exists an isometry $T : V \longrightarrow W$.

Proof. By Corollary 5.6, there are orthonormal bases $A_1, \ldots, A_n$ in V, and $B_1, \ldots, B_n$ in W. By Theorem II, 4.1, a unique linear transformation $T : V \longrightarrow W$ is determined by setting $T(A_i) = B_i$ for $i = 1, \ldots, n$. If $B = \Sigma_{i=1}^n b_i B_i$ is a vector of W, then $T(\Sigma b_i A_i) = B$; hence $\operatorname{im} T = W$. Let X, Y be vectors in V. For suitable x's and y's, we have

$$X = \Sigma_{i=1}^n x_i A_i, \quad Y = \Sigma_{i=1}^n y_i A_i \ .$$

Then

$$T(X) = \Sigma_{i=1}^{n} x_i B_i, \quad T(Y) = \Sigma_{i=1}^{n} y_i B_i \ .$$

Because the A's are orthonormal, Theorem 5.9 gives

$$X \cdot Y = \Sigma_{i=1}^{n} x_i y_i \ .$$

Because the B's are also orthonormal,

$$T(X) \cdot T(Y) = \Sigma_{i=1}^{n} x_i y_i = X \cdot Y \ .$$

Therefore T is an isometry.

6.4. Corollary. Any n-dimensional vector space with a scalar product is isometric to R^n.

This corollary clinches the matter: an n-dimensional vector space with a scalar product is nothing more than another copy of euclidean n-dimensional space.

6.5. Definition. If V has a scalar product, an orthogonal transformation of V is an endomorphism $V \longrightarrow W$ which is an isometry.

By Proposition 6.2, an orthogonal transformation must be an automorphism. Hence the set of orthogonal transformations is a subset of the general linear group $A(V)$. It is in fact a subgroup (cf. Theorem II, 9.2). It is easily shown that if S and T are orthogonal so also is their composition ST. It is obvious that I_V is orthogonal. Finally, the orthogonality of T implies that of T^{-1} by Proposition 6.2. Thus we may speak of the orthogonal group of V.

6.6. Theorem. Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be two orthonormal bases in V. Then there is one and only one orthogonal transformation T of V such that $T(A_i) = B_i$ for $i = 1, \ldots, n$.

The proof is similar to that of Theorem 6.3.

6.7. Corollary. If $A_1, \ldots, A_n$ is an orthonormal basis in V, a one-to-one correspondence between the group of orthogonal transformations and the set of all orthonormal bases in V is defined by assigning to each orthogonal T the basis $T(A_1), \ldots, T(A_n)$.

6.8. Orthogonal matrices. Recall (II, §5) that a basis $A_1, \ldots, A_n$ and an endomorphism $T : V \longrightarrow V$ determine a matrix (α_{ji}) by

$$T(A_j) = \Sigma_{i=1}^{n} \alpha_{ji} A_i \ . \tag{1}$$

If V has a scalar product, and the A's form an orthonormal basis, then Proposition 5.2 gives

$$\alpha_{ji} = T(A_j) \cdot A_i \ . \tag{2}$$

The components of $T(A_j)$ with respect to the A's are the numbers α_{ji} for $i = 1, \ldots, n$.

By Theorem 6.6, T is orthogonal if and only if $T(A_1), \ldots, T(A_n)$ is an orthonormal set of vectors. Thus T is orthogonal if and only if each two rows of its matrix are orthogonal and each row is a unit vector:

$$\Sigma_{i=1}^{n} \alpha_{ji}\alpha_{ki} = \begin{cases} 0 & \text{if } j \neq k \ , \\ 1 & \text{if } j = k \ . \end{cases} \tag{3}$$

This formula provides an easy test for recognizing that a transformation given by a matrix is orthogonal.

§7. Exercises

1. If $\dim U$ is finite, show that $(U^{\perp})^{\perp} = U$.

2. Let U and V be as in Proposition 5.4. Let $X \in V$, and let X' be the component of X in U. Show that

$$|X - X'| \leq |X - A| \qquad \text{for each } A \in U ,$$

thus proving that X' is the vector of U which is <u>nearest</u> to X.

3. Show that the basis $A_1, \ldots, A_n$ in Theorem 5.5 is uniquely determined, up to $\pm$ signs, by formula (5) of §5.

4. In R^3, let $B_1 = (1, 1, 0)$, $B_2 = (-1, 2, 3)$ and $B_3 = (0, 1, 2)$. Apply the Gram-Schmidt orthonormalization process of Theorem 5.5 to obtain the orthonormal basis

$$A_1 = \frac{1}{\sqrt{2}}(1, 1, 0), \quad A_2 = \frac{1}{\sqrt{6}}(-1, 1, 2), \quad A_3 = \frac{1}{\sqrt{3}}(1, -1, 1) .$$

5. Let V be the vector space of continuous, real-valued functions defined on the interval $[0, 1]$, with the scalar product $f \cdot g = \int_0^1 f(x)g(x)dx$ (see Exercise 4.11). The functions $1, x, x^2, x^3$ are vectors of V. Apply the Gram-Schmidt orthonormalization, and find the resulting orthonormal polynomials.

6. Let V be the vector space of continuous functions on the interval $[-\pi, \pi]$ with the scalar product described in Exercise 4.11. Show that the functions

$$1, \sin x, \cos x, \sin 2x, \cos 2x, \ldots, \sin nx, \cos nx, \ldots$$

form an orthogonal set of vectors in V. What are their lengths? Let U_n be the subspace spanned by the first $2n + 1$ of these vectors. Let $f \in V$, and let f_n be the component of f in U_n. Then f_n is a linear combination

$$f_n(x) = a_0 + \Sigma_{k=1}^{n} (a_k \cos kx + b_k \sin kx) .$$

Using Proposition 5.2, find formulas for the coefficients a_k, b_k as suitable integrals. These coefficients are called the <u>Fourier coefficients</u> of f.

7. Find the Fourier coefficients (see Exercise 6) of each of the following functions: x, x^2, $\cos^2 x$.

IV. VECTOR PRODUCTS IN R^3

§1. Introduction

Throughout this chapter V will denote a 3-dimensional vector space with a scalar product. The reason for the restriction is that the precise notion of vector product does not generalize to vector spaces of dimension different from three, although partially analogous products can be considered in such vector spaces (see Chapter IX).

In the geometric approach, the product $A \times B$ of two vectors is defined to be a vector which is perpendicular to both A and B, and whose length is $|A||B| \sin \theta$, where θ is the angle between A and B. An immediate difficulty arises: there are two such vectors (differing only in sign). A uniform choice of $A \times B$ must be made for all pairs A, B. This is done by imposing the "right (or left) hand rule" for selecting the direction of $A \times B$. The difference between left and right handed configurations is certainly clear to the intuition, but its mathematical basis is not so clear.

The approach we shall take is the axiomatic one. We shall describe what will be called a vector product if there is one; we shall show that at least one such product exists; and we shall conclude by showing that there are precisely two vector products. One of these can be called the right-handed product, and the other, the left.

1.1. Definition. A vector product in V is a function

which assigns to each ordered pair of vectors A, B in V a vector $A \times B$, and which has the following five properties:

Axiom V1. For all A, B in V, $A \times B = - B \times A$.

Axiom V2. For all A, B, C in V,

$$A \times (B + C) = A \times B + A \times C .$$

Axiom V3. For all A, B in V and $x \in R$,

$$x(A \times B) = (xA) \times B .$$

Axiom V4. For all A, B in V, $A \cdot (A \times B) = 0$.

Axiom V5. For all non-zero A, B, in V, $|A \times B| = |A||B| \sin \theta$, where (Definition III, 3.4) $\cos \theta = A \cdot B/|A||B|$.

Axiom V1 asserts that the function of two variables is <u>skew-symmetric</u>, and replaces the usual commutative law for multiplication (which does not hold). Skew-symmetry implies that $A \times A = \tilde{0}$ for all A in V.

Using V1 and V2, we can prove

V2'. For all A, B, C in V,

$$(A + B) \times C = A \times C + B \times C .$$

For,

$$(A + B) \times C = - C \times (A + B) = - (C \times A + C \times B)$$
$$= - C \times A - C \times B = A \times C + B \times C .$$

Similarly, V1 and V3 imply

V3'. For all A, B in V, and $x \in R$,

$$x(A \times B) = A \times (xB) .$$

<u>Remark</u>. The properties V2, V3, V2', V3' are paraphrased by saying that $A \times B$ is a <u>bilinear</u> function of A and B. Precisely, if A is fixed in V, the function $T: V \longrightarrow V$,

defined by $T(X) = A \times X$ for each X, is a linear transformation. Similarly, the function $S: V \longrightarrow V$, defined by $S(X) = X \times A$ for each X, is linear.

Again, using V1 and V4, one can prove

V4'. For all A, B in V, $B \cdot (A \times B) = 0$.

Thus $A \times B$ is orthogonal to both A and B.

Axiom V5 involves the angle θ. Recall (III, 3.4) that θ is the principal value so $0 \leqq \theta \leqq \pi$, and therefore $\sin \theta \geqq 0$. Using the identity $\cos^2\theta + \sin^2\theta = 1$, Axiom V5 can be shown to be completely equivalent to

V5'. For all A, B in V,

$$|A \times B|^2 + (A \cdot B)^2 = |A|^2 |B|^2 \quad .$$

The advantage of V5' is that θ has disappeared, and the relation remains true when A or B is $\tilde{\tilde{0}}$.

Axiom V5 gives a geometric interpretation of $|A \times B|$. Consider the parallelogram whose sides are A and B. If A is called the base, then the altitude of the parallelogram is $|B| \sin \theta$. Therefore $|A||B| \sin \theta = |A \times B|$ is the area of the parallelogram.

At the present stage, we do not know that a vector product exists. However, if one does exist and it is denoted by $A \times B$, then a second one $\tilde{\times}$ also exists. In fact, if we define $A \tilde{\times} B$ by $A \tilde{\times} B = - A \times B$ for all A, B, it is easily verified that $\tilde{\times}$ satisfies Axioms V1 through V5 (restated with $\tilde{\times}$ in place of $\times$).

§2. The triple product

It is assumed in this section that V has a vector product, and that $A \times B$ is one.

2.1. Proposition. $A \times B = \tilde{\tilde{0}}$ if and only if A and B are dependent.

Proof. By Axiom V5, $|A \times B| = 0$ if and only if

$$|A| = 0, \quad \text{or} \quad |B| = 0, \quad \text{or} \quad \sin\theta = 0 ;$$

this is true if and only if

$$A = \tilde{\tilde{0}}, \quad \text{or} \quad B = \tilde{\tilde{0}}, \quad \text{or} \quad \theta = 0 \quad \text{or} \quad \pi ;$$

that is, by III, 3.5, if and only if A and B are dependent.

2.2. Definition. If A, B, C are vectors in V, their triple product, denoted by $[A, B, C]$, is defined by

$$[A, B, C] = A\cdot(B \times C) . \tag{1}$$

In the expression $A\cdot(B \times C)$ the parentheses can, and will, be omitted because $(A\cdot B) \times C$ has no meaning.

The number $|A\cdot B \times C|$ has a geometric interpretation: it is the volume of the parallelepiped whose edges are A, B, and C. To see this, let θ be the angle between B and C, and φ the angle between A and $B \times C$. Then

$$|A\cdot B \times C| = (|B||C| \sin\theta)|A||\cos\varphi| .$$

We have already seen that $|B||C| \sin\theta$ is the area of the parallelogram spanned by B and C, which may be taken as the base of the parallelepiped. The altitude with respect to this base is $|A||\cos\varphi|$, because φ is the angle between the third side A and a perpendicular to the base. Since the volume is the product of the area of the base and the altitude, the assertion follows.

2.3. Theorem. The triple product has the following properties:

(i) It is skew-symmetric in all three variables, i.e. an interchange of any two variables reverses the sign.

(ii) It is trilinear, i.e. it is linear in each variable.

(iii) It is zero if and only if A, B, and C are dependent.

<u>Proof</u>. To prove (i), we must prove, for all A, B, C in V, that

(2) $[A,B,C] = -[B,A,C] = [B,C,A] = -[C,B,A] = [C,A,B] = -[A,C,B]$.

By Axiom V4,

$$(A + B)\cdot(A + B) \times C = 0 .$$

Expand the left side using the linearity of the vector product and then of the scalar product. This gives

$$A\cdot A \times C + B\cdot A \times C + A\cdot B \times C + B\cdot B \times C = 0 .$$

By Axiom V4, the first and fourth terms are zero. Therefore

(3) $A\cdot B \times C = -B\cdot A \times C$ for all A, B, C in V.

This gives the first equality in (2). Using Axiom V1,

(4) $$A\cdot B \times C = A\cdot(-C \times B) = -A\cdot C \times B .$$

Thus the first term of (2) equals the sixth. So far we have shown that switching the first two or the second two variables reverses the sign. This is enough, since each arrangement in (2) is obtained from the preceding by such an interchange. For example the last equality $[C, A, B] = -[A, C, B]$ is true since it results from switching the first two variables.

Property (ii), linearity in each variable, follows from the fact that both the scalar product and the vector product are bilinear.

Suppose A, B, and C are dependent. Then one depends on the other two, and by skew-symmetry, it suffices to consider the case that A depends on B and C, i.e. $A = bB + cC$. Using Axion V4,

$$A \cdot B \times C = (bB + cC) \cdot B \times C = bB \cdot B \times C + cC \cdot B \times C = 0 \quad .$$

Conversely, suppose A, B, C are such that $A \cdot B \times C = 0$. If $B \times C = \tilde{0}$, Proposition 2.1 asserts that B and C are dependent; then A, B and C are dependent. If $B \times C \neq \tilde{0}$, then Proposition 2.1 asserts that B and C are independent, so $L(B, C)$ has dimension 2. Let U be its orthogonal complement (see Proposition III, 5.3). Since V has dimension 3, the dimension of U is 1. But $B \times C$ is in U and is not $\tilde{0}$, so $U = L(B \times C)$. The assumption $A \cdot B \times C = 0$ implies that A is orthogonal to $B \times C$, and hence to U. Thus A is in the orthogonal complement of U, which is $L(B, C)$ by Exercise III, 7.1. This proves that A depends on B and C, and completes the proof of the theorem.

2.4. <u>Corollary</u>. $A \cdot B \times C = A \times B \cdot C$ for A, B, C in V.

<u>Proof</u>.

$$A \cdot B \times C = [A, B, C] = [C, A, B] = C \cdot A \times B = A \times B \cdot C \quad .$$

2.5. <u>Theorem</u>. Let i, j, k be an orthonormal basis in V. Then $i \times j$ is either k or $-k$. In case $i \times j = k$, then $j \times k = i$ and $k \times i = j$ and, if

$$A = a_1 i + a_2 j + a_3 k, \qquad B = b_1 i + b_2 j + b_3 k \quad ,$$

then

(5) $A \times B = (a_2b_3 - a_3b_2)i + (a_3b_1 - a_1b_3)j + (a_1b_2 - a_2b_1)k$.

Proof. Since i and j are orthogonal, their angle is $\pi/2$, and its sine is 1. Since i and j have length 1, it follows from Axiom V5 that $i \times j$ has length 1. Thus $i \times j$ is a unit vector in the orthogonal complement of $L(i, j)$, that is, in $L(k)$. Since k has length 1, we conclude that $i \times j$ is k or $-k$. By symmetry, $j \times k$ is i or $-i$, and $k \times i$ is j or $-j$.

Suppose $i \times j = k$. Then

$$[k, i, j] = k \cdot i \times j = k \cdot k = 1 \quad .$$

By skew-symmetry of the triple product,

$$j \cdot k \times i = [j, k, i] = [k, i, j] = 1 \quad .$$

Then $k \times i$ cannot be $-j$, since $j \cdot (-j) = -1$, so $k \times i = j$. Similarly $j \times k$ must be i and not $-i$.

The formula (5) for $A \times B$ is obtained by expanding

$$(a_1i + a_2j + a_3k) \times (b_1i + b_2j + b_3k) \quad ,$$

using bilinearity. There are nine terms in all. Three are zero since $i \times i$, $j \times j$, and $k \times k$ are zero (Axiom V1) and the remaining six terms can be computed from $i \times j = k$, $j \times i = -k$, etc.

2.6. Corollary. A 3-dimensional vector space with a scalar product has at most two vector products, and these can differ only in sign.

Observe that the formula for $A \times B$ was determined by knowing that $i \times j = k$. In the other case: $i \times j = -k$, we

would obtain the same formula with sign reversed.

§3. Existence of a vector product

3.1. Theorem. Let V be a 3-dimensional vector space with a scalar product, and let i, j, k be an orthonormal basis in V. Let $A \times B$ be defined by formula (5) of Theorem 2.5. Then $A \times B$ is a vector product in V.

Proof. Since $a_2b_3 - a_3b_2 = -(b_2a_3 - b_3a_2)$, etc., it follows that $A \times B = -B \times A$, i.e. Axiom V1 is satisfied.

Axiom V2 follows from

$$a_2(b_3 + c_3) - a_3(b_2 + c_2) = (a_2b_3 - a_3b_2) + (a_2c_3 - a_3c_2)$$

and similar relations for the other two components.

Axiom V3 follows from

$$(xa_2)b_3 - (xa_3)b_2 = x(a_2b_3 - a_3b_2)$$

and similar relations on the remaining components.

By Theorem III, 5.9, $A \cdot A \times B$ is given by

$$a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1) = 0 \quad ,$$

so Axiom V4 holds.

Property V5' (equivalent to Axiom V5) is proved by verifying the identity

$$\begin{aligned}(a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\ + (a_1b_1 + a_2b_2 + a_3b_3)^2 \\ = (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) \quad .\end{aligned}$$

This is done by brute force, and is left to the student.

3.2. Corollary. A 3-dimensional vector space with a

scalar product has precisely two vector products, and these differ only in sign.

Remarks. The two vector products are equivalent in the following sense. Let $T: V \longrightarrow V$ be the linear transformation defined by $T(i) = i$, $T(j) = j$ and $T(k) = -k$, where i, j, k is an orthonormal basis in V. Because $i, j, -k$ is also orthonormal, T is an isometry. If $\times$ is defined by $i \times j = k$, and $\tilde{\times}$ by $i \tilde{\times} j = -k$, it is easily seen that

$$T(A) \tilde{\times} T(B) = T(A \times B) ;$$

that is, T transforms $\times$ into $\tilde{\times}$. If a given choice of $\times$ is called "right-handed", then an orthonormal basis i, j, k is called "right-handed" if and only if $i \times j = k$. If $i \times j = -k$, the orthonormal basis i, j, k would then be called "left-handed".

§4. Properties of the vector product

4.1. Proposition. If A and B are vectors such that $A \cdot B = 0$ and $A \times B = \tilde{0}$, then either $A = \tilde{0}$ or $B = \tilde{0}$.

Proof. In fact, $A \cdot B = 0$ implies $A = \tilde{0}$ or $B = \tilde{0}$ or $\cos \theta = 0$. Only the last case need be considered. Then $\sin \theta = 1$, so $A \times B = \tilde{0}$ gives $|A \times B| = |A||B| = 0$. Thus, either $|A| = 0$ or $|B| = 0$, and therefore $A = \tilde{0}$ or $B = \tilde{0}$.

Remark. If a product of real numbers $ab = 0$, and a is not zero, we may divide by a and conclude that $b = 0$. Now the ability to divide out a non-zero factor fails for each of the products $A \cdot B$ and $A \times B$. If $A \neq \tilde{0}$ and $A \cdot B = 0$, we know only that B is orthogonal to A. If $A \neq \tilde{0}$ and $A \times B = \tilde{0}$, we know only that B depends on A. Thus there is no reasonable notion

of <u>division</u> to go with either the scalar or the vector product.

4.2. <u>Proposition</u>. For any vectors A, B, C, we have

$$A \times (B \times C) = (A \cdot C)B - (A \cdot B)C \quad . \tag{1}$$

<u>Remark</u>. It should be emphasized that the vector product is not associative. For example

$$i \times (i \times j) = i \times k = -k \times i = -j \quad ,$$
$$(i \times i) \times j = \tilde{0} \times j = \tilde{0} \quad .$$

<u>Proof</u>. Let $A = a_1 i + a_2 j + a_3 k$, etc. Using formula (5) of Theorem 2.5,

$$\begin{aligned} A \times (B \times C) = {} & \{a_2(b_1c_2 - b_2c_1) - a_3(b_3c_1 - b_1c_3)\}i \\ & + \{a_3(b_2c_3 - b_3c_2) - a_1(b_1c_2 - b_2c_1)\}j \\ & + \{a_1(b_3c_1 - b_1c_3) - a_2(b_2c_3 - b_3c_2)\}k \quad . \end{aligned}$$

On the other hand,

$$(A \cdot C)B = (a_1c_1 + a_2c_2 + a_3c_3)(b_1 i + b_2 j + b_3 k) \quad ,$$
$$(A \cdot B)C = (a_1b_1 + a_2b_2 + a_3b_3)(c_1 i + c_2 j + c_3 k) \quad .$$

If we multiply out the last two lines and subtract, the resulting coefficients of i, j, k coincide with those in $A \times (B \times C)$.

§5. <u>Analytic geometry</u>

Much of the ordinary solid analytic geometry can be expressed in abbreviated style using vector algebra. The following paragraphs are samples.

The distance between two points represented by vectors A and B is $|B - A|$.

The area of a triangle whose vertices are represented by

vectors A, B, C is

$$\frac{1}{2}\,|(B - A) \times (C - A)| \quad ,$$

since the area of the triangle is half that of the parallelogram whose sides are $B - A$ and $C - A$. (see §1).

The volume of a tetrahedron whose vertices are A, B, C, D is

$$\frac{1}{6}\,|[B - A,\ C - A,\ D - A]| \quad ,$$

since the triple product gives the volume of the parallelepiped, and the latter divides into six tetrahedra of equal volumes.

The perpendicular distance from a point A to the line through B and C is just the altitude of the parallelogram on the vectors $A - B$ and $C - B$. The quotient of the area of the parallelogram by the length of the base gives the altitude:

$$|(A - B) \times (C - B)|/|C - B| \ .$$

The perpendicular distance from a point A to the plane through the points B, C, D is the altitude of a parallelepiped. Dividing the volume by the area of the base gives

$$|[A - B,\ C - B,\ D - B]|/|(C - B) \times (D - B)| \ .$$

Suppose we wish to find the shortest distance between the line through the points A, B and the line through the points C, D. The vector $(B - A) \times (C - D)$ is perpendicular to both lines. The perpendicular projection of the vector $C - A$ on the common perpendicular is the required length:

$$|(C - A)\cdot(B - A) \times (C - D)|/|(B - A) \times (C - D)| \ .$$

Let A be a non-zero vector. The set of vectors X satisfying $A \cdot X = 0$ is the linear subspace (plane through $\tilde{0}$) perpendicular to $L(A)$. If B is any vector, the set of vectors Y satisfying

$$A \cdot (Y - B) = 0$$

forms a plane through B perpendicular to $L(A)$. This is seen by setting $X = Y - B$; for as X ranges over the plane $A \cdot X = 0$, $Y = X + B$ ranges over the parallel plane through B.

Let A be a non-zero vector. The vectors X satisfying $A \times X = \tilde{0}$ are those dependent on A. They form the line $L(A)$. If B is any vector, the set of vectors Y satisfying

$$A \times (Y - B) = \tilde{0}$$

forms a line through B parallel to $L(A)$. This is seen by setting $X = Y - B$ as in the preceding paragraph.

If A is a vector and $r \in R$, it is obvious that

$$|X - A| = r$$

is the equation of a sphere with center at the point A and radius r. If A and B are vectors, the equation

$$(X - A) \cdot (X - B) = 0$$

is that of a sphere having as a diameter the segment connecting the points A and B. This seen by recalling that a triangle with vertices on a sphere, and with one side a diameter, is a right triangle.

§6. Exercises

1. For any vectors A, B, and C show that

$$A \times (B \times C) + C \times (A \times B) + B \times (C \times A) = \tilde{\tilde{0}} .$$

This is called the Jacobi identity (cf. Exercise II, 10.8).

2. Let A be a non-zero vector, let B be orthogonal to A, and let $c \in R$. Show that the pair of equations

$$A \cdot X = c, \qquad A \times X = B$$

has one and only one solution X.

3. In R^3, let $A * B$ denote a multiplication which assigns to each two vectors A, B a vector $A * B$, and which satisfies the conditions of bilinearity, namely:

$$x(A * B) = (xA) * B = A * (xB), \text{ all vectors } A, B, \text{ and } x \in R,$$
$$A * (B + C) = A * B + A * C, \text{ all vectors } A, B, C ,$$
$$(A + B) * C = A * C + B * C, \text{ all vectors } A, B, C .$$

Show that there are two non-zero vectors X, Y such that $X * Y = \tilde{\tilde{0}}$.

4. In R^3, let

$$A = (1,0,-2), \quad B = (-1,1,0), \quad C = (2,-1,1), \quad D = (0,3,1) .$$

(i) Find the area of the triangle A, B, C.

(ii) Find the perpendicular distance from A to the line through B, C.

(iii) Find the volume of the tetrahedron A, B, C, D .

(iv) Find the perpendicular distance from A to the plane through B, C, D.

(v) Find the perpendicular distance between the lines through A, B and through C, D.

5. Describe the locus represented by each of the following equations where X is a variable vector, A, B are constant

vectors, and $c \in R$ is constant:

(i) $$A \cdot X = c \ ,$$

(ii) $$X \times (X - A) = \tilde{0} \ ,$$

(iii) $$A \times X = B \qquad \text{(assuming } A \cdot B = 0\text{)} \ ,$$

(iv) $$X \cdot (X - A) = c \ ,$$

(v) $$|X - A| + |X - B| = 2c \ ,$$

(vi) $$|X \times A| = 1 \ .$$

Seven-dimensional cross product

From Wikipedia, the free encyclopedia

In mathematics, the **seven-dimensional cross product** is a bilinear operation on vectors in seven dimensional Euclidean space. It assigns to any two vectors **a**, **b** in $\mathbb{R}^7$ a vector $\mathbf{a} \times \mathbf{b}$ also in $\mathbb{R}^7$.[1] Like the cross product in three dimensions the seven-dimensional product is anticommutative and $\mathbf{a} \times \mathbf{b}$ is orthogonal to both **a** and **b**. Unlike in three dimensions, it does not satisfy the Jacobi identity. And while the three-dimensional cross product is unique up to a change in sign, there are many seven-dimensional cross products. The seven-dimensional cross product has the same relationship to octonions as the three-dimensional product does to quaternions.

The seven-dimensional cross product is one way of generalising the cross product to other than three dimensions, and it turns out to be the only other non-trivial bilinear product of two vectors that is vector valued, anticommutative and orthogonal.[2] In other dimensions there are vector-valued products of three or more vectors that satisfy these conditions, and binary products with bivector results.

Contents

Example

The postulates underlying construction of the seven-dimensional cross product are presented in the section Definition. As context for that discussion, the historically first example of the cross product is tabulated below using $\mathbf{e}_1$ to $\mathbf{e}_7$ as basis vectors.[3][4] This table is one of 480 independent multiplication tables fitting the pattern that each unit vector appears once in each column and once in each row.[5] Thus, each unit vector appears as a product in the table six times, three times with a positive sign and three with a negative sign because of antisymmetry about the diagonal of zero entries. For example, $\mathbf{e}_1 = \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_4 \times \mathbf{e}_5 = \mathbf{e}_7 \times \mathbf{e}_6$ and the negative entries are the reversed cross products.

Number	1	2	3	4	5	6	7
Letter	*i*	*j*	*k*	*l*	*il*	*jl*	*kl*

Alternate	*i*	*j*	*k*	*l*	*m*	*n*	*o*

Alternate indexing schemes

×	$\mathbf{e_1}$	$\mathbf{e_2}$	$\mathbf{e_3}$	$\mathbf{e_4}$	$\mathbf{e_5}$	$\mathbf{e_6}$	$\mathbf{e_7}$
$\mathbf{e_1}$	0	$\mathbf{e}_3$	$-\mathbf{e}_2$	$\mathbf{e}_5$	$-\mathbf{e}_4$	$-\mathbf{e}_7$	$\mathbf{e}_6$
$\mathbf{e_2}$	$-\mathbf{e}_3$	0	$\mathbf{e}_1$	$\mathbf{e}_6$	$\mathbf{e}_7$	$-\mathbf{e}_4$	$-\mathbf{e}_5$
$\mathbf{e_3}$	$\mathbf{e}_2$	$-\mathbf{e}_1$	0	$\mathbf{e}_7$	$-\mathbf{e}_6$	$\mathbf{e}_5$	$-\mathbf{e}_4$
$\mathbf{e_4}$	$-\mathbf{e}_5$	$-\mathbf{e}_6$	$-\mathbf{e}_7$	0	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e_5}$	$\mathbf{e}_4$	$-\mathbf{e}_7$	$\mathbf{e}_6$	$-\mathbf{e}_1$	0	$-\mathbf{e}_3$	$\mathbf{e}_2$
$\mathbf{e_6}$	$\mathbf{e}_7$	$\mathbf{e}_4$	$-\mathbf{e}_5$	$-\mathbf{e}_2$	$\mathbf{e}_3$	0	$-\mathbf{e}_1$
$\mathbf{e_7}$	$-\mathbf{e}_6$	$\mathbf{e}_5$	$\mathbf{e}_4$	$-\mathbf{e}_3$	$-\mathbf{e}_2$	$\mathbf{e}_1$	0

Cayley's sample multiplication table

Entries in the interior give the product of the corresponding vectors on the left and the top in that order (the product is anti-commutative). Some entries are highlighted to emphasize the symmetry.

The table can be summarized by the relation[4]

$$\mathbf{e}_i \times \mathbf{e}_j = \varepsilon_{ijk}\mathbf{e}_k \ ,$$

where ε_{ijk} is a completely antisymmetric tensor with a positive value +1 when *ijk* = 123, 145, 176, 246, 257, 347, 365. By picking out the factors leading to the unit vector $\mathbf{e}_1$, for example, one finds the formula for the $\mathbf{e}_1$ component of $\mathbf{x} \times \mathbf{y}$. Namely

$$(\mathbf{x} \times \mathbf{y})_1 = x_2y_3 - x_3y_2 + x_4y_5 - x_5y_4 + x_7y_6 - x_6y_7 = -(\mathbf{y} \times \mathbf{x})_1 \ .$$

The top left 3 × 3 corner of the table is the same as the cross product in three dimensions. It also may be noticed that orthogonality of the cross product to its constituents **x** and **y** is a requirement upon the entries in this table. However, because of the many possible multiplication tables, general results for the cross product are best developed using a basis-independent formulation, as introduced next.

Definition

We can define a cross product on a Euclidean space *V* as a bilinear map from *V* × *V* to *V* mapping vectors **x** and **y** in *V* to another vector **x** × **y** also in *V*, where **x** × **y** has the properties[1][6]

- **orthogonality:**

$$\mathbf{x} \cdot (\mathbf{x} \times \mathbf{y}) = (\mathbf{x} \times \mathbf{y}) \cdot \mathbf{y} = 0,$$

- **magnitude:**

$$|\mathbf{x} \times \mathbf{y}|^2 = |\mathbf{x}|^2|\mathbf{y}|^2 - (\mathbf{x} \cdot \mathbf{y})^2$$

where ($\mathbf{x}\cdot\mathbf{y}$) is the Euclidean dot product and $|\mathbf{x}|$ is the vector norm. The first property states that the cross product is perpendicular to its arguments, while the second property gives the magnitude of the cross product. An equivalent expression in terms of the angle θ between the vectors[7] is[8]

$$|\mathbf{x} \times \mathbf{y}| = |\mathbf{x}||\mathbf{y}| \sin\theta,$$

or the area of the parallelogram in the plane of **x** and **y** with the two vectors as sides.[9] As a third alternative the following can be shown to be equivalent to either expression for the magnitude:[10]

$$|\mathbf{x} \times \mathbf{y}| = |\mathbf{x}||\mathbf{y}| \text{ if } (\mathbf{x} \cdot \mathbf{y}) = 0.$$

Consequences of the defining properties

Given the three basic properties of (i) bilinearity, (ii) orthogonality and (iii) magnitude discussed in the section on definition, a nontrivial cross product exists only in three and seven dimensions.[2][8][10] This restriction upon dimensionality can be shown by postulating the properties required for the cross product, then deducing an equation which is only satisfied when the dimension is 0, 1, 3 or 7. In zero dimensions there is only the zero vector, while in one dimension all vectors are parallel, so in both these cases a cross product must be identically zero.

The restriction to 0, 1, 3 and 7 dimensions is related to Hurwitz's theorem, that normed division algebras are only possible in 1, 2, 4 and 8 dimensions. The cross product is derived from the product of the algebra by considering the product restricted to the 0, 1, 3, or 7 imaginary dimensions of the algebra. Again discarding trivial products the product can only be defined this way in three and seven dimensions.[11]

In contrast with three dimensions where the cross product is unique (apart from sign), there are many possible binary cross products in seven dimensions. One way to see this is to note that given any pair of vectors **x** and **y** $\in \mathbb{R}^7$ and any vector **v** of magnitude $|\mathbf{v}| = |\mathbf{x}||\mathbf{y}| \sin\theta$ in the five dimensional space perpendicular to the plane spanned by **x** and **y**, it is possible to find a cross product with a multiplication table (and an associated set of basis vectors) such that $\mathbf{x} \times \mathbf{y} = \mathbf{v}$. That leaves open the question of just how many vector pairs like **x** and **y** can be matched to specified directions like **v** before the limitations of any particular table intervene.

Another difference between the three dimensional cross product and a seven dimensional cross product is:[8]

> "…for the cross product $\mathbf{x} \times \mathbf{y}$ in $\mathbb{R}^7$ there are also other planes than the linear span of **x** and **y** giving the same direction as $\mathbf{x} \times \mathbf{y}$"
>
> —Pertti Lounesto, *Clifford algebras and spinors, p. 97*

This statement is exemplified by every multiplication table, because any specific unit vector selected as a product occurs as a mapping from three *different* pairs of unit vectors, once with a plus sign and once with a minus sign. Each of these different pairs, of course, corresponds to another plane being mapped into the same direction.

Further properties follow from the definition, including the following identities:

1. **Anticommutativity:**

$$\mathbf{x} \times \mathbf{y} = -\mathbf{y} \times \mathbf{x},$$

2. **Scalar triple product:**

$$\mathbf{x} \cdot (\mathbf{y} \times \mathbf{z}) = \mathbf{y} \cdot (\mathbf{z} \times \mathbf{x}) = \mathbf{z} \cdot (\mathbf{x} \times \mathbf{y})$$

3. **Malcev identity:**[8]

$$(\mathbf{x} \times \mathbf{y}) \times (\mathbf{x} \times \mathbf{z}) = ((\mathbf{x} \times \mathbf{y}) \times \mathbf{z}) \times \mathbf{x} + ((\mathbf{y} \times \mathbf{z}) \times \mathbf{x}) \times \mathbf{x} + ((\mathbf{z} \times \mathbf{x}) \times \mathbf{x}) \times \mathbf{y}$$

$$\mathbf{x} \times (\mathbf{x} \times \mathbf{y}) = -|\mathbf{x}|^2 \mathbf{y} + (\mathbf{x} \cdot \mathbf{y})\mathbf{x}.$$

Other properties follow only in the three dimensional case, and are not satisfied by the seven dimensional cross product, notably,

1. **Vector triple product:**

$$\mathbf{x} \times (\mathbf{y} \times \mathbf{z}) = (\mathbf{x} \cdot \mathbf{z})\mathbf{y} - (\mathbf{x} \cdot \mathbf{y})\mathbf{z}$$

2. **Jacobi identity:**[8]

$$\mathbf{x} \times (\mathbf{y} \times \mathbf{z}) + \mathbf{y} \times (\mathbf{z} \times \mathbf{x}) + \mathbf{z} \times (\mathbf{x} \times \mathbf{y}) = 0$$

Coordinate expressions

To define a particular cross product, an orthonormal basis $\{\mathbf{e}_j\}$ may be selected and a multiplication table provided that determines all the products $\{\mathbf{e}_i \times \mathbf{e}_j\}$. One possible multiplication table is described in the Example section, but it is not unique.[5] Unlike three dimensions, there are many tables because every pair of unit vectors is perpendicular to five other unit vectors, allowing many choices for each cross product.

Once we have established a multiplication table, it is then applied to general vectors **x** and **y** by expressing **x** and **y** in terms of the basis and expanding $\mathbf{x} \times \mathbf{y}$ through bilinearity.

Using $\mathbf{e}_1$ to $\mathbf{e}_7$ for the basis vectors a different multiplication table from the one in the Introduction, leading to a different cross product, is given with anticommutativity by[8]

$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_4, \quad \mathbf{e}_2 \times \mathbf{e}_4 = \mathbf{e}_1, \quad \mathbf{e}_4 \times \mathbf{e}_1 = \mathbf{e}_2,$$

$$\mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_5, \quad \mathbf{e}_3 \times \mathbf{e}_5 = \mathbf{e}_2, \quad \mathbf{e}_5 \times \mathbf{e}_2 = \mathbf{e}_3,$$

$$\mathbf{e}_3 \times \mathbf{e}_4 = \mathbf{e}_6, \quad \mathbf{e}_4 \times \mathbf{e}_6 = \mathbf{e}_3, \quad \mathbf{e}_6 \times \mathbf{e}_3 = \mathbf{e}_4,$$

$$\mathbf{e}_4 \times \mathbf{e}_5 = \mathbf{e}_7, \quad \mathbf{e}_5 \times \mathbf{e}_7 = \mathbf{e}_4, \quad \mathbf{e}_7 \times \mathbf{e}_4 = \mathbf{e}_5,$$

$$\mathbf{e}_5 \times \mathbf{e}_6 = \mathbf{e}_1, \quad \mathbf{e}_6 \times \mathbf{e}_1 = \mathbf{e}_5, \quad \mathbf{e}_1 \times \mathbf{e}_5 = \mathbf{e}_6,$$

$$\mathbf{e}_6 \times \mathbf{e}_7 = \mathbf{e}_2, \quad \mathbf{e}_7 \times \mathbf{e}_2 = \mathbf{e}_6, \quad \mathbf{e}_2 \times \mathbf{e}_6 = \mathbf{e}_7,$$

$\times$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$	$\mathbf{e}_7$
$\mathbf{e}_1$	0	$\mathbf{e}_4$	$\mathbf{e}_7$	$-\mathbf{e}_2$	$\mathbf{e}_6$	$-\mathbf{e}_5$	$-\mathbf{e}_3$
$\mathbf{e}_2$	$-\mathbf{e}_4$	0	$\mathbf{e}_5$	$\mathbf{e}_1$	$-\mathbf{e}_3$	$\mathbf{e}_7$	$-\mathbf{e}_6$
$\mathbf{e}_3$	$-\mathbf{e}_7$	$-\mathbf{e}_5$	0	$\mathbf{e}_6$	$\mathbf{e}_2$	$-\mathbf{e}_4$	$\mathbf{e}_1$
$\mathbf{e}_4$	$\mathbf{e}_2$	$-\mathbf{e}_1$	$-\mathbf{e}_6$	0	$\mathbf{e}_7$	$\mathbf{e}_3$	$-\mathbf{e}_5$
$\mathbf{e}_5$	$-\mathbf{e}_6$	$\mathbf{e}_3$	$-\mathbf{e}_2$	$-\mathbf{e}_7$	0	$\mathbf{e}_1$	$\mathbf{e}_4$
$\mathbf{e}_6$	$\mathbf{e}_5$	$-\mathbf{e}_7$	$\mathbf{e}_4$	$-\mathbf{e}_3$	$-\mathbf{e}_1$	0	$\mathbf{e}_2$
$\mathbf{e}_7$	$\mathbf{e}_3$	$\mathbf{e}_6$	$-\mathbf{e}_1$	$\mathbf{e}_5$	$-\mathbf{e}_4$	$-\mathbf{e}_2$	0

Lounesto's multiplication table

$$\mathbf{e}_7 \times \mathbf{e}_1 = \mathbf{e}_3, \quad \mathbf{e}_1 \times \mathbf{e}_3 = \mathbf{e}_7, \quad \mathbf{e}_3 \times \mathbf{e}_7 = \mathbf{e}_1 .$$

More compactly this rule can be written as

$$\mathbf{e}_i \times \mathbf{e}_{i+1} = \mathbf{e}_{i+3}$$

with $i = 1...7$ modulo 7 and the indices i, $i + 1$ and $i + 3$ allowed to permute evenly. Together with anticommutativity this generates the product. This rule directly produces the two diagonals immediately adjacent to the diagonal of zeros in the table. Also, from an identity in the subsection on consequences,

$$\mathbf{e}_i \times (\mathbf{e}_i \times \mathbf{e}_{i+1}) = -\mathbf{e}_{i+1} = \mathbf{e}_i \times \mathbf{e}_{i+3} \ ,$$

which produces diagonals further out, and so on.

The $\mathbf{e}_j$ component of cross product $\mathbf{x} \times \mathbf{y}$ is given by selecting all occurrences of $\mathbf{e}_j$ in the table and collecting the corresponding components of $\mathbf{x}$ from the left column and of $\mathbf{y}$ from the top row. The result is:

$$\begin{aligned}
\mathbf{x} \times \mathbf{y} = {} & (x_2y_4 - x_4y_2 + x_3y_7 - x_7y_3 + x_5y_6 - x_6y_5)\,\mathbf{e}_1 \\
& + (x_3y_5 - x_5y_3 + x_4y_1 - x_1y_4 + x_6y_7 - x_7y_6)\,\mathbf{e}_2 \\
& + (x_4y_6 - x_6y_4 + x_5y_2 - x_2y_5 + x_7y_1 - x_1y_7)\,\mathbf{e}_3 \\
& + (x_5y_7 - x_7y_5 + x_6y_3 - x_3y_6 + x_1y_2 - x_2y_1)\,\mathbf{e}_4 \\
& + (x_6y_1 - x_1y_6 + x_7y_4 - x_4y_7 + x_2y_3 - x_3y_2)\,\mathbf{e}_5 \\
& + (x_7y_2 - x_2y_7 + x_1y_5 - x_5y_1 + x_3y_4 - x_4y_3)\,\mathbf{e}_6 \\
& + (x_1y_3 - x_3y_1 + x_2y_6 - x_6y_2 + x_4y_5 - x_5y_4)\,\mathbf{e}_7.
\end{aligned}$$

As the cross product is bilinear the operator $\mathbf{x}\times-$ can be written as a matrix, which takes the form[*citation needed*]

$$T_{\mathbf{x}} = \begin{bmatrix}
0 & -x_4 & -x_7 & x_2 & -x_6 & x_5 & x_3 \\
x_4 & 0 & -x_5 & -x_1 & x_3 & -x_7 & x_6 \\
x_7 & x_5 & 0 & -x_6 & -x_2 & x_4 & -x_1 \\
-x_2 & x_1 & x_6 & 0 & -x_7 & -x_3 & x_5 \\
x_6 & -x_3 & x_2 & x_7 & 0 & -x_1 & -x_4 \\
-x_5 & x_7 & -x_4 & x_3 & x_1 & 0 & -x_2 \\
-x_3 & -x_6 & x_1 & -x_5 & x_4 & x_2 & 0
\end{bmatrix}.$$

The cross product is then given by

$$\mathbf{x} \times \mathbf{y} = T_{\mathbf{x}}(\mathbf{y}).$$

Different multiplication tables

Two different multiplication tables have been used in this article, and there are more.[5][12] These multiplication tables are characterized by the Fano plane,[13][14] and these are shown in the figure for the two tables used here: at top, the one described by Sabinin, Sbitneva, and Shestakov, and at bottom that described by Lounesto. The

numbers under the Fano diagrams (the set of lines in the diagram) indicate a set of indices for seven independent products in each case, interpreted as $ijk \rightarrow \mathbf{e}_i \times \mathbf{e}_j = \mathbf{e}_k$. The multiplication table is recovered from the Fano diagram by following either the straight line connecting any three points, or the circle in the center, with a sign as given by the arrows. For example, the first row of multiplications resulting in $\mathbf{e}_1$ in the above listing is obtained by following the three paths connected to $\mathbf{e}_1$ in the lower Fano diagram: the circular path $\mathbf{e}_2 \times \mathbf{e}_4$, the diagonal path $\mathbf{e}_3 \times \mathbf{e}_7$, and the edge path $\mathbf{e}_6 \times \mathbf{e}_1 = \mathbf{e}_5$ rearranged using one of the above identities as:

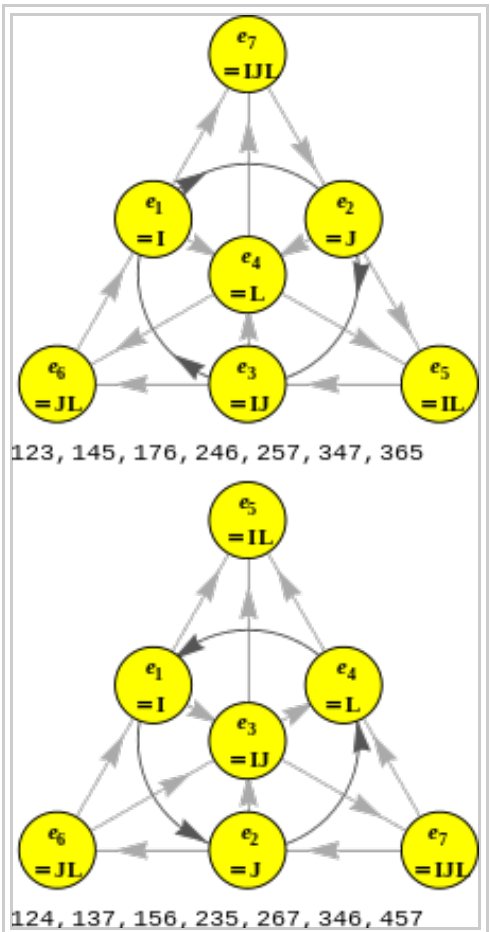
$$\mathbf{e_6} \times (\mathbf{e_6} \times \mathbf{e_1}) = -\mathbf{e_1} = \mathbf{e_6} \times \mathbf{e_5},$$

or

$$\mathbf{e_5} \times \mathbf{e_6} = \mathbf{e_1},$$

also obtained directly from the diagram with the rule that any two unit vectors on a straight line are connected by multiplication to the third unit vector on that straight line with signs according to the arrows (sign of the permutation that orders the unit vectors).

It can be seen that both multiplication rules follow from the same Fano diagram by simply renaming the unit vectors, and changing the sense of the center unit vector. The question arises: how many multiplication tables are there?[14]



Fano planes for the two multiplication tables used here.

> The question of possible multiplication tables arises, for example, when one reads another article on octonions, which uses a different one from the one given by [Cayley, say]. Usually it is remarked that all 480 possible ones are equivalent, that is, given an octonionic algebra with a multiplication table and any other valid multiplication table, one can choose a basis such that the multiplication follows the new table in this basis. One may also take the point of view, that there exist different octonionic algebras, that is, algebras with different multiplication tables. With this interpretation...all these octonionic algebras are isomorphic.
>
> —Jörg Schray, Corinne A Manogue, *Octonionic representations of Clifford algebras and triality (1994)*

Using geometric algebra

The product can also be calculated using geometric algebra. The product starts with the exterior product, a bivector valued product of two vectors:

$$\mathbf{B} = \mathbf{x} \wedge \mathbf{y} = \frac{1}{2}(\mathbf{xy} - \mathbf{yx}).$$

This is bilinear, alternate, has the desired magnitude, but is not vector valued. The vector, and so the cross product, comes from the product of this bivector with a trivector. In three dimensions up to a scale factor there is only one trivector, the pseudoscalar of the space, and a product of the above bivector and one of the two unit trivectors gives the vector result, the dual of the bivector.

A similar calculation is done is seven dimensions, except as trivectors form a 35-dimensional space there are many trivectors that could be used, though not just any trivector will do. The trivector that gives the same product as the above coordinate transform is

$$\mathbf{v} = \mathbf{e}_{124} + \mathbf{e}_{235} + \mathbf{e}_{346} + \mathbf{e}_{457} + \mathbf{e}_{561} + \mathbf{e}_{672} + \mathbf{e}_{713}.$$

This is combined with the exterior product to give the cross product

$$\mathbf{x} \times \mathbf{y} = -(\mathbf{x} \wedge \mathbf{y}) \lrcorner \mathbf{v}$$

where $\lrcorner$ is the left contraction operator from geometric algebra.[8][15]

Relation to the octonions

Just as the 3-dimensional cross product can be expressed in terms of the quaternions, the 7-dimensional cross product can be expressed in terms of the octonions. After identifying $\mathbb{R}^7$ with the imaginary octonions (the orthogonal complement of the real line in **O**), the cross product is given in terms of octonion multiplication by

$$\mathbf{x} \times \mathbf{y} = \operatorname{Im}(\mathbf{xy}) = \frac{1}{2}(\mathbf{xy} - \mathbf{yx}).$$

Conversely, suppose V is a 7-dimensional Euclidean space with a given cross product. Then one can define a bilinear multiplication on $\mathbb{R} \oplus V$ as follows:

$$(a, \mathbf{x})(b, \mathbf{y}) = (ab - \mathbf{x} \cdot \mathbf{y}, a\mathbf{y} + b\mathbf{x} + \mathbf{x} \times \mathbf{y}).$$

The space $\mathbb{R} \oplus V$ with this multiplication is then isomorphic to the octonions.[16]

The cross product only exists in three and seven dimensions as one can always define a multiplication on a space of one higher dimension as above, and this space can be shown to be a normed division algebra. By Hurwitz's theorem such algebras only exist in one, two, four, and eight dimensions, so the cross product must be in zero, one, three or seven dimensions. The products in zero and one dimensions are trivial, so non-trivial cross products only exist in three and seven dimensions.[17][18]

The failure of the 7-dimension cross product to satisfy the Jacobi identity is due to the nonassociativity of the octonions. In fact,

$$\mathbf{x} \times (\mathbf{y} \times \mathbf{z}) + \mathbf{y} \times (\mathbf{z} \times \mathbf{x}) + \mathbf{z} \times (\mathbf{x} \times \mathbf{y}) = -\frac{3}{2}[\mathbf{x}, \mathbf{y}, \mathbf{z}]$$

where $[\mathbf{x}, \mathbf{y}, \mathbf{z}]$ is the associator.

Rotations

In three dimensions the cross product is invariant under the group of the rotation group, SO(3), so the cross product of **x** and **y** after they are rotated is the image of $\mathbf{x} \times \mathbf{y}$ under the rotation. But this invariance is not true in seven dimensions; that is, the cross product is not invariant under the group of rotations in seven dimensions, SO(7). Instead it is invariant under the exceptional Lie group G_2, a subgroup of SO(7).[8][16]

Generalizations

Non-trivial binary cross products exist only in three and seven dimensions. But if the restriction that the product is binary is lifted, so products of more than two vectors are allowed, then more products are possible.[19][20] As in two dimensions the product must be vector valued, linear, and anti-commutative in any two of the vectors in the product.

The product should satisfy orthogonality, so it is orthogonal to all its members. This means no more than $n - 1$ vectors can be used in n dimensions. The magnitude of the product should equal the volume of the parallelotope with the vectors as edges, which is can be calculated using the Gram determinant. So the conditions are

- orthogonality:

$$(\mathbf{a_1} \times \cdots \times \mathbf{a_k}) \cdot \mathbf{a_j} = 0$$

- the Gram determinant:

$$|\mathbf{a_1} \times \cdots \times \mathbf{a_k}|^2 = \det(\mathbf{a_i} \cdot \mathbf{a_j}) = \begin{vmatrix} \mathbf{a_1} \cdot \mathbf{a_1} & \mathbf{a_1} \cdot \mathbf{a_2} & \dots & \mathbf{a_1} \cdot \mathbf{a_k} \\ \mathbf{a_2} \cdot \mathbf{a_1} & \mathbf{a_2} \cdot \mathbf{a_2} & \dots & \mathbf{a_2} \cdot \mathbf{a_k} \\ \dots & \dots & \dots & \dots \\ \mathbf{a_k} \cdot \mathbf{a_1} & \mathbf{a_k} \cdot \mathbf{a_2} & \dots & \mathbf{a_k} \cdot \mathbf{a_k} \end{vmatrix}$$

The Gram determinant is the squared volume of the parallelotope with $\mathbf{a}_1, \dots, \mathbf{a}_k$ as edges. If there are just two vectors $\mathbf{x}$ and $\mathbf{y}$ it simplifies to the condition for the binary cross product given above, that is

$$|\mathbf{x} \times \mathbf{y}|^2 = \begin{vmatrix} \mathbf{x} \cdot \mathbf{x} & \mathbf{x} \cdot \mathbf{y} \\ \mathbf{y} \cdot \mathbf{x} & \mathbf{y} \cdot \mathbf{y} \end{vmatrix} = |\mathbf{x}|^2|\mathbf{y}|^2 - (\mathbf{x} \cdot \mathbf{y})^2,$$

With these conditions a non-trivial cross product only exists:

- as a binary product in three and seven dimensions
- as a product of $n - 1$ vectors in $n > 3$ dimensions
- as a product of three vectors in eight dimensions

The product of $n - 1$ vectors is in n dimensions is the Hodge dual of the exterior product of $n - 1$ vectors. One version of the product of three vectors in eight dimensions is given by

$$\mathbf{a} \times \mathbf{b} \times \mathbf{c} = (\mathbf{a} \wedge \mathbf{b} \wedge \mathbf{c}) \lrcorner (\mathbf{w} - \mathbf{v}\mathbf{e}_8)$$

where $\mathbf{v}$ is the same trivector as used in seven dimensions, $\lrcorner$ is again the left contraction, and $\mathbf{w} = -\mathbf{v}\mathbf{e}_{12\dots7}$ is a 4-vector.

See also

- Composition algebra

Notes

1. ^ *a b* WS Massey (1983). "Cross products of vectors in higher dimensional Euclidean spaces". *The American Mathematical Monthly* (Mathematical Association of America) **90** (10): 697–701. doi:10.2307/2323537 (http://dx.doi.org/10.2307%2F2323537). JSTOR 2323537 (//www.jstor.org/stable/2323537).
2. ^ *a b* WS Massey (1983). "Cross products of vectors in higher dimensional Euclidean spaces". *The American Mathematical Monthly* **90** (10): 697–701. doi:10.2307/2323537 (http://dx.doi.org/10.2307%2F2323537). JSTOR 2323537 (//www.jstor.org/stable/2323537). "If one requires only three basic properties of the cross product ... it turns out that a cross product of vectors exists only in 3-dimensional and 7-dimensional Euclidean space."
3. ^ This table is due to Arthur Cayley (1845) and John T. Graves (1843). See G Gentili, C Stoppato, DC Struppa and F Vlacci (2009). "Recent developments for regular functions of a hypercomplex variable" (http://books.google.com/?id=H-5v6pPpyb4C&pg=PA168). In Irene Sabadini, M Shapiro, F Sommen. *Hypercomplex analysis* (Conference on quaternionic and Clifford analysis; proceedings ed.). Birkaüser. p. 168. ISBN 978-3-7643-9892-7.
4. ^ *a b* Lev Vasil′evitch Sabinin, Larissa Sbitneva, I. P. Shestakov (2006). "§17.2 Octonion algebra and its regular bimodule representation" (http://books.google.com/?id=_PEWt18egGgC&pg=PA235). *Non-associative algebra and its applications*. CRC Press. p. 235. ISBN 0-8247-2669-3
5. ^ *a b c* Rafał Abłamowicz, Pertti Lounesto, Josep M. Parra (1996). "§ Four ocotonionic basis numberings" (http://books.google.com/?id=OpbY_abijtwC&pg=PA202). *Clifford algebras with numeric and symbolic computations*. Birkhäuser. p. 202. ISBN 0-8176-3907-1.
6. ^ Mappings are restricted to be bilinear by (Massey 1993) and Robert B Brown and Alfred Gray (1967). "Vector cross products" (http://www.springerlink.com/content/a42n878560522255/). *Commentarii Mathematici Helvetici* (Birkhäuser Basel) **42** (1/December): 222–236. doi:10.1007/BF02564418 (http://dx.doi.org/10.1007%2FBF02564418)..
7. ^ The definition of angle in *n*-dimensions ordinarily is defined in terms of the dot product as:

$$(\mathbf{x}\cdot\mathbf{y}) = |\mathbf{x}||\mathbf{y}|\ \cos\theta\ ,\ \text{in the range}\ (-\pi < \theta \leq \pi)\ ,$$

where θ is the angle between the vectors. Consequently, this property of the cross product provides its magnitude as:

$$|\mathbf{x}\times\mathbf{y}|^2 = |\mathbf{x}|^2|\mathbf{y}|^2\left(1-\cos^2\theta\right)\ .$$

From the Pythagorean trigonometric identity this magnitude equals

$$|\mathbf{x}\times\mathbf{y}| = |\mathbf{x}||\mathbf{y}|\sin\theta.$$

See Francis Begnaud Hildebrand (1992). *Methods of applied mathematics* (http://books.google.com/?id=17EZkWPz_eQC&pg=PA24) (Reprint of Prentice-Hall 1965 2nd ed.). Courier Dover Publications. p. 24. ISBN 0-486-67002-3.
8. ^ *a b c d e f g h* Lounesto, pp. 96–97
9. ^ Kendall, M. G. (2004). *A Course in the Geometry of N Dimensions* (http://books.google.com/?id=_dFJ6pSzRLkC&pg=PA19). Courier Dover Publications. p. 19. ISBN 0-486-43927-5.
10. ^ *a b* Z.K. Silagadze (2002). "Multi-dimensional vector product". *Journal of Physics A: Mathematical and General* **35** (23): 4949. arXiv:math.RA/0204357 (//arxiv.org/abs/math.RA/0204357). doi:10.1088/0305-4470/35/23/310 (http://dx.doi.org/10.1088%2F0305-4470%2F35%2F23%2F310).
11. ^ Nathan Jacobson (2009). *Basic algebra I* (http://books.google.com/?id=_K04QgAACAAJ&dq=isbn=0-486-47189-6&cd=1) (Reprint of Freeman 1974 2nd ed.). Dover Publications. pp. 417–427. ISBN 0-486-47189-6.
12. ^ Further discussion of the tables and the connection of the Fano plane to these tables is found here: Tony Smith. "Octonion products and lattices" (http://www.valdostamuseum.org/hamsmith/480op.html). Retrieved 2010-07-11.
13. ^ Rafał Abłamowicz, Bertfried Fauser (2000). *Clifford Algebras and Their Applications in Mathematical Physics: Algebra and physics* (http://books.google.com/?id=yvCC94xzJG8C&pg=PA26). Springer. p. 26. ISBN 0-8176-4182-3.
14. ^ *a b* Jörg Schray, Corinne A. Manogue (1996). "Octonionic representations of Clifford algebras and triality" (http://www.springerlink.com/content/w1884mlmj88u5205/). *Foundations of physics* (Springer) **26** (1/January): 17–70. doi:10.1007/BF02058887 (http://dx.doi.org/10.1007%2FBF02058887). Available as ArXive preprint (http://arxiv.org/abs/hep-th/9407179v1) Figure 1 is located here (http://arxiv.org/PS_cache/hep-th/ps/9407/9407179v1.fig1-1.png).
15. ^ Bertfried Fauser (2004). "§18.4.2 Contractions" (http://books.google.com/?id=b6mbSCv_MHMC&pg=PA292). In

Pertti Lounesto, Rafał Abłamowicz. *Clifford algebras: applications to mathematics, physics, and engineering*. Birkhäuser. pp. 292 *ff*. ISBN 0-8176-3525-4.
16. ^ *a* *b* John C. Baez (2001). "The Octonions" (http://math.ucr.edu/home/baez/octonions/oct.pdf). *Bull. Amer. Math.* **39**: 38.
17. ^ Elduque, Alberto (2004). *Vector cross products* (http://www.unizar.es/matematicas/algebra/elduque/Talks/crossproducts.pdf).
18. ^ Darpö, Erik (2009). "Vector product algebras". *Bulletin of the London Mathematical Society* **41** (5): 898–902. doi:10.1112/blms/bdp066 (http://dx.doi.org/10.1112%2Fblms%2Fbdp066). See also: *Real vector product algebras*. CiteSeerX: 10.1.1.66.4 (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.66.4).
19. ^ Lounesto, §7.5: *Cross products of k vectors in* $\mathbb{R}^n$, p. 98
20. ^ Jean H. Gallier (2001). "Problem 7.10 (2)" (http://books.google.com/?id=CTHaW9ft1ZMC&pg=PA244#v=onepage&q). *Geometric methods and applications: for computer science and engineering*. Springer. p. 244. ISBN 0-387-95044-3.

References

- Brown, Robert B.; Gray, Alfred (1967). "Vector cross products". *Commentarii Mathematici Helvetici* **42** (1): 222–236. doi:10.1007/BF02564418 (http://dx.doi.org/10.1007%2FBF02564418).
- Lounesto, Pertti (2001). *Clifford algebras and spinors* (http://books.google.com/?id=kOsybQWDK4oC). Cambridge, UK: Cambridge University Press. ISBN 0-521-00551-5.
- Silagadze, Z.K. (2002). "Multi-dimensional vector product" (http://iopscience.iop.org/0305-4470/35/23/310). *J Phys A: Math Gen* **35** (23): 4949. doi:10.1088/0305-4470/35/23/310 (http://dx.doi.org/10.1088%2F0305-4470%2F35%2F23%2F310). Also available as ArXiv reprint arXiv:math.RA/0204357.
- Massey, W.S. (1983). "Cross products of vectors in higher dimensional Euclidian spaces". *The American Mathematical Monthly* **90** (10): 697–701. doi:10.2307/2323537 (http://dx.doi.org/10.2307%2F2323537). JSTOR 2323537 (//www.jstor.org/stable/2323537).

Categories: Bilinear operators | Binary operations | Octonions | Linear algebra | Vectors

---

V. ENDOMORPHISMS

§1. Introduction

Throughout this chapter, V will be a finite dimensional vector space with a scalar product. Our objective is to display the structure of certain kinds of endomorphisms (II, §9) of V, namely, the symmetric and skew-symmetric endomorphisms. However, certain results will be proved only for $\dim V \leq 3$. This is because the determinant of an endomorphism will not be defined in general until Chapter IX. For $\dim V \leq 3$, the notion of determinant can be derived from the triple product $A \cdot B \times C = [A, B, C]$ of Definition IV, 2.2, although it is actually independent of the notion of scalar or vector product (cf. Exercise 3.4).

§2. The determinant

To simplify the notation, we shall abbreviate $T(X)$ by TX.

2.1. Theorem. Let $T \in E(V)$, where $\dim V = 3$. Let A, B, C and A', B', C' be any two sets of independent vectors in V. Then

$$(1) \qquad \frac{[TA, TB, TC]}{[A, B, C]} = \frac{[TA', TB', TC']}{[A', B', C']} .$$

Proof. Note that $[A, B, C] \neq 0$ because A, B, C are independent (Theorem IV, 2.3, (iii)). Let i, j, k be an orthonormal basis in V with $k = i \times j$. If we can prove that

(2) $$[TA, TB, TC] = [A, B, C][Ti, Tj, Tk]$$

for any three vectors A, B, C, then formula (1) will follow because both ratios must equal $[Ti, Tj, Tk]$. In terms of the basis, let

$$A = a_1 i + a_2 j + a_3 k, \quad B = b_1 i + \ldots, \quad \text{etc.}$$

Since T is linear, the left side of (2) becomes

$$[a_1 Ti + a_2 Tj + a_3 Tk,\ b_1 Ti + b_2 Tj + b_3 Tk,\ c_1 Ti + c_2 Tj + c_3 Tk] .$$

Using the trilinearity of the triple product, Theorem IV, 2.3 (ii), we can expand this into a sum of 27 terms. By Theorem IV, 2.3 (iii), all but 6 of the 27 terms are zero because of a repeated vector in a triple product, so the sum equals

$$a_1 b_2 c_3 [Ti, Tj, Tk] + a_1 b_3 c_2 [Ti, Tk, Tj] + a_2 b_1 c_3 [Tj, Ti, Tk]$$
$$+ a_2 b_3 c_1 [Tj, Tk, Ti] + a_3 b_1 c_2 [Tk, Ti, Tj] + a_3 b_2 c_1 [Tk, Tj, Ti] .$$

Each of the triple products involve the same three vectors and by skew-symmetry, Theorem IV, 2.3 (i), equals $\pm [Ti, Tj, Tk]$. Thus the sum equals

$$(a_1 b_2 c_3 - a_1 b_3 c_2 - a_2 b_1 c_3 + a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1)[Ti, Tj, Tk] .$$

If we make the analogous calculation for $[A, B, C]$, the only change is that Ti, Tj, Tk are replaced by i, j, k. Since $k = i \times j$, we have $[i, j, k] = 1$; that is, the term in parenthesis above is exactly $[A, B, C]$. This proves (2) and completes the proof of the theorem.

2.2. <u>Definition</u>. Let $T \in E(V)$, where $\dim V = 3$. The

ratio (1) of Theorem 2.1 is called the <u>determinant</u> of T and is denoted by $\det T$.

<u>Remarks</u>. If, in (1), we take A, B, and C to be i, j, and $k = i \times j$, then the denominator is 1, and it follows that

$$\det T = [Ti, Tj, Tk] = Ti \cdot Tj \times Tk \quad . \tag{3}$$

Recall (IV, §2) that the triple product $[A, B, C]$ is, in absolute value, the volume of the parallelepiped spanned by A, B, C. Now T transforms a parallelepiped into another. Thus, if P is a parallelepiped with a vertex at the origin, $|\det T|$ is the ratio of the volumes of $T(P)$ and P. The restriction that P have a vertex at the origin can be deleted. In fact, adding a fixed vector A_0 to P gives a translated parallelepiped $A_0 + P$ having the same volume. The image of $A_0 + P$ is $T(A_0) + T(P)$ which again has the same volume as $T(P)$. Therefore

$$\text{vol}\,(T(P)) = |\det T|\ \text{vol}\,(P) \tag{4}$$

for any parallelepiped P.

2.3. <u>Proposition</u>. $\det T = 0$ if and only if T is singular, i.e. if and only if $\ker T \neq \tilde{0}$.

<u>Proof</u> ($\dim V = 3$). Let i, j, k be an orthonormal basis. Then $\det T = 0$ is equivalent, by (3), to $[Ti, Tj, Tk] = 0$. By Theorem IV, 2.3 (iii), this last is equivalent to the statement that Ti, Tj, Tk are dependent, i.e. $aTi + bTj + cTk = \tilde{0}$ for some numbers a, b, c not all zero. But this is equivalent to $T(ai + bj + ck) = \tilde{\tilde{0}}$, i.e. $\ker T$ contains a non-zero vector.

2.4. <u>Theorem</u>. Let $S, T \in E(V)$. Then

(i) $$\det(ST) = (\det S)(\det T) \quad ,$$

(ii) $$\det I = 1 \quad ,$$

(iii) if $T \in A(V)$,

$$\det(T^{-1}) = 1/\det T \quad .$$

<u>Proof</u> ($\dim V = 3$). If T is singular, so also is the composition ST, and (i) holds in this case. If T is non-singular, then, if A, B, C are independent, so are TA, TB, TC. Hence

$$\det(ST) = \frac{[STA, STB, STC]}{[A, B, C]}$$

$$= \frac{[STA, STB, STC]}{[TA, TB, TC]} \cdot \frac{[TA, TB, TC]}{[A, B, C]} = (\det S)(\det T) \ ,$$

so (i) holds in all cases. Formula (ii) is obvious from the definition of the determinant, since $IA = A$, etc. If T is an automorphism, take S to be T^{-1}; then $ST = T^{-1}T = I$, and (iii) follows from (i) and (ii).

2.5. <u>The case</u> $\dim V < 3$. If $\dim V = 2$, then V may be considered as a subspace of $V' = V \oplus R$ (Definition II, 11.8). Given $T \in E(V)$, let $T' \in E(V')$ be defined by $T'A = TA$ for $A \in V$ and $T'A = A$ for $A \in R$; then define

$$\det T = \det T' \quad .$$

Note that, if P is the parallelogram spanned by $A, B \in V$, then

(5) $$|\det T| = \frac{\text{area}\ (T(P))}{\text{area}\ (P)} \quad .$$

This follows from (4) and the fact that area $(P) = \text{vol}\ (P')$,

where P' is the parallelepiped spanned by $A, B \in V$ and a unit vector in R. If $\dim V = 1$, then any $T \in E(V)$ is of the form $TX = aX$, for some real number a, and we take $\det T = a$, with $|\det T| = \text{length}\ (TA)/\text{length}\ (A) = |TA|/|A|$ for any $A \in V$.

It is left as an exercise to give the proofs of Proposition 2.3 and Theorem 2.4 in the cases $\dim V = 1, 2$.

§3. Exercises

1. If T is orthogonal, Definition III, 6.5, show that $\det T = \pm 1$.

2. If $T : R^2 \longrightarrow R^2$ is linear, and (α_{ij}) is its matrix with respect to a basis, show that

$$\det T = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} .$$

3. If $T : R^3 \longrightarrow R^3$ is linear, and (α_{ij}) is its matrix with respect to a basis, show that

$$\det T = \alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} + \alpha_{12}\alpha_{23}\alpha_{31}$$
$$- \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31} .$$

4. Show that $\det T$ is independent of the choice of the vector and scalar products in V, where $\dim V = 3$.

§4. Proper vectors

4.1. Definition. A non-zero vector $A \in V$ is called a proper vector of $T \in E(V)$ if TA depends on A, i.e. if $TA = \lambda A$ for some $\lambda \in R$. The number λ is called a proper value of T.

Remarks. The adjective "characteristic" is often used

in the literature in place of "proper". To illustrate the concept, we consider some examples.

Suppose $a \in R$ and T is defined by $TX = aX$ for all X. Then each non-zero vector is proper, and each proper value is a.

An endomorphism may have no proper vectors, e.g. a rotation in R^2 through an angle not a multiple of π.

If $\dim V$ is 1, each non-zero vector is proper for any T and the proper value is $\det T$.

In R^3, let T be the reflection in the plane of i and j, that is, $Ti = i$, $Tj = j$, $Tk = -k$, where $k = i \times j$. Then each vector in the plane is proper with the value 1. The only other proper vectors are of the form ak with proper value -1.

Zero is a proper value of an endomorphism T if and only if T is singular.

4.2. <u>Proposition</u>. If λ is a proper value of T, then the set of vectors $X \in V$ such that $TX = \lambda X$ is a linear subspace of V of dimension ≥ 1.

<u>Proof</u>. Suppose $a \in R$, $TX = \lambda X$, and $TY = \lambda Y$. Using the linearity of T, we have

$$T(aX) = aTX = a\lambda X = \lambda(aX) \quad .$$

$$T(X + Y) = TX + TY = \lambda X + \lambda Y = \lambda(X + Y) \quad .$$

This shows that these vectors form a linear subspace. Since λ is a proper value, there is some corresponding proper vector which, by definition, is not zero. Hence the dimension is at

least one.

4.3. Proposition. If $A_1, \ldots, A_k$ are proper vectors of T having distinct proper values $\lambda_1, \ldots, \lambda_k$, then $A_1, \ldots, A_k$ are independent.

Proof. Suppose, to the contrary, that they are dependent. Let h be the smallest index such that A_h depends on $A_1, \ldots, A_{h-1}$ (see Proposition I, 9.4). Then

$$A_h = \Sigma_{i=1}^{h-1} x_i A_i \quad .$$

Applying T to both sides gives

$$\lambda_h A_h = \Sigma_{i=1}^{h-1} x_i \lambda_i A_i \quad ;$$

but also

$$\lambda_h A_h = \lambda_h \Sigma_{i=1}^{h-1} x_i A_i = \Sigma_{i=1}^{h-1} x_i \lambda_h A_i \quad ;$$

and therefore

$$\Sigma_{i=1}^{h-1} x_i(\lambda_i - \lambda_h)A_i = \tilde{0} \quad .$$

Since $A_1, \ldots, A_{h-1}$ are independent, each coefficient is zero: $x_i(\lambda_i - \lambda_h) = 0$. But λ_h is distinct from each λ_i for $i < h$. Therefore each $x_i = 0$. This means $A_h = \tilde{0}$. This contradiction proves the proposition.

4.4. Corollary. If $\dim V = n$, then T has at most n distinct proper values.

4.5. Definition. Let $T \in E(V)$ and, for each $x \in R$, define $T_x \in E(V)$ by $T_x = T - xI$. Then $\det T_x$, as a function of x, is a polynomial of degree $n = \dim V$ called the characteristic polynomial of T.

<u>Proof</u> ($n = 3$). For any independent $A, B, C \in V$,

$$\begin{aligned}[A, B, C] \det T_x &= [TA - xA, TB - xB, TC - xC] \\ &= [TA, TB, TC] \\ &\quad - x\,\{[A, TB, TC] + [TA, B, TC] + [TA, TB, C]\} \\ &\quad + x^2\{[A, B, TC] + [A, TB, C] + [TA, B, C]\} \\ &\quad - x^3[A, B, C] \quad ,\end{aligned}$$

so $\det T_x$ is a polynomial in x of degree 3.

For general values of n, we have

$$(1) \qquad \det T_x = (-1)^n x^n + (-1)^{n-1}(\text{trace } T)x^{n-1} + \ldots + \det T \quad ,$$

where

4.6. <u>Definition</u>. The <u>trace</u> of an endomorphism is the sum of the roots of its characteristic polynomial.

Note that multiple roots must be included in the sum according to their multiplicities, and that complex (non-real) roots occur as conjugate pairs, since the coefficients in (1) are real numbers.

4.7. <u>Theorem</u>. The set of proper values of T coincides with the set of real roots of the characteristic polynomial of T.

<u>Proof</u>. The statement "λ is a root of the characteristic polynomial" means that $\det T_\lambda = 0$. By Proposition 2.3, this is equivalent to the existence of a non-zero vector A such that $T_\lambda A = \tilde{\tilde{0}}$. By definition of T_λ, this is equivalent to $TA = \lambda A$; and hence equivalent to the statement that λ is a proper value of T.

Since a polynomial of odd degree with real coefficients has at least one real root, we have

4.8. Corollary. If $T \in E(V)$ where $n = \dim V$ is odd, then the characteristic polynomial of T has at least one real root, and T has at least one proper vector.

4.9. Theorem. Let $T \in E(V)$, $\dim V = n$. If the characteristic polynomial of T has n distinct roots $\lambda_1, \ldots, \lambda_n$, then there is a basis $A_1, \ldots, A_n$ for V such that $TA_i = \lambda_i A_i$ for $i = 1, \ldots, n$.

Proof. By Theorem 4.7, each root λ_i is a proper value of T, so there exists a non-zero vector A_i such that $TA_i = \lambda_i A_i$. By Proposition 4.3, the vectors $A_1, \ldots, A_n$ are independent, since the λ_i are distinct, and therefore form a basis for V.

§5. The adjoint

Let $T \in E(V)$. For each $Y \in V$, consider the linear transformation $V \longrightarrow R$ defined by $X \longrightarrow TX \cdot Y$. By Theorem III, 5.8, there is a unique vector, which will be denoted by T^*Y, such that this transformation is given by $X \longrightarrow X \cdot T^*Y$, that is,

$$(1) \qquad TX \cdot Y = X \cdot T^*Y \qquad \text{for all } X \in V .$$

5.1. Definition. The function T^* which assigns to $X \in V$ the vector $T^*X \in V$ is called the adjoint of T.

Note that this assignment depends, in general, on the scalar product in V.

5.2. Lemma. If $A, B, \in V$, and $X \cdot A = X \cdot B$ for

all $X \in V$, then $A = B$.

Proof. $X \cdot A = X \cdot B$ implies $X \cdot (A - B) = 0$. Since this holds for all X, it holds for $X = A - B$:

$$0 = (A - B) \cdot (A - B) = |A - B|^2 \quad .$$

Therefore $|A - B| = 0$; hence $A - B = \tilde{0}$.

5.3. Proposition. The adjoint T^* of T is linear; that is, $T^* \in E(V)$.

Proof. If $a \in R$ and $X, Y \in V$, then

$$X \cdot T^*(aY) = TX \cdot aY = a(TX \cdot Y) = a(X \cdot T^*Y) = X \cdot aT^*Y \quad .$$

Since this holds for each X, Lemma 5.2 gives $T^*(aY) = aT^*Y$.

Now let $X, Y, Y' \in V$. Then

$$X \cdot T^*(Y + Y') = TX \cdot (Y + Y') = TX \cdot Y + TX \cdot Y'$$

$$= X \cdot T^*Y + X \cdot T^*Y' = X \cdot (T^*Y + T^*Y') \quad .$$

Since this holds for each X, Lemma 5.2 gives $T^*(Y + Y') = T^*Y + T^*Y'$.

5.4. Lemma. With respect to an orthonormal basis, the matrix (α^*_{ij}) of T^* is the transpose of the matrix (α_{ij}) of T, that is,

$$\alpha^*_{ij} = \alpha_{ji} \quad .$$

Proof. If $A_1, \ldots, A_n$ is the orthonormal basis, we apply II, §5, and Proposition III, 5.2, and obtain

$$\alpha^*_{ij} = T^*A_j \cdot A_i = A_i \cdot T^*A_j = TA_i \cdot A_j = \alpha_{ji} \quad .$$

Using this lemma, we can give examples of adjoint transformations as follows.

If $n = 1$, then $T^* = T$ for any T because $\alpha^*_{11} = \alpha_{11}$.

If $n = 2$, and T is the rotation about $\tilde{0}$ through the angle θ, the matrix of T with respect to the standard basis is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

The transpose is

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} .$$

Therefore T^* is the rotation through the angle $-\theta$. If $\theta = \pi/2$, then $T^* = -T$. If $\theta = \pi$, then $T^* = T$.

If $n = 3$, and T is defined by $TX = A \times X$ for a fixed vector A, then $T^* = -T$ because

$$X \cdot T^* Y = TX \cdot Y = A \times X \cdot Y = X \cdot (-A \times Y) .$$

5.5. <u>Proposition</u>. As a function from $E(V)$ to $E(V)$, the passage to the adjoint has the following properties:

(i) $(S + T)^* = S^* + T^*$ for all $S, T \in E(V)$,

(ii) $(aT)^* = aT^*$ for all $T \in E(V)$, $a \in R$.

(iii) $T^{**} = T$ for all $T \in E(V)$.

<u>Remarks</u>. Because of (iii), this function is bijective, by Exercise II, 3.2. Since it is also linear, by (i) and (ii), it is an automorphism of $E(V)$, considered as a vector space, and an involution of $E(V)$ (cf. Exercise II, 10.5) or an automorphism "of period 2".

<u>Proof</u>. For any $X, Y \in V$, we have

$$X\cdot(S + T)^*Y = (S + T)X\cdot Y = SX\cdot Y + TX\cdot Y$$
$$= X\cdot S^*Y + X\cdot T^*Y = X\cdot(S^*Y + T^*Y)$$
$$= X\cdot(S^* + T^*)Y \quad .$$

Since, for a fixed Y, this holds for all X, Lemma 5.2 gives

$$(S + T)^*Y = (S^* + T^*)Y \quad .$$

Since this holds for each Y, (i) is true.

For any $X, Y \in V$, $a \in R$, we have

$$X\cdot(aT)^*Y = (aT)X\cdot Y = a(TX)\cdot Y = a(TX\cdot Y)$$
$$= a(X\cdot T^*Y) = X\cdot aT^*Y \quad .$$

Since this holds for all X, Lemma 5.2 gives $(aT)^*Y = aT^*Y$, and therefore (ii) is true.

For any $X, Y \in V$, we have

$$X\cdot T^{**}Y = T^*X\cdot Y = Y\cdot T^*X = TY\cdot X = X\cdot TY \quad ,$$

from which (iii) follows.

5.6. <u>Definition</u>. An endomorphism T of V is said to be <u>symmetric</u> (or self-adjoint) if $T^* = T$. It is said to be <u>skew-symmetric</u> if $T^* = -T$.

The examples given above should be reviewed; some are symmetric, and some are skew-symmetric.

5.7. <u>Proposition</u>. For any $T \in E(V)$ the endomorphisms

$$\frac{1}{2}(T + T^*) \qquad \text{and} \qquad \frac{1}{2}(T - T^*)$$

are symmetric and skew-symmetric, respectively. [They are called the <u>symmetric</u> and <u>skew-symmetric parts</u> of T, since $T = \frac{1}{2}(T + T^*) + \frac{1}{2}(T - T^*)$].

Proof. Using Proposition 5.5, we have

$$\left(\tfrac{1}{2}\,(T \pm T^*)\right)^* = \tfrac{1}{2}\,(T \pm T^*)^* = \tfrac{1}{2}\,(T^* \pm T^{**}) = \tfrac{1}{2}\,(T^* \pm T)$$

$$= \pm\,\tfrac{1}{2}\,(T \pm T^*) \quad .$$

§6. Exercises

1. Let $T : R^3 \longrightarrow R^3$ be linear, and let (α_{ij}) be its matrix with respect to a basis. Give the matrix representing $T_x = T - xI$ with respect to the same basis. Compute the characteristic polynomial of T by evaluating $\det T_x$, and show that

$$\text{trace } T = \alpha_{11} + \alpha_{22} + \alpha_{33} \quad .$$

2. Let $S, T \in E(V)$. Show that $(ST)^* = T^*S^*$.

3. If $T \in A(V)$ is orthogonal, show that $T^* = T^{-1}$.

4. If $TA = \lambda A$ and $T^*B = \mu B$ with $\lambda \neq \mu$, show that $A \cdot B = 0$.

5. If T is both symmetric and skew-symmetric, show that $TX = \tilde{0}$ for all X.

6. If $\dim V = 3$, show that T and T^* have the same determinant and the same characteristic polynomial. Hence they have the same trace and the same proper values.

7. If $\dim V = 3$, show that the trace of any skew-symmetric endomorphism is zero.

8. If $\dim V = 3$, show that T and $\frac{1}{2}\,(T + T^*)$ have the same trace.

§7. Symmetric endormorphisms

7.1. Proposition. Let T be a symmetric endomorphism

of V, and let T be a linear subspace of V such that $T(U) \subset U$. Then the orthogonal complement $U^{\perp}$ of U (see Proposition III, 5.3) has the analogous property: $T(U^{\perp}) \subset U^{\perp}$.

Proof. Suppose $A \in U^{\perp}$, and $B \in U$. The hypothesis $T(U) \subset U$ gives $TB \in U$, so $A \cdot TB = 0$. Since T is symmetric, $TA \cdot B = A \cdot TB$. Hence $TA \cdot B = 0$ for all $B \in U$; that is, $TA \in U^{\perp}$.

The following "structure theorem" for symmetric transformations is true for a V of any finite dimension; but we give a proof only for the dimensions 1, 2, and 3.

7.2. Theorem. Let $n = \dim V$, and let T be a symmetric endomorphism of V. Then there exist an orthonormal basis $A_1, \ldots, A_n$ and real numbers $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct) such that $TA_i = \lambda_i A_i$ for $i = 1, \ldots, n$.

Remark. If the characteristic polynomial of T has n distinct real roots $\lambda_1, \ldots, \lambda_n$ (in particular, if $n = 1$), the required orthonormal basis can be constructed as follows: The basis $A_1, \ldots, A_n$ of Theorem 4.9 can be chosen so that $A_i \cdot A_i = 1$, $i = 1, \ldots, n$ (why?). Then $A_i \cdot A_j = 0$ for $i \neq j$ by Exercise 6.4, since $T = T^*$. However, the theorem is to be proved for any symmetric T and is, in fact, equivalent to the statement that the characteristic polynomial of a symmetric endomorphism T has n real roots (not necessarily distinct).

Proof ($n = 2, 3$). Let $n = 2$. By Lemma 5.4, the matrix of T with respect to an orthonormal basis must be symmetric, that is, of the form

$$\begin{pmatrix} a & c \\ c & b \end{pmatrix}$$

Its characteristic polynomial is therefore

$$(a - x)(b - x) - c^2 = x^2 - (a + b)x + ab - c^2 \quad .$$

The roots are $\frac{1}{2}(a + b \pm \sqrt{(a - b)^2 + 4c^2})$, and these are real. The case that the roots are distinct is covered in the Remark above. If the roots coincide, then $(a - b)^2 + 4c^2 = 0$, whence $a = b$ and $c = 0$, and $TX = aX$ for all X. In this case $\lambda_1 = \lambda_2 = a$ and A_1, A_2 can be any orthonormal basis.

Let $n = 3$. By Corollary 4.8, there is a proper vector A_1 of T with proper value λ_1. We can assume that $|A_1| = 1$. Let U be the linear subspace of vectors orthogonal to A_1. By Proposition 7.1, T transforms U into U. Since T gives a symmetric endomorphism of U and $\dim U = 2$, the result already proved gives an orthonormal basis A_2, A_3 in U of proper vectors with proper values λ_2, λ_3. Then A_1, A_2, A_3 is the required basis in V.

§8. Skew-symmetric endomorphisms

8.1. Proposition. If $T \in E(V)$ is skew-symmetric, then $TX \cdot X = 0$ for all $X \in V$.

Proof. Skew-symmetry gives $TX \cdot X = - X \cdot TX$. The scalar product is commutative, so $TX \cdot X = X \cdot TX$. Hence $X \cdot TX = - X \cdot TX$, and the only number which equals its negative is 0.

8.2. Corollary. A skew-symmetric endomorphism has no proper value different from zero.

Proof. Let $T \in E(V)$ be skew-symmetric and suppose that a non-zero vector A satisfies $TA = \lambda A$. By Proposition 8.1, we have $0 = TA \cdot A = \lambda(A \cdot A)$, from which follows $\lambda = 0$.

8.3. Theorem. If $\dim V = 2$, and T is skew-symmetric, then T is a $90°$ rotation followed by multiplication by a scalar.

Proof. With respect to an orthonormal basis, the matrix of T must be skew-symmetric: $\alpha_{ij} = -\alpha_{ji}$. In particular $\alpha_{ii} = 0$. Hence it has the form

$$\begin{pmatrix} 0 & -a \\ a & 0 \end{pmatrix} = \begin{pmatrix} a\cos 90° & -a\sin 90° \\ a\sin 90° & a\cos 90° \end{pmatrix}.$$

8.4. Theorem. If $\dim V = 3$, and T is skew-symmetric, then there exists a vector $A \in V$ such that $TX = A \times X$ for all X.

Proof. By Corollary 4.8, T has a proper vector, which we choose of unit length and denote by k. By Corollary 8.2, $Tk = \tilde{0}$. Let U be the linear subspace of vectors orthogonal to k. For any $X \in V$, we have

$$TX \cdot k = X \cdot (-Tk) = X \cdot \tilde{0} = 0 \quad .$$

That is, $T(V) \subset U$. In particular, T gives a skew-symmetric endomorphism of U. Let i, j be an orthonormal basis in U, chosen so that $i \times j = k$. Then, by Theorem 8.3, $Ti = aj$ and $Tj = -ai$ for some scalar a. Set $A = ak$, and let $S \in E(V)$ be defined by $SX = A \times X$. Then it is easily verified that $Si = Ti$, $Sj = Tj$, $Sk = Tk$, from which follows $S = T$.

§9. Exercises

1. Show that an endomorphism satisfying the conclusion of Theorem 7.2 is always symmetric.

2. Show that the vector A in Theorem 8.4 is unique.

3. With respect to the standard basis i, j and $k = i \times j$ in R^3, let T be represented by the matrix

$$\begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}.$$

Let $A = -ci + bj - ak$. Show that $TX = A \times X$ for all X.

4. With respect to the standard basis in R^3, let T be given by the matrix

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

Find the matrices of the symmetric and skew-symmetric parts T', T'' of T. Find the proper vectors and proper values of T'.

VI. VECTOR-VALUED FUNCTIONS OF A SCALAR

Our main objective is the study of the calculus of functions $F: D \longrightarrow W$ where D is a subset of a finite dimensional vector space V, and W is a finite dimensional vector space. In this chapter, we shall concentrate on the special case $\dim V = 1$ with D an interval. Then F defines a curve in W, and we are led to the consideration of velocities and accelerations along curves. In the next chapter, we shall consider the special case $\dim W = 1$. Then F is a scalar function of a vector. Chapter VIII will treat the general case where F is a vector-valued function of a vector.

§1. Limits and continuity

In this section V and W are vector spaces, on each of which a scalar product is given, and D is a subset of V.

1.1. Definition. Let $F: D \longrightarrow W$, and let $A \in V$. We then say that F has the limit $B \in W$ as X tends toward A, written

$$\lim_{X \longrightarrow A} F(X) = B, \quad \text{or} \quad \lim_A F = B \quad ,$$

if, for each positive number ε, there is a positive number δ such that $X \in D$ and $0 < |X - A| < \delta$ imply $|F(X) - B| < \varepsilon$.

Remark. In the above definition, it is assumed that the set of vectors $X \in V$ satisfying $X \in D$ and $0 < |X - A| < \delta$ is not the empty set.

1.2. Definition. The function $F: D \longrightarrow W$ is said to be continuous at $A \in D$ if $\lim_A F$ exists and is $F(A)$. The

function F is said to be <u>continuous in D</u> if it is continuous at each vector of D.

<u>Remark</u>. These definitions are precisely the usual ones when $V = W = R$, and require only the notion of absolute value or distance (here defined by means of the scalar product) in the domain and range to make sense. It will be shown in Chapter X that the notions of limit and continuity are independent of the choice of scalar product if V and W are finite dimensional.

1.3. <u>Theorem</u>. Let F and G be functions from D to W, and let $f: D \longrightarrow R$ be a scalar function. Let $A \in D$ and assume that $\lim_A F = B$, $\lim_A G = C$, and $\lim_A f = d$. Then the following limits exist and have the indicated values:

(i) $\lim_A (F + G) = B + C$,

(ii) $\lim_A fF = dB$,

(iii) $\lim_A (F \cdot G) = B \cdot C$,

(iv) $\lim_A (F \times G) = B \times C$, (if $\dim W = 3$).

1.4. <u>Corollary</u>. If F, G and f are continuous in D, then so also are $F + G$, fF, $F \cdot G$ and $F \times G$ (if $\dim W = 3$).

<u>Proof</u>. If we examine the standard proofs of (i) and (ii) in the case $V = W = R$, it will be observed that the only properties of addition, absolute value and multiplication needed in the proofs are possessed by these operations in any vector space W, whether the multiplication be multiplication by a scalar, or scalar product, or vector product (if $\dim W = 3$). Therefore the proofs hold without change. We illustrate this wholesale method by giving the proof of (iv) in detail.

For any $X \in D$,

$$(1) \qquad |F(X) \times G(X) - B \times C|$$

$$= |F(X) \times G(X) - F(X) \times C + F(X) \times C - B \times C|$$

$$= |F(X) \times (G(X) - C) + (F(X) - B) \times C|$$

$$\leq |F(X) \times (G(X) - C)| + |(F(X) - B) \times C|$$

$$(2) \qquad \leq |F(X)|\ |G(X) - C| + |F(X) - B|\ |C| \quad .$$

The first step above uses properties of addition (Axioms 4 and 3 of Definition I, 1.1.). The second step uses the linearity of $\times$ (Axiom V2 of Definition IV, 1.1). The third is a property of absolute value (Theorem III, 3.2 (v)). The fourth follows from Axiom V5 of Definition IV, 1.1.

Let an $\varepsilon > 0$ be given. Set $\varepsilon_1 = \mathrm{Min}(\varepsilon/(2|C| + 2), 1)$. Then, since $\lim_A F = B$, there is a $\delta_1 > 0$ such that $X \in D$ and $0 < |X - A| < \delta_1$ imply

$$|F(X) - B| < \mathrm{Min}\left(\frac{\varepsilon}{2\ |C| + 2}, 1\right) ,$$

and therefore

$$(3) \qquad |F(X) - B|\ |C| < \tfrac{1}{2}\ \varepsilon \quad ,$$

and

$$(4) \qquad |F(X)| < |B| + 1 \quad .$$

Since $\lim_A G = C$, there is a $\delta_2 > 0$ such that $X \in D$ and $0 < |X - A| < \delta_2$ imply

$$(5) \qquad |G(X) - C| < \frac{\varepsilon}{2\ |B| + 2} \quad .$$

Now let $\delta = \mathrm{Min}\ (\delta_1, \delta_2)$. Then $X \in D$ and $0 < |X - A| < \delta$ imply (3), (4) and (5). Next, (4) and (5) together give

(6) $$|F(X)|\ |G(X) - C| < \frac{1}{2}\ \varepsilon\ .$$

Finally, (3) and (6) show that (2) is less than ε, and therefore (1) is less than ε. This concludes the proof.

Now suppose that W is finite dimensional, and let $B_1, \ldots, B_n$ be a basis in W. For any $F : D \longrightarrow W$, we can define $f_i : D \longrightarrow R$ by

(7) $$F(X) = \Sigma_{i=1}^{n} f_i(X)B_i, \qquad \text{for all } X \in D\ ,$$

where, if the basis is orthonormal,

(8) $$f_i(X) = F(X)\cdot B_i\ , \qquad i = 1, \ldots, n\ .$$

1.5. <u>Definition</u>. The scalar-valued functions f_i are called the <u>component functions</u> (or components) of F relative to the basis $B_1, \ldots, B_n$.

1.6. <u>Proposition</u>. $\lim_A F$ exists if and only if $\lim_A f_i$ exists for each $i = 1, \ldots, n$, and then

(9) $$\lim_A F = \Sigma_{i=1}^{n} (\lim_A f_i)B_i\ .$$

In particular, $F: D \longrightarrow W$ is continuous if and only if its component functions, relative to any basis in W, are continuous.

<u>Proof</u>. The constant function $G: D \longrightarrow W$ obtained by setting $G(X) = B_i$ is obviously continuous. If the basis is orthonormal, then $\lim_A F = C$ implies, by Theorem 1.3 (iii), that

$$\lim_A f_i = \lim_A F\cdot B_i = C\cdot B_i\ .$$

If the given basis is not orthonormal, a component function is a linear combination of the component functions relative to an orthonormal basis and we again have, by Theorem 1.3, (i) and (ii), that $\lim_A f_i$ exists if $\lim_A F$ exists.

Conversely, suppose $\lim_A f_i$ exists for each $i = 1, \ldots, n$. Applying Theorem 1.3, (i) and (ii), we have

$$\lim_A F = \lim_A (\Sigma_{i=1}^n f_i B_i)$$
$$= \Sigma_{i=1}^n \lim_A (f_i B_i) = \Sigma_{i=1}^n (\lim_A f_i) B_i \quad .$$

Thus $\lim_A F$ exists and is given by (9).

1.7. Proposition. If $T: V \longrightarrow W$ is linear where V and W are finite dimensional, then T is continuous.

Proof. By Proposition 1.6, it suffices to prove the proposition when $W = R$. By Theorem III, 5.8, there is a vector $B \in V$ such that $TX = B \cdot X$ for each X. Then, for $A \in V$, we have

$$|TX - TA| = |T(X - A)| = |B \cdot (X - A)| \leq |B| \; |X - A| \quad .$$

Therefore $|X - A| < \delta = \varepsilon/(|B| + 1)$ implies $|TX - TA| < \varepsilon$.

§2. The derivative

2.1. Definition. Let $[a, b]$ denote the closed interval of real numbers t such that $a \leq t \leq b$, $a < b$. Let W be a vector space. A continuous function $F: [a, b] \longrightarrow W$ is called a curve in W connecting $F(a)$ to $F(b)$.

2.2. Definition. For a fixed t in $[a, b]$ the difference quotient

$$\frac{F(t+h) - F(t)}{h}$$

is a vector-valued function of the scalar h defined for $-(t - a) \leq h \leq b - t$, excepting $h = 0$. If the difference quotient has a limit as h tends to zero, then F is said to

have a <u>derivative</u> $F'(t)$ at t, where

$$F'(t) = \lim_{h \to 0} \frac{F(t+h) - F(t)}{h} .$$

2.3. <u>Definition</u>. The curve is said to be <u>smooth</u> if it has a derivative $F'(t)$ for each t in $[a, b]$ and if $F': [a, b] \longrightarrow W$ is continuous in $[a, b]$. A curve is said to be <u>piecewise smooth</u> if its domain $[a, b]$ can be subdivided into a finite number of intervals over each of which the curve is smooth (i.e. the curve is smooth except for a finite number of "corners").

2.4. <u>Physical interpretation</u>. If the variable t represents time, then the (terminal point of the) vector $F(t)$ traverses the curve as t varies from a to b. The difference quotient represents the average velocity during the time interval from t to $t + h$, and its limit $F'(t)$ is the (<u>instantaneous</u>) <u>velocity vector</u>. It is customary to picture $F'(t)$ as a vector whose initial point is (the terminal point of) $F(t)$. If t is such that $F'(t) \neq \tilde{0}$, then $F'(t)$ spans a 1-dimensional linear subspace. The line through $F(t)$ parallel to $F'(t)$ is called the <u>tangent</u> to the curve at $F(t)$. The length $|F'(t)|$ is called the scalar velocity.

2.5. <u>Theorem</u>. Let F and G be curves: $[a, b] \longrightarrow W$, and let $f: [a, b] \longrightarrow R$ be continuous. If F, G, and f have derivatives at $t \in [a, b]$, then $F + G$, fF, $F \cdot G$ and $F \times G$ (if $\dim W = 3$) have derivatives at t, and

(i) $\quad (F + G)' = F' + G'$,

(ii) $(fF)' = f'F + fF'$,

(iii) $(F\cdot G)' = F' \cdot G + F \cdot G'$,

(iv) $(F \times G)' = F' \times G + F \times G'$ (if $\dim W = 3$) .

Proof. The standard proofs for $\dim W = 1$ remain valid without restricting the dimension. To illustrate, consider (iv). We have

$$\lim_{h \to 0} \frac{1}{h} \{F(t + h) \times G(t + h) - F(t) \times G(t)\}$$

$$= \lim_{h \to 0} \{F(t + h) \times \frac{G(t+h) - G(t)}{h} + \frac{F(t+h) - F(t)}{h} \times G(t)\} .$$

We then apply Theorem 1.3 to obtain (iv).

2.6. Proposition. If F is smooth, and if $F(t) \neq \tilde{0}$ for each $t \in [a, b]$, then the scalar function $|F(t)|$ is differentiable and

$$|F|' = F\cdot F'/|F| .$$

Proof. The usual rules of calculus, plus Theorem 2.5, (iii), give

$$|F|' = \frac{d}{dt} (F\cdot F)^{1/2} = \frac{1}{2} (F\cdot F)^{-1/2}(F'\cdot F + F\cdot F') .$$

We then use $F\cdot F' = F'\cdot F$.

Now assume that W is finite dimensional.

2.7. Proposition. Let $F: [a, b] \longrightarrow W$ be a curve. Let $B_1, \ldots, B_n$ be a basis in W, so that

(10) $$F(t) = \Sigma_{i=1}^{n} f_i(t)B_i$$

where, in the case that the basis is orthonormal,

(11) $$f_i(t) = F(t)\cdot B_i \quad .$$

Then F has a derivative at $t \in [a, b]$ if and only if f_i has a derivative at t for each $i = 1, \ldots, n$, and then

(12) $$F'(t) = \Sigma_{i=1}^{n} f_i'(t)B_i \quad .$$

Proof. The constant curve defined by $G(t) = B_i$ is easily seen to have derivative $G'(t) = \tilde{0}$. If F has a derivative at t, then Theorem 2.5 (iii), applied to (11), implies that f_i has a derivative in the orthonormal case. Conversely, if each f_i has a derivative, Theorem 2.5 (ii) implies that f_iB_i has a derivative, and then Theorem 2.5 (i), applied to (10), shows that $F'(t)$ exists and is given by (12).

2.8. Proposition. If $F'(t) = \tilde{0}$ for all $t \in [a, b]$, then F is constant over $[a, b]$.

Proof. By Proposition 2.7, $F' = \tilde{0}$ implies $f_i' = 0$ for each i. It follows that each f_i is constant, and hence also F.

§3. Arclength

Recall that the length of a curve is defined to be the limit of the lengths of approximating polygons (or broken lines) obtained from partitions of the interval $[a, b]$, as the length of the longest subinterval tends to zero, provided this limit exists. Since the length of the straight line joining $F(t)$ and $F(t + h)$ is $|F(t + h) - F(t)|$, and since we assume the standard theorems of the calculus for real-valued functions of a single real variable, we have the following result: if F is piecewise smooth, then the length L of F exists and is given by an integral:

$$L = \int_a^b |F'(t)|\, dt \ .$$

<u>Remark</u>. If W is finite dimensional, and if $B_1, \ldots, B_n$ is an orthonormal basis for W, then

$$F(t) = \Sigma_{i=1}^n f_i(t)B_i, \qquad f_i(t) = F(t)\cdot B_i \ ,$$

and the length of F can be expressed as

$$L = \int_a^b \sqrt{\Sigma_{i=1}^n (f_i'(t))^2}\ dt \ ,$$

using (12) of §2 and Theorem III, 5.9.

For each $t \in [a, b]$, set

(1) $$g(t) = \int_a^t |F'(\tau)|d\tau \ .$$

Then $s = g(t)$ is the arclength function, and

(2) $$\frac{ds}{dt} = g'(t) = |F'(t)| \ .$$

Thus the <u>velocity along the curve</u> (i.e. the rate of change of arclength) is the scalar velocity $|F'(t)|$.

We assume in the remainder of this section that F is smooth, and that $F'(t) \neq \tilde{0}$ for each $t \in [a, b]$. It follows from (1) that s is a strictly increasing function of t. Therefore $s = g(t)$ has an inverse function $t = h(s)$:

(3) $$h(g(t)) = t \ , \qquad g(h(s)) = s \ .$$

Since g' is never zero by (2), the derivative h' exists and is given by

(4) $$h'(s) = \frac{dt}{ds} = 1/\frac{ds}{dt} = 1/|F'(h(s))| \ .$$

Now let $G(s) = F(h(s))$. This function is defined for $0 \leq s \leq c$,

where $c = g(b)$, and gives the parametrization of the curve F by its arclength.

Let

(5) $$X = G(s) = F(h(s)) \quad .$$

The "function of a function" rule for differentiation (Exercise 9.1) asserts that G' exists, and that

(6) $$\frac{dX}{ds} = G'(s) = F'(h(s))h'(s) = F'(h(s))/|F'(h(s))| \quad .$$

Therefore G is a smooth curve, and

(7) $$|G'(s)| = 1 \qquad \text{for } s \in [0, c] \quad .$$

For this reason, $G'(s)$ is called the <u>unit tangent vector</u>.

If, in (5), we set $s = g(t)$, we obtain by (3) that

(8) $$X = G(g(t)) = F(t) \quad ,$$

while (6) becomes

(9) $$\frac{dX}{dt} = \frac{ds}{dt}\frac{dX}{ds} \quad .$$

This formula gives the resolution of the vector velocity as the unit tangent vector multiplied by the scalar velocity along the curve.

<u>Remark</u>. The hypothesis $F'(t) \neq \tilde{0}$ was used to ensure that dX/ds exists. The following example illustrates the need for some such restriction. Let F be the curve in the plane given by

$$F(t) = (t^3 \cos 1/t)i + (t^3 \sin 1/t)j, \qquad t \neq 0 \quad .$$

As t tends toward 0, the curve converges to $\tilde{0}$ in a spiral

which goes around $\tilde{0}$ infinitely many times. If we define $F(0) = \tilde{0}$, then F is a continuous curve defined for all t. Direct computation of the component derivatives shows that F' exists and is continuous for all t. However $F'(0) = \tilde{0}$, and it is clear that $\tilde{0}$ is a "bad point" of the curve in that there can be no tangent line to the curve at $\tilde{0}$, because the chord through $\tilde{0}$ and $F(t)$ rotates about $\tilde{0}$ infinitely many times as t tends to zero. Hence there is no unit tangent vector at $\tilde{0}$.

§4. Acceleration

Assume now that F is a curve such that the second derivative $F''(t)$ exists for each $t \in [a, b]$. The vector $F''(t)$ is called the acceleration vector, and is usually pictured as a vector whose initial point is $F(t)$. In general it is not tangent to the curve. We proceed to split the acceleration vector into the sum of a vector parallel to the tangent, called the <u>tangential acceleration</u>, and a vector perpendicular to the tangent, called the <u>centripetal acceleration</u>.

To do this we must assume, as in §3, that $F'(t) \neq \tilde{0}$ for each $t \in [a, b]$. Starting from (2) of §3, and applying Proposition 2.6 with F' in place of F, we obtain

$$\frac{d^2s}{dt^2} = F' \cdot F''/|F'| \quad . \tag{1}$$

This scalar function of t is called the <u>acceleration along the curve</u>. Since d^2s/dt^2 exists, the inverse function $t = h(s)$ has a second derivative given by

$$h''(s) = \frac{d^2t}{ds^2} = -\frac{d^2s}{dt^2} \Big/ \left(\frac{ds}{dt}\right)^3 .$$

From formula (6) of §3 we then conclude that d^2X/ds^2 exists and is given by

(2) $$\frac{d^2X}{ds^2} = F''(h(s))(h'(s))^2 + F'(h(s))h''(s) ,$$

while (9) of §3 gives

(3) $$\frac{d^2X}{dt^2} = \frac{d^2s}{dt^2}\frac{dX}{ds} + \left(\frac{ds}{dt}\right)^2 \frac{d^2X}{ds^2} .$$

This is the decomposition sought: the first term is the unit tangent vector multiplied by the scalar acceleration along the curve, and the second term is perpendicular to the tangent. To see this, recall that dX/ds has length 1:

$$\frac{dX}{ds} \cdot \frac{dX}{ds} = 1 .$$

Taking the derivative with respect to s on both sides, we obtain

$$\frac{d^2X}{ds^2} \cdot \frac{dX}{ds} + \frac{dX}{ds} \cdot \frac{d^2X}{ds^2} = 0$$

or, since the scalar product is commutative,

(4) $$\frac{d^2X}{ds^2} \cdot \frac{dX}{ds} = 0 .$$

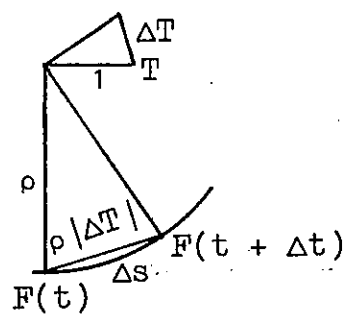
Hence the two vectors are orthogonal.

<u>Remark</u>. It can happen that d^2X/ds^2 is $\tilde{0}$. For example, if the motion takes place along a straight line, both F' and F'' are parallel to the line and so, by (2), is d^2X/ds^2. But if d^2X/ds^2 is both parallel and orthogonal to F', it must

be $\tilde{0}$.

The length of d^2X/ds^2 is called the <u>curvature</u> at X and is denoted by κ. If $\kappa \neq 0$, its reciprocal $\rho = 1/\kappa$ is called the <u>radius of curvature</u> at X.

A partial justification of these terms is obtained by considering a motion $F(t)$ which takes place in a euclidean space on a circle of radius ρ. Let $T = dX/ds$ be the unit tangent at a time t, and let ΔT be its increment corresponding to an increment Δt. From the following diagram we conclude,



by similar triangles, that the length of the chord is $\rho|\Delta T|$. Since the limit of the ratio of chord to arc is 1, it follows that

$$\left|\frac{d^2X}{ds^2}\right| = \lim_{\Delta s \longrightarrow 0} \left|\frac{\Delta T}{\Delta s}\right| = \frac{1}{\rho} \quad .$$

Thus, for circular motion, the radius of curvature is indeed the radius. The length of the centripetal acceleration vector in (3) is then

$$\kappa\left(\frac{ds}{dt}\right)^2 = \frac{1}{\rho}\left(\frac{ds}{dt}\right)^2 \quad ,$$

which agrees with the value v^2/ρ of elementary physics.

§5. Steady flows

5.1. Definition. A subset $D \subset V$ is called an open set in V if, for each vector $X \in D$, there is a positive distance r, depending on X, such that $|Y - X| < r$ implies $Y \in D$.

5.2. Definition. Let $D \subset V$ be open, let $F : D \longrightarrow V$ be continuous, and consider the differential equation

$$\frac{dY}{dt} = F(Y) \qquad \text{for } Y \in D . \tag{1}$$

A differentiable curve $Y = G(t)$ in D is called a solution of (1) if

$$\frac{d}{dt} G(t) = F(G(t)) \quad .$$

We denote by $G(X, t)$ a solution defined in an interval about $t = 0$ and satisfying the initial condition

$$G(X, 0) = X \quad , \tag{2}$$

where X is a point of D. Then

$$\frac{d}{dt} G(X, t) = F(G(X, t)) \quad , \tag{3}$$

and $Y = G(X, t)$ gives a curve which passes through X for $t = 0$ and whose derivative at a point Y on the curve is given by $F(Y)$.

5.3. Existence and uniqueness of solutions. Let V be finite dimensional, and suppose that $B_1, \ldots, B_n$ give a basis for V. If $Y = \Sigma_{i=1}^{n} y_i B_i$, then the component functions of (1) give a system of simultaneous differential equations

(4) $$\frac{dy_i}{dt} = f_i(Y) = f_i(y_1, \ldots, y_n), \qquad i = 1, \ldots, n ,$$

where we have used the same symbol f_i to denote the function of n variables $y_1, \ldots, y_n$ obtained from $f_i(Y)$. A standard existence theorem states that, if the functions f_i satisfy certain conditions (in particular, if they have partial derivatives throughout D), then there exists a unique solution

(5) $$y_i = g_i(x_1, \ldots, x_n, t), \qquad i = 1, \ldots, n ,$$

of (4), valid for sufficiently small $|t|$ (depending on $X = \Sigma_{i=1}^n x_i B_i$), and satisfying

(6) $$g_i(x_1, \ldots, x_n, 0) = x_i , \qquad i = 1, \ldots, n .$$

Then

$$G(X, t) = \Sigma_{i=1}^n g_i(X, t)B_i$$

gives a unique solution of (1) satisying the initial conditions (2).

Now let $D = V$, and suppose $G(X, t)$ satisfies (1) and (2) for $X \in V$ and $t \in R$. For each fixed t, these solutions define a transformation

$$G_t : V \longrightarrow V$$

by $X \longrightarrow G(X, t)$. If F satisfies the conditions to ensure the uniqueness of a solution of (1) and (2), then

(7) $$G_t G_s = G_{s+t} .$$

In fact, if s is a fixed real number, then the function $G(X, s + t)$ of t is easily seen to be a solution of (1) which reduces to $G(X, s)$ when $t = 0$. Therefore, by uniqueness,

(8) $$G(G(X, s), t) = G(X, s + t) \quad .$$

5.4. Proposition. The set of transformations G_t, $t \in R$, forms a (1-parameter) group of transformations of V into itself.

Proof. We must verify that the composition (7) satisfies the Axioms G1-G3 stated in Theorem II, 9.2. Axiom G1 (associativity) follows from Theorem II, 1.5. For Axiom G2, we note that the transformation G_o is I_V, by (2). For Axiom G3, we use (7) to verify that

$$G_t G_{-t} = G_{-t} G_t = G_o = I_V \quad ,$$

so G_t has an inverse G_{-t}.

Remarks. Since G_t has an inverse, the function $G_t : V \longrightarrow V$ is bijective, for each $t \in R$, by Exercise II, 3.2. By (7), $G_t G_s = G_s G_t$, so the group is called commutative (or abelian).

5.5. Steady flow. If $G(X, t)$ is thought of as giving a motion of the initial point X in V as t varies, then the relation (8) can be interpreted as follows. If C is the curve followed by a point X under $G(X, t)$, and if $Z = G(X, s)$ is a point of C, then the curve followed by Z is likewise C, i.e. G slides C along itself. There is one such curve passing through each point, namely, $G(X, t)$ passes through X when $t = 0$. If two such curves have a point in common, uniqueness implies that they coincide throughout. Thus the space V is filled with these non-intersecting curves, called the streamlines of the flow. The equation (1) asserts that the velocity

vector with which a point X passes through a point Y at a time t does not depend on t, but only on the position Y. For this reason the flow is called <u>steady</u>. In particular, the streamlines do not change as t varies.

<u>Example</u>. The simplest example is obtained by taking F in (1) to be a constant, say A. Then

$$G(X, t) = X + tA \quad .$$

For a fixed t, the function $G_t : V \longrightarrow V$ is the translation of V by the vector tA. The streamlines are the lines parallel to A.

§6. <u>Linear differential equations</u>

If V is finite dimensional and $T : V \longrightarrow V$ is a linear transformation, that is, $T \in E(V)$, the differential equation

$$\frac{dY}{dt} = TY \tag{1}$$

is said to be <u>linear and homogeneous with constant coefficients</u>.

Clearly, if we pass to the corresponding system (4) of §5, we obtain a system of simultaneous linear differential equations with constant coefficients. In this case the conditions for uniqueness are surely satisfied, and the solutions $G(X, t)$ exist for all values of t.

6.1. <u>Theorem</u>. In the case of the linear differential equation (1), the bijective transformations G_t are linear in X. Therefore

$$G(X, t) = S(t)X \quad , \tag{2}$$

where $S(t)$ is a differentiable curve in the space $A(V)$ of automorphisms of V. Further, $S(t)$ defines a 1-parameter subgroup of the group of automorphisms, since

$$S(t + t') = S(t)S(t') \quad . \tag{3}$$

<u>Proof</u>. Let $X, X' \in V$. Then

$$\frac{d}{dt}(G(X, t) + G(X', t))$$

$$= \frac{d}{dt} G(X, t) + \frac{d}{dt} G(X', t)$$

$$= TG(X, t) + TG(X', t) = T(G(X, t) + G(X', t)) \quad .$$

This shows that $G(X, t) + G(X', t)$ is a solution of (1). Since it reduces to $X + X'$ when $t = 0$, the uniqueness of the solution gives

$$G(X, t) + G(X', t) = G(X + X', t) \quad .$$

If $a \in R$, then

$$\frac{d}{dt} aG(X, t) = a \frac{d}{dt} G(X, t) = aTG(X, t)$$

$$= T(aG(X, t)) \quad .$$

This shows that $aG(X, t)$ is a solution of (1). Since it reduces to aX when $t = 0$, uniqueness gives

$$aG(X, t) = G(aX, t) \quad .$$

This proves the linearity of G_t in X, and leads to (2). The group property (3) is a restatement of formula (7) of §5, by means of (2).

6.2. <u>Theorem</u>. Let T in (1) be skew-symmetric, and let $S(t)$ be defined by (2). Then, for each t, $S(t)$ is an

orthogonal transformation of V.

Proof. We apply the formula of Theorem 2.5 (iii) for differentiating a scalar product, and then the definition of skew-symmetry (V, 5.6), to obtain

$$\frac{d}{dt}\,(G(X,\ t)\cdot G(X',\ t)) = TG(X,\ t)\cdot G(X',\ t) + G(X,\ t)\cdot TG(X',\ t)$$
$$= -\,G(X,\ t)\cdot TG(X',\ t) + G(X,\ t)\cdot TG(X',\ t) = 0.$$

Therefore $G(X,\ t)\cdot G(X',\ t)$ is independent of t, and for $t = 0$ is $X\cdot X'$. Hence

$$S(t)X\cdot S(t)X' = G(X,\ t)\cdot G(X',\ t) = X\cdot X' \quad . \tag{4}$$

This proves the theorem.

6.3. Theorem. Let $\dim V = 3$ and let $\Omega \in V$ with $\Omega \neq \tilde{\tilde{0}}$. The motion determined by the differential equation

$$\frac{dY}{dt} = \Omega \times Y \tag{5}$$

is rotation about the axis $L(\Omega)$ with the constant angular velocity $|\Omega|$, and $S(t)$ is the group of rotations about $L(\Omega)$ (with $S(t)$ corresponding to rotation through the angle $\pm\, t|\Omega|$).

Proof. Since $\Omega \times \Omega = \tilde{\tilde{0}}$, the velocity vector $\Omega \times Y$ is $\tilde{\tilde{0}}$ at $Y = a\Omega$ for each $a \in R$. It follows that the constant curves $Y = a\Omega$ give solutions of (5) and that each vector of $L(\Omega)$ remains fixed during the motion.

At a vector Y not in $L(\Omega)$, the velocity vector $\Omega \times Y$ is perpendicular to Ω. Thus the streamlines lie in planes perpendicular to Ω. Since $TY = \Omega \times Y$ is skew-symmetric, the motion preserves distance, by (4), and, in particular, distance

from the fixed vectors of $L(\Omega)$. Therefore the motion takes place along circles with centers in $L(\Omega)$.

If θ is the angle between Y and Ω, the distance between Y and the axis $L(\Omega)$ is $|Y| \sin\theta$, which is therefore the radius of the streamline through Y. The scalar velocity at Y is $|\Omega \times Y| = |\Omega|\ |Y| \sin\theta$. Thus the angular velocity about $L(\Omega)$ is $|\Omega|$. Since this value is independent of Y and t, the motion determined by (5) is a rotation about $L(\Omega)$.

The direction of the rotation is determined by the choice of $\times$.

6.4. <u>Theorem</u>. Let T in (1) be symmetric, and let $B_1, \ldots, B_n$ be an orthonormal set of proper vectors of T with proper values $\lambda_1, \ldots, \lambda_n$ respectively, where $n = \dim V$ (see Theorem V, 7.2). Let $S(t)$ be defined by (2). Then, for each t, $S(t)$ is a symmetric transformation of V for which $B_1, \ldots, B_n$ are proper vectors with proper values $e^{\lambda_1 t}, \ldots, e^{\lambda_n t}$ respectively.

<u>Proof</u>. With respect to the basis $B_1, \ldots, B_n$, the system of differential equations of the components is

$$\frac{dy_i}{dt} = \lambda_i y_i \ , \qquad\qquad i = 1, \ldots, n \ .$$

Evidently the solution is $y_i = x_i e^{\lambda_i t}$ for each i. Thus the matrix of $S(t)$ with respect to the basis $B_1, \ldots, B_n$ is the diagonal matrix with $e^{\lambda_1 t}, \ldots, e^{\lambda_n t}$ as diagonal entries.

6.5. <u>Corollary</u>. Let T in (1) be symmetric and let $S(t)$ be defined by (2). Then

$$\det S(t) = \exp(t \text{ trace } T) \quad ,$$

and

$$\frac{d}{dt} \det S(t)\Big|_{t=0} = \text{trace } T \quad .$$

<u>Proof</u>. The first relation follows from the fact that the determinant of a diagonal matrix is the product of its diagonal elements, so that $\det S(t)$ is given by

$$e^{\lambda_1 t} e^{\lambda_2 t} \cdots e^{\lambda_n t} = e^{(\lambda_1+\lambda_2+\cdots+\lambda_n)t} \quad ,$$

and from the definition of the trace of T (V, 4.6). The second relation is obtained by differentiating the first and setting $t = 0$.

Thus, when T is symmetric, the resulting flow has the following properties. The origin $\tilde{\tilde{0}}$ is fixed. The linear subspaces $L(B_1), \ldots, L(B_n)$ are streamlines. Other streamlines are generally curved (unless some of the λ_i are equal).

For example, if $n = 2$ and $\lambda_1 = 1$, $\lambda_2 = 2$, then the solution is given by

$$y_1 = x_1 e^t, \qquad y_2 = x_2 e^{2t} \quad .$$

From this it follows that

$$\frac{y_1^{\,2}}{y_2} = \frac{x_1^{\,2}}{x_2} \quad .$$

Thus if C is a constant and $(x_1, x_2) \neq \tilde{\tilde{0}}$ is on the parabola $x_2 = Cx_1^{\,2}$, then its position at any other time t is on the same parabola. The streamlines are obtained by varying C. Note that each parabola, with $\tilde{\tilde{0}}$ omitted, gives two streamlines, and

that the (unique) streamline through $\tilde{\tilde{0}}$ is the constant curve $\tilde{\tilde{0}}$.

As another example, take $n = 2$ and $\lambda_1 = 1$, $\lambda_2 = -1$. The solution is

$$y_1 = x_1 e^t, \qquad y_2 = x_2 e^{-t} \quad .$$

So $y_1 y_2 = x_1 x_2$. The streamlines are the origin $\tilde{\tilde{0}}$, the four semi-axes, and the hyperbolas obtained by setting $x_1 x_2$ equal to a non-zero constant.

Having analysed so successfully the automorphisms $S(t)$ corresponding to the solutions in the symmetric and skew-symmetric cases, it would be reasonable to expect an analysis of any linear T, because

$$T = T_1 + T_2$$

is the sum of its symmetric and skew-symmetric parts (Proposition V, 5.7). This is not so easy as it might appear at first glance. If $S_1(t)$ and $S_2(t)$ give the solutions for T_1 and T_2 respectively, one would hope (in view of Corollary 6.5) that the composition $S_1(t)S_2(t)$ might give the solution for $T_1 + T_2$; but this is not always the case. Note that the fact that (2) is a solution of (1) is equivalent to $dS/dt = TS$. Thus, if $Y = S_1(t)S_2(t)X$, then

$$\begin{aligned} \frac{dY}{dt} &= T_1 S_1(t) S_2(t) X + S_1(t) T_2 S_2(t) X \\ &= [T_1 + S_1(t) T_2 S_1(t)^{-1}] Y \quad . \end{aligned}$$

If $S_1(t)$ and T_2 commute, the second term reduces to T_2, and

then Y is indeed the solution for $T_1 + T_2$. But, in general, $S_1(t)$ and T_2 do not commute. However when t is small, $S_1(t)$ is very near $S_1(0) = I$ which commutes with all endomorphisms. Thus, for small values of t, $S_1(t)$ very nearly commutes with T_2; and $S_1(t)S_2(t)X$ is an approximate solution for small values of t. This gives the so-called <u>infinitesimal analysis</u> of the solution of $Y' = TY$: for small values of t, the solution $S(t)$ is approximately the composition of the rotation $S_2(t)$ followed by the symmetric transformation $S_1(t)$.

If $A \in V$ is a fixed vector, and T is linear, we have an inhomogeneous equation

$$\frac{dY}{dt} = A + TY \quad . \tag{6}$$

If $A = TB$ for some $B \in V$, which is certainly true if T^{-1} exists, then the solution of (6) expressed in terms of the solution $S(t)X$ of the corresponding homogeneous equation (1), is

$$Y = S(t)X + S(t)B - B \quad , \tag{7}$$

as can be seen by direct substitution. Thus, for each t, the solution (6) differs from that of (1) by $S(t)B - B$, i.e. by a translation which tends to zero with t. Thus the infinitesimal analysis in the inhomogeneous case is obtained from that of the homogeneous case by adjoining a small translation.

The results which have been obtained in this section for linear equations can be used to derive the infinitesimal analysis of the solution of the equation

$$\frac{dY}{dt} = F(Y) \tag{8}$$

of a general steady flow discussed in §5. To do this we must anticipate a result to be proved in Chapter VIII, namely, that if F is suitably differentiable in the open set D, and $X_0 \in D$, then, for Y sufficiently close to X_0, the function F has a Taylor's expansion with remainder:

$$F(Y) = F(X_0) + F'(X_0)(Y - X_0) + \varepsilon|Y - X_0| \quad .$$

In this expression $F'(X_0)$ is a linear transformation T, and ε tends to 0 as Y tends to X_0. With $A = F(X_0)$, (8) takes the form

$$\frac{d}{dt}(Y - X_0) = A + T(Y - X_0) + \varepsilon|Y - X_0| \quad . \tag{9}$$

If we omit the remainder term, (9) takes the form of (6), and the solution has the form

$$Y - X_0 = S(t)(X - X_0) + S(t)B - B \quad . \tag{10}$$

If we consider initial points X near X_0 and small values of t, then (10) gives an approximation to the solution of (9) which improves as X approaches X_0 and t approaches 0. It follows that the infinitesimal analysis of a linear flow applies to the more general case. If we picture a small spherical ball centered at X_0, then its motion under the flow for a short period of time can be described as a small rotation about X_0, followed by a symmetric transformation with origin X_0 which distorts the ball slightly, followed finally by a translation which tends to zero with t.

§7. General differential equations

A differential equation of order κ has the form

$$\frac{d^{\kappa}Y}{dt^{\kappa}} = F(t, Y, \frac{dY}{dt}, \ldots, \frac{d^{\kappa-1}Y}{dt^{\kappa-1}}) \quad . \tag{1}$$

It is understood here that Y and its derivatives are vectors in an n-dimensional space V, and F is a function of the scalar t and the indicated κ vectors, and the values of F are in V.

The equation (1) can be given the form of a steady flow

$$\frac{dZ}{dt} = H(Z) \tag{2}$$

in a vector space W of dimension $n\kappa + 1$ as follows. Choose a basis in V and in W. Then a vector Z in W determines a scalar s and κ n-dimensional vectors $Y_o, Y_1, \ldots, Y_{\kappa-1}$ by taking s to be the first component of Z, then Y_o to be the vector in V whose components are the 2nd to the $(n+1)^{st}$ of Z, then Y_1 to be the vector whose components are the $(n+2)^{nd}$ to the $(2n+1)^{st}$ of Z, etc. A scalar and κ vectors of V determine a vector of W by reversing the process. So we may write

$$Z = (s, Y_o, Y_1, \ldots, Y_{\kappa-1}) \quad ,$$

and treat s and the Y's as components of Z.

Now consider the system of simultaneous first order differential equations

$$\begin{aligned} &\frac{ds}{dt} = 1, \frac{dY_o}{dt} = Y_1, \frac{dY_1}{dt} = Y_2, \ldots, \frac{dY_{\kappa-2}}{dt} = Y_{\kappa-1} \quad , \\ &\frac{dY_{\kappa-1}}{dt} = F(s, Y_o, Y_1, \ldots, Y_{\kappa-1}) \quad . \end{aligned} \tag{3}$$

If we define $H(Z)$ in terms of components by

$$H(Z) = (1, Y_1, Y_2, \ldots, Y_{\kappa-1}, F(s, Y_o, \ldots, Y_{\kappa-1})) \quad ,$$

then the system (3) is equivalent to the single equation (2) because

$$\frac{dZ}{dt} = (\frac{ds}{dt}, \frac{dY_o}{dt}, \ldots, \frac{dY_{\kappa-1}}{dt}) \quad .$$

But in (3), we may eliminate variables by setting $s = t$, $Y_o = Y$, and $Y_i = d^iY/dt^i$ for $i = 1, \ldots, \kappa - 1$. The result is the single equation (1).

This reduction of the general equation to a steady flow is theoretically important. For example, the existence theorem for steady flows implies an existence theorem for the general equation. However, from the practical point of view of studying the properties of solutions, it is not very useful. This is mainly because $n\kappa + 1$ is much greater than n, and we want a picture of what happens in n-dimensions.

§8. Planetary motion

This section outlines the vectorial method of solving the classical "2 body problem". The filling-in of details is left to the student as an exercise.

Two points in R^3 with position vectors Y_1, Y_2 and masses m_1, m_2 are moving subject to Newton's laws of attraction. These laws give the pair of simultaneous differential equations

$$(1)\qquad \begin{aligned} m_1 \frac{d^2Y_1}{dt^2} &= Gm_1m_2 \frac{Y_2 - Y_1}{|Y_2-Y_1|^3} , \\ m_2 \frac{d^2Y_2}{dt^2} &= Gm_1m_2 \frac{Y_1 - Y_2}{|Y_2-Y_1|^3} , \end{aligned}$$

where G is the gravitational constant.

Adding the two equations gives an equation which integrates immediately into

$$(2)\qquad m_1Y_1 + m_2Y_2 = At + B ,$$

where the vectors A, B are constants of integration. The vector

$$(3)\qquad Z = \frac{m_1Y_1 + m_2Y_2}{m_1 + m_2}$$

lies on the line joining the two points, and is called their <u>center of gravity</u>. Then (2) asserts that Z moves in a straight line with constant velocity.

We now seek to determine the motion of each point relative to Z. Setting

$$Y = Y_1 - Z = \frac{m_2}{m_1 + m_2} (Y_1 - Y_2) ,$$

we have

$$\frac{d^2Y}{dt^2} = \frac{d^2Y_1}{dt^2} .$$

Then the first equation of (1) gives

$$(4)\qquad \frac{d^2Y}{dt^2} = - \kappa^2 \frac{Y}{|Y|^3} , \qquad \kappa^2 = \frac{Gm_2{}^3}{(m_1+m_2)^2}$$

The second equation of (1) gives the same equation except for a different value of the constant κ.

The integration of (4) is done in two stages. For the first, form the vector product of both sides of (4) with Y, and integrate to obtain

$$Y \times \frac{dY}{dt} = C \tag{5}$$

where the vector C is a constant of integration. This gives $Y \cdot C = 0$, which implies that the curve of motion lies in the plane through Y perpendicular to C. If we introduce polar coordinates in this plane:

$$Y = (r \cos \theta)i + (r \sin \theta)j \quad , \tag{6}$$

then (5) yields the equation

$$\frac{1}{2} r^2 \frac{d\theta}{dt} = \frac{1}{2} |C| \quad . \tag{7}$$

Since $\frac{1}{2} r^2 d\theta$ is the element of area, (7) gives Kepler's second law of planetary motion: <u>the radius vector from sun to planet sweeps out equal areas in equal times</u>.

For the second stage of the integration, we obtain from (4) and (5) that

$$\frac{d^2Y}{dt^2} \times C = - \frac{\kappa^2}{|Y|^3} Y \times (Y \times \frac{dY}{dt}) \quad . \tag{8}$$

We expand the right side (see Proposition IV, 4.2), using $r = |Y|$, $Y \cdot Y = r^2$, from which follows

$$Y \cdot \frac{dY}{dt} = r \frac{dr}{dt} \quad .$$

Then (8) becomes

$$\frac{d^2Y}{dt^2} \times C = \kappa^2 \frac{d}{dt} \frac{Y}{r} \quad ,$$

which integrates into

$$\frac{dY}{dt} \times C = \kappa^2 \left(\frac{Y}{r} + E\right) \quad . \tag{9}$$

The constant E of integration is perpendicular to C. Eliminating dY/dt between (9) and (5) by forming the triple product $[Y, dY/dt, C]$ in two ways, we obtain

$$\kappa^2(r + E\cdot Y) = C\cdot C \quad . \tag{10}$$

To interpret this equation, choose the basis i, j in the plane of motion so that $E = ei$ with $e > 0$. Then (10) gives

$$r = \frac{c^2/\kappa^2}{1 + e \cos \theta} \quad , \qquad c = |C| \quad . \tag{11}$$

This is the polar equation of a conic with focus at the origin, having eccentricity e, and with directrix parallel to j at a distance d from the origin where

$$ed = c^2/\kappa^2 \quad .$$

For $e < 1$, this is Kepler's first law: <u>the planet moves in an ellipse with the sun at a focus</u>. (Actually it is the center of gravity of sun and planet which is at the focus.) Parabolic and hyperbolic orbits $(e \geq 1)$ are followed by bodies which enter the solar system with a high enough velocity to escape from the sun's influence.

In the case of elliptical motion, the semi-major axis,

denoted by a, is

$$a = \frac{ed}{1 - e^2} ,$$

and the semi-minor axis is $b = a\sqrt{1 - e^2}$. Dividing the area πab of the ellipse by the sectorial velocity (7) gives the <u>period</u>

$$p = \frac{2\pi}{\kappa} a^{3/2} .$$

This gives Kepler's third law: <u>the square of the period of a planet is proportional to the cube of the mean distance</u>.

§9. Exercises

1. State and prove the "function of a function" rule for differentiation in the case needed in order to derive (6) of §3.

2. Show that the flow in R^3 defined by

$$\frac{dy_1}{dt} = - y_2 - 2y_3, \quad \frac{dy_2}{dt} = - 3y_3 + y_1, \quad \frac{dy_3}{dt} = 2y_1 + 3y_2$$

is a rotation about an axis at constant angular velocity. Find the components of a vector in the axis.

3. Discuss the flow in R^2 defined by

$$\frac{dy_1}{dt} = 6y_1 + 2y_2, \quad \frac{dy_2}{dt} = 2y_1 + 9y_2 .$$

Make a sketch showing the streamlines.

4. Let A, X, X' be fixed vectors. Find the integral curve of $d^2Y/dt^2 = A$ which passes through X with velocity X'

at the time $t = 0$. Show that the curve lies in a plane, and is a parabola.

5. Let X, X' be fixed vectors. Find the integral curve of $d^2Y/dt^2 = dY/dt$ which passes through X with velocity X' when $t = 0$. What is the nature of the curve?

6. Let X, X' be fixed vectors, and $a \in R$. Find the integral curve of $d^2Y/dt^2 = -a^2Y$ which passes through X with velocity X' when $t = 0$. Show that motion is periodic and that the curve is an ellipse.

7. Let $T: V \longrightarrow V$ be linear and such that $TTY = -a^2Y$ for every $Y \in V$. Find the integral curve of $dY/dt = TY$ which passes through X when $t = 0$. [Hint: Use a solution of Exercise 6 for a suitable choice of X'.]

8. Let $f: V \longrightarrow R$ be continuous. Show that each integral curve of $d^2Y/dt^2 = f(Y)Y$ lies in some plane.

$$g(t + h) - g(t) = f(X + tY + hY) - f(X + tY) \quad ,$$

so

$$g'(t) = f'(X + tY, Y) \tag{4}$$

exists for each t. By the mean value theorem for real-valued functions of a single real variable,

$$g(1) - g(0) = g'(\theta) \tag{5}$$

for some $0 < \theta < 1$. Substituting in (5) from (3) and (4) gives the required conclusion.

1.5. <u>Proposition</u>. Let $X_o \in D$ and $Y, Z \in V$. Assume that $f'(X_o, Y)$ exists and that there is an $r > 0$ such that $f'(X, Z)$ exists for all X satisfying $|X - X_o| < r$. Assume moreover that $f'(X, Z)$ is continuous in X at X_o. Then $f'(X_o, Y + Z)$ exists, and

$$f'(X_o, Y + Z) = f'(X_o, Y) + f'(X_o, Z) \quad . \tag{6}$$

<u>Proof</u>. For h sufficiently small,

$$\begin{aligned} & f(X_o + hY + hZ) - f(X_o) \\ & = [f(X_o + hY + hZ) - f(X_o + hY)] + [f(X_o + hY) - f(X_o)] \quad . \end{aligned} \tag{7}$$

By Proposition 1.3, $f'(X, hZ)$ exists for all X satisfying $|X - X_o| < r$ and, in particular, for $X = X_o + hY + thZ$, $0 \leq t \leq 1$, if h is small enough that $|hY + hZ| < r$. Then by Propositions 1.4 and 1.3,

$$f(X_o + hY + hZ) - f(X_o + hY) = f'(X_o + hY + \theta hZ, hZ)$$
$$= hf'(X_o + hY + \theta hZ, Z) \quad ,$$

for some θ (depending on h) with $0 < \theta < 1$. Thus, if we divide by h in (7), the first term in brackets on the right side has the limit $f'(X_o, Z)$ as h tends to zero, by the hypothesis on the continuity of $f'(X, Z)$. The second term in brackets has the limit $f'(X_o, Y)$, so we conclude that $f'(X_o, Y + Z)$ exists and is given by (6).

By combining Propositions 1.3 and 1.5, we obtain

1.6. <u>Theorem</u>. If, for each $Y \in V$, $f'(X, Y)$ exists for each $X \in D$ and is continuous in X, i.e. if f is continuously differentiable in D, then, for each $X \in D$, $f'(X, Y)$ is a linear function of Y.

1.7. <u>Definition</u>. If $f: D \longrightarrow R$ is continuously differentiable in D, then the <u>derivative</u> $f'(X)$ at $X \in D$ is the linear transformation

$$f'(X): V \longrightarrow R$$

defined by $f'(X)Y = f'(X, Y)$.

1.8. <u>Proposition</u>. Let $A_1, \ldots, A_n$ be a basis in V. Setting $X = \Sigma_{i=1}^{n} x_i A_i$, then $f(X)$ becomes a function of n variables $f(x_1, \ldots, x_n)$. If $f'(X, A_i)$ exists, then

$$f'(X, A_i) = \frac{\partial}{\partial x_i} f(x_1, \ldots, x_n), \qquad i = 1, \ldots, n \quad .$$

Moreover, f is continuously differentiable in D if and only if the partial derivatives $f'(X, A_i)$ exist and are continuous in D, $i = 1, \ldots, n$, and then

$$(8) \qquad f'(X, Y) = \Sigma_{i=1}^{n} y_i f'(X, A_i) = \Sigma_{i=1}^{n} y_i \frac{\partial f}{\partial x_i} , \qquad Y = \Sigma_{i=1}^{n} y_i A_i .$$

Proof. The components of $X + hA_i$ are those of X except for the ith which is $x_i + h$. Thus the difference quotient used in defining $f'(X, A_i)$ is exactly the one used in defining the partial derivative.

If f is continuously differentiable, then (8) follows from Theorem 1.6.

Conversely, if the partial derivatives $f'(X, A_i)$ exist and are continuous in D, then the same is true for $f'(X, y_i A_i)$, $y_i \in R$, for any $i = 1, \ldots, n$, by Proposition 1.3 and Corollary VI, 1.4. To prove (8) for arbitrary $Y \in V$, we use Proposition 1.5 and induction on the number of basis elements A_i used in expressing Y in terms of the basis. In fact, if $f'(X, Y)$ exists and is given by (8), and therefore continuous in D, for all $Y \in L(A_1, \ldots, A_{k-1})$, then the same statements hold for $Y \in L(A_1, \ldots, A_k)$, since

$$\begin{aligned} f'(X, Y) &= f'(X, \Sigma_{i=1}^{k-1} y_i A_i + y_k A_k) \\ &= f'(X, \Sigma_{i=1}^{k-1} y_i A_i) + f'(X, y_k A_k) \\ &= \Sigma_{i=1}^{k-1} y_i f'(X, A_i) + y_k f'(X, A_k) . \end{aligned}$$

§2. Rate of change along a curve

2.1. Theorem. Let $f: D \longrightarrow R$ be continuously differentiable in D. Let $F: [a, b] \longrightarrow D$ be a smooth curve in D, and let

$$g(t) = f(F(t)) \qquad \text{for } t \in [a, b] .$$

Then $g: [a, b] \longrightarrow R$ is differentiable, and

$$(1) \qquad g'(t) = f'(F(t), F'(t)) .$$

<u>Proof</u>. For h sufficiently small and satisfying $-(t - a) \leq h \leq b - t$, Proposition 1.4 gives

$$\begin{aligned} g(t + h) - g(t) &= f(F(t + h)) - f(F(t)) \\ &= f'(F(t) + \theta(F(t + h) - F(t)), F(t + h) - F(t)) \end{aligned}$$

where $0 < \theta < 1$. By Proposition 1.3,

$$(2) \quad \frac{g(t+h) - g(t)}{h} = f'(F(t) + \theta(F(t + h) - F(t)), \frac{F(t+h) - F(t)}{h}) ,$$

$h \neq 0$. As h tends to zero, the right-hand side of (2) has the limit $f'(F(t), F'(t))$ since, by hypothesis, f is continuously differentiable, and F is smooth. Thus the left-hand side of (2) has a limit as h tends to zero, and (1) holds.

<u>Remarks</u>. The hypothesis that f is continuously differentiable does not immediately imply the existence of the limit of the right side of (2) if F is smooth. The hypothesis gives only that $f'(X, Y)$ is continuous in X for each fixed Y and therefore, by Theorem 1.6, linear (and continuous) in Y for each fixed X. A justification of this step will be given in §3.

From Theorem 2.1, it is clear that the derivative of f at X with respect to Y, according to Definition 1.1, is just the rate of change of f along the curve $X + tY$ at $t = 0$. If f is continuously differentiable, the same value is obtained for

the rate of change of f at X along any curve whose instantaneous velocity vector at X is Y.

2.2. Corollary. If $A_1, \ldots, A_n$ is a basis for V, and $X = F(t) = \Sigma_{i=1}^{n} x_i(t)A_i$, then (1) becomes

$$\frac{dg}{dt} = \Sigma_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{dx_i}{dt} \quad . \tag{3}$$

Proof. Evaluate $f'(F(t), F'(t))$ in (1) by means of (8) of §1 and (12) of Chapter VI, §2.

§3. Gradient; directional derivative

The definition of derivative, etc., can be shown to be independent of the particular choice of scalar product on V if V is finite dimensional (cf. Chapter X). We now consider some notions which depend on the choice of the scalar product.

3.1. Proposition. Let $f: D \longrightarrow R$ be continuously differentiable, where $D \subset V$ is open, and let V have a scalar product. Then, for each $X \in D$, there is a unique vector $\nabla f(X)$ such that

$$f'(X, Y) = \nabla f(X) \cdot Y \quad , \qquad Y \in V \quad . \tag{1}$$

Proof. The existence of $\nabla f(X)$ satisfying (1) follows from Theorem III, 5.8 applied to the linear function $f'(X, Y)$ of Y.

3.2. Definition. The vector $\nabla f(X)$ of Proposition 3.1 is called the gradient of f at X, denoted also by grad f.

Remarks. It is clear from (1) that knowing the gradient $\nabla f(X)$ of f at X is equivalent to knowing the derivative

$f'(X)$ of f at X. However, $f'(X) \in L(V, R)$ while $\nabla f(X) \in V$ and may be considered as a vector with initial point at X.

3.3. Proposition. If $f: D \longrightarrow R$ is continuously differentiable, then

$$\nabla f: D \longrightarrow V$$

is continuous.

Proof. Let $A_1, \ldots, A_n$ be an orthonormal basis for V. Then

$$\nabla f(X) = \Sigma_{i=1}^{n} (\nabla f)_i(X) A_i \tag{2}$$

where

$$(\nabla f)_i(X) = \nabla f(X) \cdot A_i = f'(X, A_i) \quad . \tag{3}$$

Since $f'(X, A_i)$ is continuous by hypothesis, $i = 1, \ldots, n$, the continuity of ∇f follows from Proposition VI, 1.6.

Examples. If $A_1, \ldots, A_n$ in Proposition 1.8 is an orthonormal basis, then

$$\nabla f = \Sigma_{i=1}^{n} \frac{\partial f}{\partial x_i} A_i \quad . \tag{4}$$

In terms of the gradient of f, formula (1) of §2 for the derivative of the function $g(t) = f(F(t))$, where $F(t)$ is a smooth curve in D, may be stated as

$$\frac{dg}{dt} = \nabla f \cdot \frac{dF}{dt} \quad . \tag{5}$$

To justify the limit in the proof of Theorem 2.1, note that the right-hand member of (2) of §2 may be written as

$$\nabla f(F(t) + \theta(F(t + h) - F(t)))\cdot\left(\frac{F(t+h) - F(t)}{h}\right)$$

we then apply Theorem VI, 1.3 (iii) to obtain the limit

$$\nabla f(F(t))\cdot F'(t) = f'(F(t), F'(t)) \quad .$$

3.4. Definition. If Y is a unit vector, i.e. $|Y| = 1$, then $f'(X, Y)$ is called the directional derivative of f in the direction Y.

Since $|sY| = |s|$, it is clear from the definition

$$f'(X, Y) = \lim_{s \to 0} \frac{1}{s}[f(X + sY) - f(X)]$$

that $f'(X, Y)$ is just the rate of change of f with respect to distance along the straight line through X in the direction Y. For this reason, the directional derivative is often denoted by df/ds.

3.5. Proposition. The directional derivative df/ds in the direction Y is given by

$$\frac{df}{ds} = |\nabla f| \cos \theta \quad , \tag{6}$$

where θ is the angle between the gradient vector and Y. It follows that f changes most rapidly in the direction of ∇f, and the maximum rate of change is $|\nabla f|$. In any other direction, the rate of change is the projection of ∇f on that direction.

Proof. By (1),

$$\frac{df}{ds} = \nabla f\cdot Y$$

where, by hypothesis, $|Y| = 1$.

§4. Level surfaces

In interpreting a function $f: D \longrightarrow R$ and its derivative or gradient when $D \subset R^n$, it is convenient to use the method of inverse images for visualizing the function (see Definition II, 1.2). For each $x \in R$, $f^{-1}(x)$ is a subset of D. If $x_1 \neq x_2$, then $f^{-1}(x_1) \cap f^{-1}(x_2)$ is empty. And if $X \in D$, then $X \in f^{-1}f(X)$. Thus, the collection of all inverse images fills D and gives a decomposition of D into pairwise disjoint sets.

Examples. Let $f: R^2 \longrightarrow R$ be defined by $f(x_1, x_2) = x_1 + x_2$. Then f is linear. The kernel, $f^{-1}(0)$, is the line $x_1 = -x_2$ and, for each $c \in R$, $f^{-1}(c)$ is the parallel line $x_1 + x_2 = c$. Thus, the inverse images decompose R^2 into a family of parallel lines.

Let $f: R^2 \longrightarrow R$ be $f(x_1, x_2) = x_1^2 + x_2^2$. Then $f^{-1}(c)$ is empty if $c < 0$, and it is a circle of radius c if $c \geq 0$.

Let $f: R^2 \longrightarrow R$ be $f(x_1, x_2) = x_1x_2$. Then $f^{-1}(0)$ is the union of the two axes and, if $c \neq 0$, $f^{-1}(c)$ is the hyperbola $x_1x_2 = c$ having the axes as asymptotes. Thus R^2 is decomposed into the family of all hyperbolas having the axes as asymptotes, together with the asymptotes.

4.1. Definition. If $f: D \longrightarrow R$ and $D \subset R^n$, then $f^{-1}(c)$, for any $c \in R$, will be called a level curve of f if $n = 2$, a level surface of f if $n = 3$, and a level hypersurface if $n > 3$ (but this is usually abbreviated to level surface).

Proof. If $X = F(t)$ is a smooth curve which lies on the level surface S of f, which passes through X_o when $t = t_o$, then

$$g(t) = f(F(t)) = c \qquad \text{for all } t \ .$$

Therefore $g'(t) = 0$ for all t, so (5) of §3 gives

$$\nabla f(X_o)\cdot F'(t_o) = 0 \ ;$$

that is, the tangent vector $F'(t_o)$ of the curve at X_o lies in the tangent plane to S at X_o and is orthogonal to $\nabla f(X_o)$.

4.4. Proposition. Let $f: D \longrightarrow R$ be continuously differentiable. A solution of the differential equation

$$\frac{dY}{dt} = \nabla f(Y) \tag{1}$$

(see Definition VI, 5.2) is orthogonal, at each of its points for which $\nabla f \neq \tilde{0}$, to the level surface of f through that point.

The proof is obvious. The solutions of (1) are called the orthogonal trajectories of the family of level surfaces of f.

§5. Exercises

1. If $f(X) = X\cdot X$, show that $\nabla f(X) = 2X$.

2. Let $T: V \longrightarrow V$ be linear, and let $f(X) = TX\cdot X$. Show that $\nabla f = T + T^*$ (see Definition V, 5.1).

3. Let T be an endomorphism of R^3, and let $f(X) = A\cdot X \times TX$ where A is a fixed vector. Show that

$$\nabla f(X) = T^*(A \times X) - A \times TX \ .$$

4. Show that the product rule holds: $\nabla(fg) = g\nabla f + f\nabla g$.

5. Let $h(X) = g(f(X))$, where $f: D \longrightarrow R$ and $g: R \longrightarrow R$ are differentiable. Show that h is differentiable and that

$$\nabla h(X) = g'(f(X))\nabla f(X) \quad .$$

In the following exercises we revert to the customary notation in R^3 where $X = xi + yj + zk$, and $f(X) = f(x, y, z)$.

6. Find a vector normal to the surface $x^3 + xy - z^2 = 2$ at the point $(1, 2, 1)$.

7. If $f_1(x, y) = g_1(x + y)$ and $f_2(x, y) = g_2(x - y)$ show that $\nabla f_1 \cdot \nabla f_2 = 0$.

8. If $f(x, y, z) = g(x^2 + y^2 + z^2)$, show that $X \times \nabla f = \tilde{0}$.

9. When, where, and at what angles does the helix

$$F(t) = (3t\cos\pi t)i + (3\sin\pi t)j + 4tk$$

intersect the sphere $x^2 + y^2 + z^2 = 25$?

10. Show that the curve

$$F(t) = t^2 i + tj + (\tfrac{1}{2}\log t)k \ , \qquad t > 0,$$

meets each level surface of $f = 2x^2 + y^2 + z$ at right angles.

§6. Reconstructing a function from its gradient

6.1. Definition. A function $F: D \longrightarrow V$, where D is an open set in the vector space V, is called a vector field (or field) in D.

Remarks. The reason for the term "level curve" is seen by considering the graph $z = f(x_1, x_2)$ of f with the z-axis pictured in the vertical direction and R^2 as the horizontal plane. A horizontal plane (parallel to R^2) is defined by setting $z = c$ for some $c \in R$. Such a plane cuts the graph (which is usually a surface) in a curve. The perpendicular projection of this curve onto R^2 is the level curve $f^{-1}(c)$. Thus $f^{-1}(c)$ is the projection in R^2 of the curve on the surface $z = f(x_1, x_2)$ at the level $z = c$. Level curves are sometimes called contour curves in analogy with the contour lines of relief maps (in which case f is the altitude).

For $n = 3$, the level surfaces are ordinary surfaces, in general. For example, the level surfaces of $f = x_1 + 3x_2 - x_3$ form a family of parallel planes; those of $f = x_1^2 + x_2^2 + x_3^2$ are all spheres with center at the origin.

If $f: R^n \longrightarrow R$ is linear, then $\dim f^{-1}(0)$ is $n - 1$ (if $\operatorname{im} f \neq 0$) and the level surfaces are parallel hyperplanes (usually abbreviated to planes).

4.2. Definition. Let $f: D \longrightarrow R$ be continuously differentiable. Let $X_0 \in D$, and suppose that $\nabla f(X_0) \neq \tilde{0}$. Take $c = f(X_0)$, and let S denote the level surface $f(X) = c$. Then the line through X_0 in the direction of $\nabla f(X_0)$ is called the normal to S at X_0; the plane through X_0 perpendicular to $\nabla f(X_0)$ is called the tangent plane to S at X_0.

Examples. The equation of the tangent plane (cf. Chapter IV, §5) is

$$\nabla f(X_o)\cdot(X - X_o) = 0 \quad .$$

In R^3, the equation of the normal line is

$$\nabla f(X_o) \times (X - X_o) = \tilde{0} \quad .$$

For the special case when f is linear, it follows from Theorem III, 5.8, that there is a vector A such that $f(X) = A\cdot X$. Then

$$f(X + hY) - f(X) = A\cdot(X + hY) - A\cdot X = hA\cdot Y \quad .$$

Dividing by h and taking the limit as h tends to zero, we obtain

$$f'(X, Y) = A\cdot Y \quad .$$

Hence $\nabla f(X) = A$ for all X, and the equation for the tangent plane becomes $A\cdot(X - X_o) = 0$, or $A\cdot X = A\cdot X_o$. But this is just the equation $f(X) = c$ for S, so the tangent plane coincides with S (as it should).

The need for the assumption $\nabla f(X_o) \neq \tilde{0}$ is evident in the example $f(X) = X\cdot X$. A simple calculation gives

$$f'(X, Y) = 2X\cdot Y \,, \qquad \nabla f(X) = 2X \quad .$$

For $X_o = \tilde{0}$, we have $\nabla f(\tilde{0}) = \tilde{0}$ while the level surface S through X_o is just the single point $\tilde{0}$, so there can be no tangent plane.

4.3. <u>Proposition</u>. Let $f: D \longrightarrow R$ be continuously differentiable, let $X_o \in D$, $\nabla f(X_o) \neq \tilde{0}$, and let S be the level surface of f through X_o. Then the gradient vector at X_o is orthogonal to the tangent vectors of all curves in S through X_o.

We have already considered vector fields in Chapter VI, §5, when we studied the differential equation

$$\frac{dY}{dt} = F(Y) \quad , \tag{1}$$

with $F: D \longrightarrow V$ a continuous vector field.

In §3, we have defined a continuous vector field $\nabla f: D \longrightarrow V$ corresponding to any continuously differentiable function $f: D \longrightarrow R$, $D \subset V$, and a given choice of scalar product.

We now consider the differential equation

$$\nabla f = F \quad , \tag{2}$$

where $F: D \longrightarrow V$ is given, and f is to be found.

In terms of a basis for V, the system (1) is equivalent to a system of ordinary differential equations (one independent variable) and a solution $Y = G(t)$ of (1) always exists, at least for sufficiently small $|t|$. However, if we consider the components of (2) relative to an orthonormal basis for V, we obtain a system of partial differential equations for f (if $\dim V > 1$), and solutions f exist only for certain choices of F (see §8); that is, not every vector field is a gradient.

As an illustration, we consider the problem of reconstructing a function in R^3 from its gradient field. This is a problem of integration. We adopt the customary notation $X = xi + yj + zk$, and $f(X) = f(x, y, z)$.

Let f be the function

(1) $$f(x, y, z) = 2x^2yz - xz^2 - 3xy^2 + y - 2z + 5 \quad .$$

The component functions of the gradient

$$\nabla f = \frac{\partial f}{\partial x}i + \frac{\partial f}{\partial y}j + \frac{\partial f}{\partial z}k$$

are

(2) $$\frac{\partial f}{\partial x} = 4xyz - z^2 - 3y^2 \quad ,$$

(3) $$\frac{\partial f}{\partial y} = 2x^2z - 6xy + 1 \quad ,$$

(4) $$\frac{\partial f}{\partial z} = 2x^2y - 2xz - 2 \quad .$$

Starting with these three functions the problem is to reconstruct f. Now (2) is obtained from f by treating y, z as constants and differentiating with respect to x. If we start with (2), and form the indefinite integral with respect to x (treating y, z as constants), we obtain

(5) $$f = 2x^2yz - xz^2 - 3xy^2 + C_1(y, z)$$

where C_1 (the constant of integration) may be any function of y, z. Treating (3) and (4) similarly, we have

(6) $$f = 2x^2yz - 3xy^2 + y + C_2(x, z) \quad ,$$

(7) $$f = 2x^2yz - xz^2 - 2z + C_3(x, y) \quad .$$

Comparing the three expressions of f in (5), (6) and (7), we

must write down an explicit function which agrees with all three. The term $2x^2yz$ appears in all three. The term $-xz^2$ of (5) appears explicitly in (7), and is a term of $C_2(x, z)$ of (6). The term $-3xy^2$, appearing explicitly in (5) and (6), is a term of $C_3(x, y)$ of (7). The term y of (6) may be regarded as a term of $C_1(y, z)$ and of $C_3(x, y)$. Similarly $-2z$ of (7) may be regarded as a term of $C_1(y, z)$ and of $C_2(y, z)$. Thus

$$f = 2x^2yz - xz^2 - 3xy^2 + y - 2z + C$$

agrees with all three forms of f, and differs from (1) only by an additive constant.

It should be emphasized that not every vector field is a gradient. As an example of a vector field in R^2 which is not a gradient, take $F = xi + xyj$. If there were a function f satisfying $\nabla f = F$, then

$$\frac{\partial f}{\partial x} = x \quad , \qquad\qquad \frac{\partial f}{\partial y} = xy \quad ,$$

and therefore

$$\frac{\partial^2 f}{\partial y \partial x} = 0 \quad , \qquad\qquad \frac{\partial^2 f}{\partial x \partial y} = y \quad .$$

But this contradicts the fact that the mixed derivatives of f in the two orders should be equal. If the integration method described above is applied to this example, we obtain

$$f = \frac{1}{2}x^2 + C_1(y) \quad , \qquad\qquad f = \frac{1}{2}xy^2 + C_2(x) \quad .$$

These two forms are incompatible because $\frac{1}{2}xy^2$ is not a term of the first form. It should be clear from this example that "most" vector fields are not gradients.

§7. Line integrals

7.1. Definition. Let D be an open subset of V, and let $F: D \longrightarrow V$ be a continuous vector field in D. Let C be a piecewise smooth curve $X = X(t)$ in D defined for $t \in [a, b]$. The integral of the field F along the curve C is denoted by $\int_C F \cdot dX$, and is defined by

$$(1) \qquad \int_C F \cdot dX = \int_a^b F(X(t)) \cdot \frac{dX}{dt}\, dt \quad .$$

Remarks. Because C is piecewise smooth, and F is continuous, the integrand on the right in (1) is a piecewise continuous function of t. Hence the integral exists. The integral is also called a line integral. Integrals more general than (1) will be considered in Chapter XII.

There are various notations for the line integral. For example, the expression

$$\int (x + 2yz)dx + x^2 dy - 3xydz$$

represents the line integral of the field

$$F = (x + 2yz)i + x^2 j - 3xyk$$

along an unspecified curve, where

$$dX = dx\, i + dy\, j + dz\, k \quad .$$

<u>Example</u>. In R^3, let

$$F = x^2 i + 2yzj - 3k, \qquad X(t) = ti - t^2 j + (1 - t)k ,$$

where $0 \leq t \leq 1$. Then

$$F(X(t)) = t^2 i - 2t^2(1 - t)j - 3k ,$$

$$\frac{dX}{dt} = i - 2tj - k ,$$

and

$$\int_C F\cdot dX = \int_0^1 (t^2 + 4t^3(1 - t) + 3)dt = 53/15 .$$

The following properties of the line integral are evident from its definition in terms of an ordinary integral.

7.2. <u>Proposition</u>. The line integral of a vector field along a given curve C is a linear function of the vector field:

(2) $$\int_C aF\cdot dX = a\int_C F\cdot dX , \qquad a \in R ,$$

(3) $$\int_C (F + G)\cdot dX = \int_C F\cdot dX + \int_C G\cdot dX .$$

Also if $X(t)$ is defined for $a \leq t \leq c$, and $a < b < c$, then the curve C from $X(a)$ to $X(c)$ is the union of the curves C' and C'' obtained by the restriction of t to $a \leq t \leq b$ and to $b \leq t \leq c$. In this case,

(4) $$\int_C F\cdot dX = \int_{C'} F\cdot dX + \int_{C''} F\cdot dX .$$

7.3. <u>Proposition</u>. The line integral is invariant under a change of parameter along the curve. In particular, the arclength s may be used as a parameter, and we have

$$\int_C F\cdot dX = \int_0^e F\cdot\frac{dX}{ds}\, ds$$

where e is the length of C.

Proof. Let $X(t)$ be a piecewise smooth curve C in D defined for $a \leq t \leq b$, and let $g\colon [c, d] \longrightarrow [a, b]$ be a piecewise smooth function $t = g(\tau)$ such that $g(c) = a$, $g(d) = b$ and $g'(\tau) > 0$. Then $X(g(\tau))$ is a piecewise smooth curve C' in D which coincides as a set of points with C, and is a reparametrization of C. Since

$$\frac{dx}{d\tau} = \frac{dX}{dt}\frac{dt}{d\tau} ,$$

it follows that

$$\int_c^d F(X(g(\tau)))\cdot\frac{dX}{d\tau}\, d\tau = \int_c^d [F(X(g(\tau)))\cdot\frac{dX}{dt}]\, \frac{dt}{d\tau}\, d\tau$$

$$= \int_a^b F(X(t))\cdot\frac{dX}{dt}\, dt \quad .$$

The last step follows from the standard rule for changing an integral by a substitution. Thus

$$\int_C F\cdot dX = \int_{C'} F\cdot dX \quad .$$

Remarks. A piecewise smooth function $g\colon [c, d] \longrightarrow [a, b]$ which satisfies $g(c) = b$ and $g(d) = a$, with $g'(\tau) < 0$, does not give a reparametrization of the curve C, but defines a curve which will be denoted by $-C$, since

(5) $$\int_{-C} F\cdot dX = -\int_C F\cdot dX \quad .$$

That is, a curve appearing in a line integral is oriented.

Reversing the order of the end points reverses the orientation of the curve and the sign of the integral.

In the case that the curve is parametrized by its arc length, dX/ds is the unit tangent vector (see Chapter VI, §3) and therefore $F \cdot dX/ds$ is just the component of F along the curve.

The standard physical interpretation of the line integral is that it represents the work done by a force field F on a particle moving along C (more precisely, its negative is the work done against F in moving along C). In fact, the elementary definition of work is effective force times the distance moved. The component of F along C is the effective force, so multiplying by the increment of distance ds and summing gives the total work.

§8. The fundamental theorem of calculus

8.1. Theorem. Let $f: D \longrightarrow R$ be continuously differentiable in D. Let X_0, X_1 be two points of D which can be connected by a curve in D. Then, for any piecewise smooth curve C in D, from X_0 to X_1,

$$\int_C \nabla f \cdot dX = f(X_1) - f(X_0) \quad . \tag{1}$$

Remark. It should be observed that (1) is a generalization of the fundamental theorem of the calculus. It becomes that theorem when $V = R$, because ∇f reduces to the ordinary derivative, and the scalar product is ordinary multiplication.

Proof. Let $X(t)$, $a \leq t \leq b$ be a curve C with

$X(a) = X_0$ and $X(b) = X_1$. Let $g(t) = f(X(t))$. By (5) of §3,

$$g' = \frac{dg}{dt} = \nabla f \cdot \frac{dX}{dt} \quad ,$$

so g is piecewise continuous. The fundamental theorem of calculus gives

$$\int_a^b g'(t)dt = g(b) - g(a) \quad .$$

Substituting for g and g', this becomes

$$\int_a^b \nabla f \cdot \frac{dX}{dt}\, dt = f(X(b)) - f(X(a)) \quad .$$

But this is a restatement of (1).

8.2. <u>Definition</u>. An open set D of V is said to be <u>connected</u> if each pair of points of D can be connected by a piecewise smooth curve lying in D.

8.3. <u>Corollary</u>. If D is connected, then f is determined by its gradient up to an additive constant.

<u>Proof</u>. Choose a fixed reference point $X_0 \in D$. Define a scalar function $h: D \longrightarrow R$ by choosing, for each point $X \in D$, a path C from X_0 to X, and setting

$$h(X) = \int_C \nabla f \cdot dX \quad .$$

By (1), $h(X) = f(X) - f(X_0)$. Thus if we know the constant $f(X_0)$, and the gradient ∇f, the function f is completely determined.

8.4. <u>Theorem</u>. Let D be a connected open set in V. Then the following three properties that a continuous vector field $F: D \longrightarrow V$ may possess are equivalent:

(i) F is the gradient of a scalar function $f: D \longrightarrow R$

(ii) If C is any piecewise smooth <u>closed</u> curve in D (i.e. $X(a) = X(b)$), then

$$\int_C F \cdot dX = 0 \quad .$$

(iii) If C_1, C_2 are any two piecewise smooth curves having the same initial and terminal points, then

$$\int_{C_1} F \cdot dX = \int_{C_2} F \cdot dX \quad .$$

(This property is referred to as independence of path.)

<u>Proof</u>. The proof has three parts. We shall show that (i) implies (ii), that (ii) implies (iii), and that (iii) implies (i).

Let $F = \nabla f$ and let C be a closed curve beginning and ending at X_o. Then $X_1 = X_o$ in (1), and therefore

$$\int_C F \cdot dX = \int_C \nabla f \cdot dX = 0 \quad .$$

So (i) implies (ii).

Suppose (ii) holds, and that C_1, C_2 are two curves from X_o to X_1. Let C_1 be represented by a function $X_1(t)$, $a \leq t \leq b$, and C_2 by $X_2(\tau)$, $\alpha \leq \tau \leq \beta$. Define a composite curve C obtained by tracing first C_1 and then $-C_2$:

$$X(t) = \begin{cases} X_1(t) \ , & a \leq t \leq b \ , \\ X_2(\beta - t + b) \ , & b \leq t \leq \beta - \alpha + b \ . \end{cases}$$

Since (ii) holds and C is a closed curve, $\int_C F\cdot dX = 0$. Since $C = C_1 - C_2$, the formulas (4) and (5) of §8 give

$$0 = \int_{C_1 - C_2} F\cdot dX = \int_{C_1} F\cdot dX + \int_{-C_2} F\cdot dX$$

$$= \int_{C_1} F\cdot dX - \int_{C_2} F\cdot dX \quad .$$

Therefore (ii) implies (iii).

For the last part, let F be a vector field having property (iii). Choose a reference point $X_o \in D$. For each point X in D choose a path $C(X)$ from X_o to X in D, and define $f(X)$ by

$$f(X) = \int_{C(X)} F\cdot dX \quad .$$

Because of property (iii), $f(X)$ does not depend on the choice of $C(X)$. To compute the gradient of f at X, let $Y \in V$, and let C_h be the straight line $X(t) = X + tY$ for $t \in [0, h]$. Let $C = C(X)$. Then $C + C_h$ is a curve from X_o to $X + hY$, which lies in D when h is sufficiently small. Then

$$f(X + hY) - f(X) = \int_{C+C_h} F\cdot dX - \int_C F\cdot dX \quad .$$

Using $\int_{C+C_h} = \int_C + \int_{C_h}$, this reduces to

$$\int_{C_h} F\cdot dX = \int_o^h F(X + tY)\cdot\frac{dX}{dt}\, dt$$

$$= \int_o^h F(X + tY)\cdot Y\, dt \quad .$$

Since $g(t) = F(X + tY)\cdot Y$ is continuous, the mean value theorem for integrals states

$$\int_0^h g(t)dt = hg(t_1)$$

for some t_1, $0 < t_1 < h$. This gives

$$f(X + hY) - f(X) = h\ F(X + t_1Y)\cdot Y \quad .$$

Dividing by h and taking the limit as $h \longrightarrow 0$ gives

$$f'(X, Y) = F(X)\cdot Y \quad . \tag{2}$$

Since (2) holds for any $Y \in V$, we conclude that

$$\nabla f = F \quad .$$

Thus we have shown that (iii) implies (i).

§9. Exercises

1. For each of the following vector fields either show that it is not a gradient, or find a scalar function of which it is the gradient.

(a) $2xy^3i + x^2z^3j + 3x^2yz^2k$, in R^3 ,

(b) $yi + zj + xk$, in R^3 ,

(c) $(y^2 - 4y/x^3)i + (2xy + 2/x^2)j$, in R^2 ,

(d) X , in V ,

(e) $i \times X$, in R^3 .

2. If T is an endomorphism of V, show that the vector field $F(X) = TX$ is a gradient if and only if T is symmetric.

3. Evaluate the line integrals

$$I = \int(2x + y)dx + xdy \ , \qquad I' = \int ydx - xdy \ ,$$

along each of the following three paths in R^2 from $(0, 0)$ to (x_o, y_o):

(a) $X(t) = x_o t i + y_o t j$, $0 \leq t \leq 1$,

(b) $X(t) = x_o t^2 i + y_o t j$, $0 \leq t \leq 1$,

(c) the broken line consisting of the segment $0 \leq y \leq y_o$ of the y-axis, and the segment $0 \leq x \leq x_o$ of the line $y = y_o$.

4. Evaluate the line integral

$$\int \frac{xdy - ydx}{x^2 + y^2}$$

(a) once around the circle $x^2 + y^2 = a^2$,

(b) once around the circle $(x - 2)^2 + y^2 = 1$,

(c) along the arc of $x^2 + y^2 = 1$ in the first quadrant from $(1, 0)$ to $(0, 1)$,

(d) along the path from $(1, 0)$ to $(0, 1)$ made up of segments on the lines $x = 1$ and $y = 1$.

VIII. VECTOR-VALUED FUNCTIONS OF A VECTOR

§1. The derivative

1.1. Definition. Let V, W be vector spaces. Let D be an open set in V and let

$$F: D \longrightarrow W$$

be continuous. If $X \in D$ and $Y \in V$, the derivative of F at X with respect to Y is denoted by $F'(X, Y)$ and is defined to be the limit

$$F'(X, Y) = \lim_{h \longrightarrow 0} \frac{1}{h} [F(X + hY) - F(X)]$$

whenever the limit exists.

1.2. Definition. The function $F: D \longrightarrow W$ is said to be differentiable in D if $F'(X, Y)$ exists for each $X \in D$ and $Y \in V$. It is said to be continuously differentiable in D if, for each $Y \in V$, $F'(X, Y)$ exists for each $X \in D$ and is continuous in X.

It is to be observed that these definitions reduce to those given in Chapter VII if $\dim W = 1$.

Examples. If $F: D \longrightarrow W$ is constant, i.e. $F(X) = B \in W$ for each X, then the difference quotient is zero, and hence also its limit. That is, $F'(X, Y) = \tilde{0}_W$ for $X \in D$ and $Y \in V$.

Let $T: V \longrightarrow W$ be linear, and take $D = V$, $F(X) = TX$ for each $X \in V$. Then

$$F(X + hY) - F(X) = T(X + hY) - TX = TX + hTY - TX = hTY \quad .$$

Dividing by h and taking the limit as h tends to zero, we have

$$F'(X, Y) = TY \qquad \text{for all } X, Y \text{ in } V .$$

1.3. <u>Theorem</u>. Let $F, G: D \longrightarrow W$ and $f: D \longrightarrow R$ be continuous. If F, G, f have derivatives at $X \in D$ with respect to $Y \in V$, then $F + G$, fF, $F \cdot G$, and $F \times G$ (if $\dim W = 3$) have derivatives at X with respect to Y, and

(i) $(F + G)'(X, Y) = F'(X, Y) + G'(X, Y)$,

(ii) $(fF)'(X, Y) = f'(X, Y)F(X) + f(X)F'(X, Y)$,

(iii) $(F \cdot G)'(X, Y) = F'(X, Y) \cdot G(X) + F(X) \cdot G'(X, Y)$,

(iv) $(F \times G)'(X, Y) = F'(X, Y) \times G(X) + F(X) \times G'(X, Y)$ (if $\dim W = 3$).

The proof is essentially the same as for Theorem VI, 2.5 and is left as an exercise.

Now assume that W is finite dimensional.

1.4. <u>Proposition</u>. Let $F: D \longrightarrow W$ be continuous. Let $B_1, \ldots, B_n$ be a basis in W, so that

$$F(X) = \Sigma_{i=1}^{n} f_i(X)B_i \tag{1}$$

where the component functions $f_i: D \longrightarrow R$ are continuous and, if the basis is orthonormal, satisfy

$$f_i(X) = F(X) \cdot B_i, \qquad i = 1, \ldots, n . \tag{2}$$

Then the derivative $F'(X, Y)$ exists if and only if the derivatives $f_i'(X, Y)$ exist for each $i = 1, \ldots, n$, and then

(3) $$F'(X, Y) = \Sigma_{i=1}^{n} f_i'(X, Y)B_i \quad .$$

Proof. The constant function $G: D \longrightarrow B_i$ has $G'(X, Y) = \tilde{0}_W$. If $F'(X, Y)$ exists and if the basis is orthonormal, then Theorem 1.3 (iii) applied to (2) shows that $f_i'(X, Y)$ exists and is given by

(4) $$f_i'(X, Y) = F'(X, Y)\cdot B_i \ , \qquad i = 1, \ldots, n.$$

If the given basis is not orthonormal, then f_i is a linear combination of the component functions with respect to some orthonormal basis, from which it follows that $f_i'(X, Y)$ exists. Conversely, if $f_i'(X, Y)$ exists, $i = 1, \ldots, n$, then (3) follows from (1) by Theorem 1.3, (i) and (ii).

1.5. Corollary. The function $F: D \longrightarrow W$ is continuously differentiable in D if and only if its component functions, relative to any basis for W, are continuously differentiable in D.

Proof. If the component functions are continuously differentiable, then (3) shows that F is continuously differentiable, by Proposition VI, 1.6. If F is continuously differentiable, then (4) implies that the component functions f_i relative to an orthonormal basis for W are continuously differentiable, by Corollary VI, 1.4, and this implies that the component functions relative to an arbitrary basis for W are continuously differentiable.

1.6. Theorem. If $F: D \longrightarrow W$ is continuously differentiable in D, then, for each $X \in D$, $F'(X, Y)$ is a linear

function of $Y \in V$.

Proof. Let $B_1, \ldots, B_n$ be a basis for W. By Corollary 1.5, the component functions f_i of F relative to this basis are continuously differentiable in D and therefore, by Theorem VII, 1.6, linear in Y. Then (3) implies that $F'(X, Y)$ is linear in Y.

1.7. Definition. If $F: D \longrightarrow W$, $D \subset V$, is continuously differentiable in D, then the derivative $F'(X)$ of F at X is the linear transformation

$$F'(X): V \longrightarrow W$$

defined by $F'(X)Y = F'(X, Y)$ for each $Y \in D$. Thus,

$$F': D \longrightarrow L(V, W) \quad .$$

Examples. The derivative of a constant is zero. In fact, in this case, $F'(X)$ transforms all of V into $\tilde{0}_W$, so $F'(X)$ is the zero transformation in $L(V, W)$ for each $X \in D$, and $F': D \longrightarrow L(V, W)$ is the constant function zero.

The derivative of a linear function is a constant. For, if $F: V \longrightarrow W$ is defined by $F(X) = TX$ where $T \in L(V, W)$, then

$$F'(X) = T \qquad \text{for each } X \in V \quad .$$

In particular, taking T to be the identity transformation $IX = X$, one can say that the derivative of X is I.

If $F: D \longrightarrow V$, then F is a vector field (Definition VII, 6.1) and $F': D \longrightarrow E(V)$.

Let $F: D \longrightarrow W$, $D \subset V$, be (continuously) differentiable

in D. Let $A_1, \ldots, A_k$ be a basis for V and $B_1, \ldots, B_n$ a basis for W, and set

$$X = \Sigma_{j=1}^{k} x_j A_j, \quad Y = \Sigma_{j=1}^{k} y_j A_j, \quad F(X) = \Sigma_{i=1}^{n} f_i(X) B_i \quad .$$

Then, by (3), and (8) of §1 of Chapter VII,

$$F'(X)Y = \Sigma_{i=1}^{n} \; \Sigma_{j=1}^{k} \; y_j \frac{\partial f_i}{\partial x_j} B_i \quad .$$

In particular, the matrix representation (Definition II, 5.1) of $F'(X)$, relative to the given choices of bases in V and W, is

$$(5) \qquad \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_k} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_k} \\ \cdot & \cdot & \cdots & \cdot \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_k} \end{pmatrix}$$

where each entry in the matrix is a (continuous) scalar-valued function of the vector $X \in D$.

1.8. <u>Definition</u>. The matrix (5) is called the <u>Jacobian matrix</u> of F relative to the given choices of bases in V and W.

§2. Taylor's expansion

Let $F: D \longrightarrow W$ be continuously differentiable in D, and let $X_0 \in D$. Let $D_0 \subset V$ consist of the points of D, excepting X_0. Define a function $E(X, X_0): D_0 \longrightarrow W$ by the identity

(1) $$F(X) = F(X_o) + F'(X_o)(X - X_o) + E(X, X_o)|X - X_o| \quad .$$

If $X \neq X_o$ and $X \in D$, i.e. if $X \in D_o$, it is clear that the equation can be solved for $E(X, X_o)$.

2.1. <u>Proposition</u>. If E is defined as above and if W is finite dimensional, then

(2) $$\lim_{X \to X_o} E(X, X_o) = \tilde{0}_W \quad .$$

<u>Proof</u>. Let $B_1, \ldots, B_n$ be a basis in W. Then

$$F(X) = \Sigma_{i=1}^n f_i(X)B_i, \qquad F'(X)Y = \Sigma_{i=1}^n f_i'(X, Y)B_i \quad ,$$

and

$$E(X, X_o) = \frac{F(X) - F(X_o)}{|X-X_o|} - F'(X_o)\frac{X - X_o}{|X-X_o|}$$

$$= \Sigma_{i=1}^n e_i(X, X_o)B_i \quad ,$$

where

$$e_i(X, X_o) = \frac{f_i(X) - f_i(X_o)}{|X-X_o|} - f_i'(X_o, \frac{X - X_o}{|X-X_o|}) \quad .$$

It is clearly sufficient to show that

(3) $$\lim_{X \to X_o} e_i(X, X_o) = 0, \qquad i = 1, \ldots, n \quad .$$

By the mean value theorem (VII, 1.4), we have

$$\frac{f_i(X) - f_i(X_o)}{|X-X_o|} = f_i'(X_o + \theta(X - X_o), \frac{X - X_o}{|X-X_o|})$$

for some θ with $0 < \theta < 1$. Thus

$$e_i(X, X_o) = [\nabla f_i(X_o + \theta(X - X_o)) - \nabla f_i(X_o)] \cdot \frac{X - X_o}{|X-X_o|} ,$$

so

$$(4) \qquad |e_i(X, X_o)| \leq |\nabla f_i(X_o + \theta(X - X_o)) - \nabla f_i(X_o)|$$

by Theorem III, 3.2 (iv), since $(X - X_o)/|X - X_o|$ has length 1. Now ∇f_i is continuous at X_o so the right side of (4) — and therefore the left side also — tends to zero as X tends to X_o. This completes the proof of Proposition 2.1.

In the case of an ordinary function f of a real variable, the value $f'(x_o)$ is called the slope of the graph of f at the point $(x_o, f(x_o))$. The affine function

$$g(x) = f(x_o) + f'(x_o)(x - x_o)$$

has, as graph, a straight line called the tangent at $(x_o, f(x_o))$.

The mathematical justification of these terms is based on Taylor's formula with remainder

$$f(x) = f(x_o) + f'(x_o)(x - x_o) + e(x, x_o)|x - x_o|,$$

from which it follows that $f(x) - g(x)$ tends to zero, as x tends to x_o, more rapidly than $|x - x_o|$, i.e.

$$\lim_{x \to x_o} \frac{f(x) - g(x)}{|x-x_o|} = \lim_{x \to x_o} e(x, x_o) = 0 .$$

Furthermore, this property characterizes $g(x)$ within the class of all affine functions, for, if

$$h(x) = a + b(x - x_o) ,$$

then

$$\lim_{x \to x_0} \frac{f(x) - h(x)}{|x-x_0|}$$

is infinite unless $a = f(x_0)$; if $a = f(x_0)$, then the limit is $f'(x_0) - b$, which can be zero only if $b = f'(x_0)$.

This formulation of the tangent concept generalizes immediately to vector functions by virtue of Proposition 2.1 and formula (1).

2.2. <u>Theorem</u>. Let $F: D \longrightarrow W$ be continuously differentiable in D, where W is finite dimensional, and let $X_0 \in D$. Then the affine function $G: V \longrightarrow W$ defined by

$$G(X) = F(X_0) + F'(X_0)(X - X_0)$$

is the best affine approximation to F in a neighborhood of X_0 in the sense that it is the only affine function with the property

$$\lim_{X \to X_0} \frac{F(X) - G(X)}{|X-X_0|} = \tilde{0}_W \quad . \tag{5}$$

The above theorem is independent of the particular choices of scalar products in V and W if V is also finite dimensional. (Although the function $E(X, X_0)$ of (1) depends on the scalar product in V, the actual "remainder" $E(X, X_0)|X - X_0|$ clearly does not, nor does the best affine approximation to F near X_0.) This depends on the fact, to be shown in Chapter X, that the notions of limit and continuity, etc. are independent of the particular choices of scalar product in V and W if these vector spaces are finite dimensional. Intuitively, if the values of

$|X - X_0|$, as measured by one scalar product in V, tend to zero, then the same statement will hold for another scalar product on V, even though the actual values assigned to the symbol $|X - X_0|$ may be different.

§3. Exercises

1. Prove Theorem 1.3 (ii).

2. In R^3, let $F(X) = X \times TX$, where T is linear. Show by direct computation that

$$F'(X)Y = Y \times TX + X \times TY \quad .$$

3. Let A be a fixed vector, and $F(X) = (A \cdot X)TX$ where T is linear. Show by direct computation that

$$F'(X)Y = (A \cdot Y)TX + (A \cdot X)TY \quad .$$

4. Let $f(X) = (X \cdot X)^a$ where $a \in R$. Show that

$$\nabla f(X) = 2a(X \cdot X)^{a-1}X$$

and

$$(\nabla f)'(X)Y = 4a(a - 1)(X \cdot X)^{a-2}(X \cdot Y)X + 2a(X \cdot X)^{a-1}Y \quad .$$

Show that X and any vector perpendicular to X are proper vectors of $(\nabla f)'(X)$.

5. Let D be connected and let F and G be functions from D to W such that $F'(X) = G'(X)$ for each $X \in D$. Show that $F - G$ is a constant.

6. Let $F: D \longrightarrow W$ be continuously differentiable in D, where W is finite dimensional. Let $G: [a, b] \longrightarrow D$ be a smooth curve in D. Show that $FG: [a, b] \longrightarrow W$ is smooth and

that

$$(FG)'(t) = F'(G(t), G'(t)) \quad .$$

7. Complete the proof of Theorem 2.2 by showing that the affine function

$$H(X) = A + B(X - X_o) \quad ,$$

with $A \in W$, $B \in L(V, W)$, has the property (5) of §2 only if $A = F(X_o)$, $B = F'(X_o)$.

8. Write out Taylor's expansion in terms of components and partial derivatives in the case $\dim V = 2$ and $\dim W = 3$.

9. Let D be an open set in R^3, and let $f: D \longrightarrow R$ be continuous. Let X_o be a point in D where f has a gradient $\nabla f(X_o) \neq \tilde{0}$. Define $g: R^3 \longrightarrow R$ by

$$g(X) = f(X_o) + \nabla f(X_o)\cdot(X - X_o) \quad .$$

Show that $g^{-1}f(X_o)$ is a plane which is tangent to the level surface of f through X_o.

§4. Divergence and curl

We shall consider now the special case $V = W$. Then $F: D \longrightarrow V$, $D \subset V$, is a vector field in D, and its derivative at a point $X \in D$ is an endormorphism of V. Thus the results of Chapter V concerning endomorphisms can be used to give further interpretations of certain aspects of the derivative in this case.

4.1. Definition. If the vector field F is differentiable in D, the divergence of F at $X \in D$, denoted by $\operatorname{div} F(X)$, is defined by

$$\operatorname{div} F(X) = \operatorname{trace} F'(X)$$

(cf. Definition V, 4.6).

4.2. Proposition. If the vector field F is continuously differentiable in D, then

$$\operatorname{div} F: D \longrightarrow R$$

is continuous.

Proof. Let $A_1, \ldots, A_n$ be a basis for V, with

$$F(X) = \Sigma_{i=1}^{n} f_i(X)A_i, \qquad X = \Sigma_{j=1}^{n} x_j A_j \quad ,$$

and consider the $n \times n$ Jacobian matrix (Definition 1.8) of $F'(X)$, relative to this choice of basis. Then the trace of the endomorphism $F'(X)$ is equal to the sum of the diagonal entries in this matrix (cf. Exercise VI, 6.1 for the case $\dim V = 3$), so

$$(1) \qquad \operatorname{div} F = \frac{\partial f_1}{\partial x_1} + \ldots + \frac{\partial f_n}{\partial x_n} \quad ,$$

where each summand on the right side of (1) is a continuous function of X.

Remark. It is clear that $\operatorname{div} F$ is only a partial representation of F' since, of the n^2 functions $\partial f_i/\partial x_j$ needed to express F' completely, only n appear in (1), and these are not uniquely determined by giving their sum.

Corresponding to a fixed choice of scalar product on V, we have defined the adjoint T^* of any $T \in E(V)$ (Definition V, 5.1) and a unique decomposition (Proposition V, 5.7) of $T \in E(V)$ into the sum of its symmetric and skew-symmetric parts. Applied

to $F'(X)$, this gives

$$F'(X) = F'_+(X) + F'_-(X) \quad , \tag{2}$$

where

$$F'_+(X) = \tfrac{1}{2}(F'(X) + F'(X)^*), \quad F'_-(X) = \tfrac{1}{2}(F'(X) - F'(X)^*) \tag{3}$$

and $F'_+(X)$ is a symmetric endomorphism of V, and $F'_-(X)$ a skew-symmetric endomorphism of V.

Note that we also have $\operatorname{div} F = \operatorname{trace} F'_+$ (cf. Exercise V, 6.8 for the case $\dim V = 3$).

4.3. <u>Proposition</u>. If $F: D \longrightarrow R$ is a scalar function such that its gradient $\nabla f: D \longrightarrow V$ is continuously differentiable in D, then, for each $X \in D$, $(\nabla f)'(X)$ is a symmetric endomorphism of V.

<u>Proof</u>. With respect to an orthonormal basis $A_1, \ldots, A_n$ in V, the component functions of ∇f are $\partial f/\partial x_i$ for $i = 1, \ldots, n$ (see formula (4) of VII, §3); then the Jacobian matrix of $(\nabla f)'$ has as entries the second derivatives $\partial^2 f/\partial x_i \partial x_j$. Since the continuity of these derivatives implies that $\partial^2 f/\partial x_i \partial x_j = \partial^2 f/\partial x_j \partial x_i$, we have $F'(X)^* = F'(X)$, by Proposition V, 5.4.

Now let $\dim V = 3$, and suppose that a scalar product has been chosen on V, together with one of the two vector products corresponding to the chosen scalar product.

4.4. <u>Definition</u>. Let $F: D \longrightarrow V$ be a differentiable vector field in D. Then, for each $X \in D$, the <u>curl</u> of F at X is the unique vector $\operatorname{curl} F(X)$ such that

so the matrix representation of $2F'_{-}(X)$ is

$$\begin{pmatrix} 0 & \frac{\partial f_1}{\partial y} - \frac{\partial f_2}{\partial x} & \frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x} \\ \frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} & 0 & \frac{\partial f_2}{\partial z} - \frac{\partial f_3}{\partial y} \\ \frac{\partial f_3}{\partial x} - \frac{\partial f_1}{\partial y} & \frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z} & 0 \end{pmatrix} .$$

Then, by Exercise V, 9.3,

$$(6) \qquad \text{curl } F(X) = \left(\frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z}\right)i + \left(\frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x}\right)j + \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y}\right)k \quad ,$$

so is continuous by Proposition VI, 1.6.

4.7. <u>Symbolic notation</u>. For dim $V = 3$, let i, j, k be an orthonormal basis for V, such that $i \times j = k$. Write $X = xi + yj + zk$, and let

$$(7) \qquad \nabla = i\frac{\partial}{\partial x} + j\frac{\partial}{\partial y} + k\frac{\partial}{\partial z}$$

denote a symbolic differential operator. In analogy with the formulas given in Theorem III, 5.9 and in Theorem IV, 2.5 for the scalar and vector products in terms of an orthonormal basis, the formulas (1) and (6) may be written symbolically as

$$(1') \qquad \text{div } F = \nabla \cdot F \quad ,$$

$$(6') \qquad \text{curl } F = \nabla \times F \ ,$$

where $F: D \longrightarrow V$ is a vector field in $D \subset V$.

(4) $$2F'_-(X, Y) = \text{curl } F(X) \times Y, \qquad Y \in V .$$

The existence of a vector curl $F(X)$ satisfying (4) follows from Theorem V, 8.4, since $2F'_-(X)$ is a skew-symmetric endomorphism. By Exercise V, 9.2, this vector is uniquely determined (and is therefore zero if and only if F'_- is the zero endormorphism of V).

4.5. <u>Proposition</u>. Let $D \subset V$, $\dim V = 3$. If $f: D \longrightarrow R$ has a continuously differentiable gradient, then

(5) $$\text{curl } (\text{grad } f) = \tilde{0} .$$

<u>Proof</u>. If $F = \text{grad } f$, then F' is symmetric, by Proposition 4.3, so $F'_-(X) = \tilde{0}$.

4.6. <u>Proposition</u>. If $F: D \longrightarrow V$ is continuously differentiable, then the vector field

$$\text{curl } F: D \longrightarrow V$$

is continuous.

<u>Proof</u>. Let i, j, k be an orthonormal basis for V, such that $i \times j = k$. Write

$$X = xi + yj + zk, \quad F(X) = f_1(X)i + f_2(X)j + f_3(X)k .$$

Then the Jacobian matrix of $F'(X)$ is

$$\begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{pmatrix}$$

In the same way, if $f: D \longrightarrow R$ is a scalar function on D, then the notation

$$\text{grad } f = \nabla f \tag{8}$$

adopted in Chapter VII is obtained from the analogue of multiplication of a vector by a scalar. In this notation, the identity (5) becomes

$$\nabla \times \nabla f = \tilde{0} \ . \tag{5'}$$

<u>Remarks</u>. The symbolic notation makes it easy to reproduce the expression of the divergence or the curl of a vector field in terms of an orthonormal basis and to recall, for example, the fact that the divergence of a vector field is a scalar, or the identity (5').

However, it is customary in texts on advanced calculus to take the formulas (1') and (6') with ∇ defined by (7), as the definitions of the divergence and curl of a vector field, and to omit completely the derivative $F'(X)$. Such an approach has the advantage of arriving more quickly at the computational stage. It has the disadvantage of appearing to be an arbitrary procedure. It is not clear why a vector field should have two different derivatives, nor why these particular ones are chosen. The usual answer to these questions is that there is just one differential operator $\nabla = i\partial/\partial x + j\partial/\partial y + k\partial/\partial z$, but there are two products: the scalar and the vector product. Such an answer is nonsense. There are many scalar products; for each scalar product, there are two associated vector products; and, finally, there are infinitely

$$\frac{d}{dt}(Y - X_o) = F_-'(X_o)(Y - X_o) = \frac{1}{2} \text{ curl } F(X_o) \times (Y - X_o)$$

is a small rotation of our small spherical ball, about an axis through X_o parallel to curl $F(X_o)$, with angular velocity $\frac{1}{2}$ |curl $F(X_o)$|. By Theorem VI, 6.2, this motion is volume preserving (see also Exercise 7.2).

The flow determined by

$$\frac{d}{dt}(Y - X_o) = F_+'(X_o)(Y - X_o)$$

is described by Theorem VI, 6.4, since $F_+'(X_o)$ is symmetric, and is a symmetric transformation centered at X_o, having the same proper vectors as $F_+'(X_o)$, and whose proper values are $e^{\lambda_1 t}$, $e^{\lambda_2 t}$, $e^{\lambda_3 t}$ where λ_1, λ_2, λ_3 are the proper values of $F_+'(X_o)$. This distorts the spherical ball slightly into an ellipsoid. The volume of the ball changes by a factor which is the determinant of the transformation (see Chapter IV, §2). By Corollary VI, 6.5, this determinant is exp (t trace $F_+'(X_0)$) and the time rate of change of this factor, when $t = 0$, is trace $F_+'(X_o)$ = trace $F'(X_o)$ = div $F(X_o)$. Finally, the ellipsoid is subjected to a small translation with velocity vector $F(X_o)$.

Thus, the curl of F measures the local rotational effect of the flow whose velocity field is F, and the divergence of F measures the local rate of change of the volume factor. These facts explain the use of the terms divergence and curl.

§6. Harmonic fields

Now in passing from F' to div F and curl F, we are discarding a large part of F'. Specifying div F and curl F

gives only four equations for the nine component functions of F'. So we should not expect to be able to recreate F' and then F from a knowledge of div F and curl F alone.

To determine the extent to which a knowledge of the divergence and curl of a vector field fail to determine the vector field, let F and G be two vector fields in D having the same divergence and curl. Then the difference $H = F - G$ satisfies

$$\nabla\cdot H = \nabla\cdot F - \nabla\cdot G = 0, \quad \nabla \times H = \nabla \times F - \nabla \times G = \tilde{0}$$

(cf. Exercise 7.1).

6.1. <u>Definition</u>. A vector field F in $D \subset V$, dim $V = 3$, is called a <u>harmonic field</u> if and only if

$$(1) \qquad \nabla\cdot F = 0, \qquad \nabla \times F = \tilde{0} \ .$$

6.2. <u>Definition</u>. A function f with a continuously differentiable gradient is called a <u>harmonic function</u> if

$$(2) \qquad \Delta f = \nabla\cdot\nabla f = 0.$$

6.3. <u>Proposition</u>. The gradient of a harmonic function is a harmonic vector field.

<u>Proof</u>. The equations (1) for $F = \nabla f$ follow from (2), and (5') of §4.

<u>Examples</u>. In terms of an orthonormal basis, with the usual notation, we have

$$(3) \qquad \nabla\cdot\nabla f = \Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \ .$$

Thus, the equation $\Delta f = 0$ is the Laplace equation, and any solution of this equation is a harmonic function. There are many such functions, e.g. any linear function $ax + by + cz$, special quadratics such as $ax^2 + by^2 - (a + b)z^2$, and other functions such as $(x^2 + y^2 + z^2)^{-1/2}$, and $e^x \cos y$.

A harmonic function (or its negative in certain applications) is called the <u>potential function</u> of the steady flow whose velocity field is $F = \nabla f$, and the level surfaces of f are then called <u>equipotential surfaces</u>.

<u>Remark</u>. For a certain class of domains D, which will be defined in Chapter XII, every harmonic field is the gradient of a harmonic function in D. The additional assumption about D is needed to ensure that the necessary condition $\nabla \times F = \tilde{0}$ for F to be a gradient is also sufficient to ensure that F have the property (ii) or (iii) of Theorem VII, 8.4.

§7. <u>Exercises</u>

1. Take V of dimension 3, and let F, G be differentiable vector fields in V, and $a \in R$. Show that

$$\nabla \cdot aF = a(\nabla \cdot F), \qquad \nabla \times aF = a(\nabla \times F) \quad ,$$

$$\nabla \cdot (F + G) = \nabla \cdot F + \nabla \cdot G, \qquad \nabla \times (F + G) = \nabla \times F + \nabla \times G \quad .$$

2. Show that, for any vector field F in V,

$$\nabla \cdot (\nabla \times F) = 0$$

whenever the left side of this equation is defined.

3. Show that the "gravitational" field

$$F(X) = -\frac{X}{|X|^3}, \qquad X \neq \tilde{0} ,$$

in R^3 is a harmonic field. Find a scalar function whose gradient is F.

4. Let A be a fixed non-zero vector in R^3, and take D to be the points of R^3 not in $L(A)$. Define $F: D \longrightarrow R^3$ by

$$F(X) = \frac{A \times X}{|A \times X|^2} .$$

Show that F is a harmonic field.

IX. TENSOR PRODUCTS AND THE STANDARD ALGEBRAS

§1. Introduction

In order to enlarge the interpretations and applications of the notions developed in earlier chapters, we shall study further the construction of new vector spaces from given vector spaces, and the relationships between the spaces so constructed. In this introduction we shall discuss several concepts which will be needed later.

Certain constructions have been considered previously. For example, given a vector space V, a linear subspace U of V is another vector space, and we can construct the quotient vector space V/U (Definition II, 11.1). Again, given two vector spaces U and W, we may construct their direct sum $V = U \oplus W$ (Definition II, 11.8).

Or, if D is any non-empty set and W is a vector space, the set W^D of all functions with domain D and range W is a vector space, with addition and multiplication by scalars in W^D derived from these operations in the range W (Exercise I, 4.12). We have also considered certain linear subspaces of W^D in cases where more is known about the domain D. For example, if D is an open subset of a vector space, the linear subspace of continuous W-valued functions is defined, or the smaller subspace consisting of the W-valued functions which are

continuously differentiable. If V is a vector space, then $L(V, W)$ is a linear subspace of W^V, etc.

We have noted that vector spaces which are isomorphic are essentially equivalent, in that every element or operation in the one has a precise counterpart in the other, and vice versa. Also, an injective linear transformation determines an isomorphism between its domain and a linear subspace of its range.

If we are primarily interested in studying the relationship between two vector spaces which happen to be isomorphic, it is not always advisable to make the identification, made possible by the existence of an isomorphism, of the two vector spaces. On the one hand, any relationship becomes trivial after identification. On the other hand, some isomorphisms are better suited for this purpose than others.

1.1. <u>Definition</u>. Let V and W be vector spaces. A linear transformation $T: V \longrightarrow W$ is called <u>canonical</u> if it depends only on the properties of V and W as vector spaces (and not on some further choice, as of scalar products, or of bases, etc.). In general the word "canonical" is applied to any object, not necessarily a linear transformation, whose definition is independent of arbitrary choices.

<u>Remarks</u>. A linear transformation T may be defined in terms of a choice, and nevertheless be canonical if the resulting T is independent of the particular choice made. As examples of canonical linear transformations we may mention the inclusion $\iota: U \longrightarrow V$, where U is a linear subspace of V (Definition

II, 11.4) or $j: V \longrightarrow V/U$.

In developing a general theory, identification of vector spaces is made only by means of canonical isomorphisms and only in cases where an essential relationship is not absorbed by the identification. In such a case, we write $V = W$ rather than $V \longrightarrow W$ and the isomorphism, once it has been demonstrated, is not explicitly mentioned thereafter.

In cases where the relationship between vector spaces is an essential aspect, different notations will be used to help keep the relationships clear, reserving the usual notation A, B, X, Y, etc. for vectors in the given vector space V. This has already been done, for example, in using T, S to denote linear transformations from V into W, even though T, S are themselves elements of the vector space $L(V, W)$.

We shall illustrate some of the above notions by considering the case of the dual vector space $V^* = L(V, R)$ (Definition II, 8.4) of a given vector space V.

The elements of V^* are often called <u>linear forms</u> on V, rather than linear transformations on V with range R. We shall denote elements of V^* by ω or other lower case Greek letters, rather than by T, S, etc., and adopt a standard notation

$$\omega A = \langle A, \omega \rangle , \qquad A \in V, \ \omega \in V^* ,$$

for the image of A under $\omega: V \longrightarrow R$. [Note: lower case Greek letters, such as ι, κ, μ, ν are also used to denote certain canonical transformations.]

If V is finite dimensional, then V and V^* are isomorphic, by Corollary II, 8.5 and Theorem II, 7.5. However, there is no canonical isomorphism, and V and V^* are not isomorphic if V is not finite dimensional.

1.2. Proposition. Let V be finite dimensional, and let $A_1,\ldots,A_k$ be a basis for V. For each $i = 1,\ldots,k$, let $\omega^i \in V^*$ denote the linear form $\omega^i: V \longrightarrow R$ determined by the values

$$(1) \qquad \langle A_j, \omega^i \rangle = \delta^i_j = \begin{cases} 1, & \text{if } j = i, \\ 0, & \text{if } j \neq i. \end{cases}$$

Then $\omega^1,\ldots,\omega^k$ is a basis for V^*.

1.3. Definition. The basis of V^* constructed in Proposition 1.2 is called the dual basis of V^* relative to the given basis for V.

Proof of Proposition 1.2. (a) The conditions (1) determine a unique linear form ω^i for each fixed i, by Theorem II, 4.1, with $W = R$. (b) The elements $\omega^1,\ldots,\omega^k$ are independent, that is,

$$(2) \qquad \omega = \Sigma_{i=1}^k a_i\omega^i, \qquad a_i \in R,$$

and $\omega = \tilde{0}$ imply $a_i = 0$, $i = 1,\ldots,k$. In fact, for each $j = 1,\ldots,k$, we can evaluate the linear form ω^i on the basis element A_j and we obtain

$$0 = \langle A_j, \omega \rangle = \langle A_j, \Sigma_{i=1}^k a_i\omega^i \rangle$$
$$= \Sigma_{i=1}^k a_i \langle A_j, \omega^i \rangle$$
$$= \Sigma_{i=1}^k a_i\delta_j^i = a_j \quad .$$

(c) The elements $\omega^1,\ldots,\omega^k$ span V^*, that is, any linear form ω on V can be expressed in the form (2). In fact, if

$$a_i = \langle A_i, \omega \rangle , \qquad i = 1,\ldots,k ,$$

the right side of (2) gives a linear form which coincides with ω, by the uniqueness statement of Theorem II, 4.1.

<u>Remarks</u>. If $\dim V = 1$, then any non-zero element A can be taken as a basis for V, and the corresponding basis ω for V^* may be called "the" dual of A. (Note, however, that the correspondence which sends each non-zero element of V into its dual is not linear. Why?)

The isomorphism $T: V \longrightarrow V^*$ determined by the values $TA_j = \omega^j$, $j = 1,\ldots,k$, is not canonical. For example, let $\dim V = 2$. Then $\tilde{A}_1 = A_1 + A_2$, $\tilde{A}_2 = A_2$ is another basis for V, whose dual basis is $\tilde{\omega}^1 = \omega^1$, $\tilde{\omega}^2 = \omega^2 - \omega^1$. The analogous $\tilde{T}: V \longrightarrow V^*$ defined by the new bases sends A_2 into $\omega^2 - \omega^1$ rather than into ω^2. That is, $\tilde{T} \neq T$.

In this book, we shall not prove the theorem that <u>every</u> vector space V has a basis, that is, a subset D which is independent and satisfies $L(D) = V$. The set D is, of course, not finite in general. [However, we shall give an example below

(Definition 1.5) in which a non-finite basis can be exhibited.] In the general case we can again define the dual of a basis element (relative to a given basis) to be the linear form which has the value 1 on the given basis element and 0 on all other basis elements. The resulting collection of linear forms is independent, by the same argument as in (b) above, but does not give a basis for V^* if V is not finite dimensional. For example, the linear form which has the value 1 on all basis elements in V cannot be expressed as a <u>finite</u> linear combination of linear forms of the constructed type.

A choice of scalar product on V determines a linear transformation $S: V \longrightarrow V^*$, where $\omega = SA$ is defined by the values

$$(3) \qquad \langle X, \omega \rangle = X \cdot A , \qquad X \in V$$

By Lemma V, 5.2, S is injective and, by Theorem III, 5.8, is an isomorphism if V is finite dimensional. The inverse of S is used in defining the gradient $\nabla f \in V$ of a scalar-valued function on V from the derivative $f' \in V^*$. In the case $V = R^n$, with the standard scalar product, the above isomorphism is often used to identify V with V^*. This identification is not justified in a wider context, since S is not canonical. In fact, let $\tilde{S}: V \longrightarrow V^*$ be the analogous transformation corresponding to another choice of scalar product on V, and let $A \in V$ be such that $A \cdot A \neq A \tilde{\cdot} A$. Then $\langle A, SA \rangle \neq \langle A, \tilde{S}A \rangle$, that is, $SA \neq \tilde{S}A$, or $S \neq \tilde{S}$.

1.4. Proposition. Let $V^{**} = L(V^*, R)$ and let $\mu: V \longrightarrow V^{**}$ be defined as follows: μA is the linear form on V^* determined by

$$< \omega, \mu A > = < A, \omega > , \qquad \omega \in V^* .$$

Then μ is a canonical injective linear transformation which is an isomorphism if V is finite dimensional.

Proof. (a) μ is linear since, for fixed $A, B \in V$, $r \in R$, and all $\omega \in V^*$,

$$< \omega, \mu(A + B) > = < A + B, \omega > = < A, \omega > + < B, \omega >$$

$$= < \omega, \mu A > + < \omega, \mu B > = < \omega, \mu A + \mu B >$$

and

$$< \omega, \mu(rA) > = < rA, \omega > = r < A, \omega >$$

$$= < A, r\omega > = < r\omega, \mu A > = < \omega, r\mu A > .$$

(b) μ is injective, that is, $\ker \mu = \tilde{0}$. For, if $A \neq \tilde{0}$, there is a linear form ω such that $< A, \omega > \neq 0$. Then, for this choice of ω, we have $< \omega, \mu A > \neq 0$, that is, $\mu A \neq \tilde{0}_{V^{**}}$. [To show the existence of ω, choose a basis for V. Then the given element A is expressed as a finite sum $\Sigma_{\alpha=1}^{n} a^{\alpha} A_{\alpha}$ where the A_{α}'s are elements of the chosen basis. (The indices α here represent an arbitrary numbering of the finite number of basis elements involved in the expression of the given element A.) Let ω^{α} be the dual of A_{α} (relative to the chosen basis) and take $\omega = \Sigma_{\beta=1}^{n} a^{\beta} \omega^{\beta}$. Then

$$< A, \omega > = < \Sigma_{\alpha=1}^{n} a^{\alpha} A_{\alpha} , \Sigma_{\beta=1}^{n} a^{\beta}\omega^{\beta} > = \Sigma_{\alpha=1}^{n} (a^{\alpha})^2 ,$$

so $< A, \omega > = 0$ only if $A = \tilde{0}$.] (c) If V is finite dimensional, then $\dim V = \dim V^* = \dim V^{**}$, and μ is an isomorphism by Theorem II, 7.3.

Remarks. The canonical transformation μ of Proposition 1.4 is always used to identify V with a linear subspace of V^{**}. In practice, this identification means that the symbol $< A, \omega >$, whose values for fixed $\omega \in V^*$ and varying $A \in V$ express $\omega: V \longrightarrow R$, also expresses $A: V^* \longrightarrow R$ for fixed A and varying $\omega \in V^*$. However, if V is not finite dimensional, not all linear forms on V^* can be obtained in this way.

1.5. Definition. The vector space generated by an arbitrary (non-empty) set D is the linear subspace of R^D consisting of the real-valued functions on D which have the value 0 on all but a finite number of elements of D. This vector space will be denoted by $(R^D)_o$.

It is left as an exercise to verify that these elements do indeed form a vector space. Note that no special property of D is needed in order to define this linear subspace of R^D (which could also be defined if R is replaced by an arbitrary vector space W, but would not have all the properties to be demonstrated below for the case $W = R$).

1.6. Proposition. For each $X \in D$, let $[X]$ denote the function which has the value 1 at X and 0 at all other elements of D. Then the subset $[D] \subset (R^D)_o$ consisting of the

elements $[X]$, for all $X \in D$, is a basis for $(R^D)_o$ (which is infinite unless the set D has only a finite number of elements, in which case $(R^D)_o$ coincides with R^D).

Proof. In view of the definition of addition and multiplication by a scalar in R^D, a finite linear combination

$$f = \Sigma_{\alpha=1}^{n} a^{\alpha}[X_{\alpha}] \tag{4}$$

of elements of $[D]$ is the function which has the value a^{α} at the point $X_{\alpha} \in D$, $\alpha = 1,\ldots,n$, and is zero at any other points of D; in particular, $f \in (R^D)_o$. Clearly, if f is the function which has the value zero at every $X \in D$, then each a^{α} is zero. Thus $[D]$ is an independent set (Definition I, 9.1) since no finite subset is dependent. Moreover, any $f \in (R^D)_o$ can be expressed in the form (4), so $L([D]) = (R^D)_o$. That is, $[D]$ gives a basis for $(R^D)_o$.

Remarks. The basis demonstrated in Proposition 1.6 will be called the standard basis for $(R^D)_o$. If the basis is infinite, any coefficient appearing in an expression (4) for $f \in (R^D)_o$ is uniquely determined, but the number of basis elements used is not, since an arbitrary (but finite) number of additional terms can always be included in the expression, corresponding to points of D at which f is zero.

It is customary to use "generate" as synonymous with "span". This would be true in the present case after an identification of X with $[X]$, and of D with $[D]$. If D happens to be a vector space, it should be noted that the inclusion

$$\kappa : D \longrightarrow (R^D)_o \ , \qquad \kappa(X) = [X] \ ,$$

is not a linear transformation. For example, $rX \in D$ would be defined for $X \in D$, $r \in R$, but $[rX] \neq r[X]$. However, when addition and multiplication by a scalar have no meaning in D, there is no risk of confusion if we omit the brackets, that is, if we identify D with a subset of $(R^D)_o$.

1.7. Proposition. Let D be a non-empty set and let W be an arbitrary vector space. For each function $F \in W^D$, there is a unique $T \in L((R^D)_o, W)$ such that $F = T\kappa$. [If κ is taken as an identification, then T is an extension of F from the subset D to the whole of $(R^D)_o$.]

Proof. This is essentially Theorem II, 4.1 without the assumption that the domain of T is finite dimensional. The function T is defined on the standard basis of $(R^D)_o$ by

$$T[X] = F(X) \ ,$$

and extended to $(R^D)_o$ by linearity; that is, for $f \in (R^D)_o$ given by (4), we define

$$Tf = \Sigma_{\alpha=1}^{n} a^{\alpha} T[X_\alpha] = \Sigma_{\alpha=1}^{n} f(X_\alpha) F(X_\alpha) \quad .$$

Note that Tf is well-defined, since any variations in the expression used for f have zero coefficients. The linearity and uniqueness of T follow as in the proof of Theorem II, 4.1.

1.8. Definition. If D and E are arbitrary sets,

the <u>product</u> of D and E, denoted by $D \times E$, is the set whose elements are pairs (X, Y), $X \in D$, $Y \in E$.

<u>Remark</u>. If V and W are vector spaces, the set $V \times W$, which consists of pairs of vectors, is not a vector space. The operations of addition and multiplication by a scalar can be defined for elements of $V \times W$ by taking

$$(A, B) + (A', B') = (A + A', B + B') ,$$

$$r(A, B) = (rA, rB) ,$$

where $A, A' \in V$, $B, B' \in W$, $r \in R$; but the vector space so constructed is denoted by $V \oplus W$ (Definition II, 11.8) rather than by $V \times W$.

However, if Z is an arbitrary vector space, the fact that V and W are vector spaces is used to define an important linear subspace of the vector space $Z^{V \times W}$ In general, any function F in $Z^{D \times E}$ determines a function $F_A \in Z^E$ for each $A \in D$ and a function $F_B \in Z^D$ for each $B \in E$ by defining $F_A(Y) = F(A, Y)$ and $F_B(X) = F(X, B)$.

1.9. <u>Definition</u>. Let V, W, Z be vector spaces. An element T of $Z^{V \times W}$ is called a <u>bilinear transformation</u> if $T_A \in L(W, Z)$ for each fixed $A \in V$ and if $T_B \in L(V, Z)$ for each fixed $B \in W$. The vector space of the bilinear transformations from $V \times W$ to Z will be denoted by $L(V, W; Z)$.

<u>Remarks</u>. The explicit conditions that $T \in Z^{V \times W}$ be a bilinear transformation are

$$
(6) \qquad \begin{aligned} T(X + X', Y) &= T(X, Y) + T(X', Y) \ , \\ T(X, Y + Y') &= T(X, Y) + T(X, Y') \ , \\ T(rX, Y) &= rT(X, Y) = T(X, rY) \ , \end{aligned}
$$

for $X, X' \in V$, $Y, Y' \in W$, $r \in R$. It is left as an exercise to verify that $L(V, W; Z)$ is a linear subspace of $Z^{V \times W}$ (and therefore a vector space).

If $T: V \times W \longrightarrow Z$ is bilinear, it is not true, in general, that $\operatorname{im} T$ is a linear subspace of Z, as in the case of a linear transformation. For example, the conditions (6) need not imply that $T(X, Y) + T(X', Y')$ can be written in the form $T(A, B)$ for some $A \in V$, $B \in W$. Thus, the function $T_1: V \times W \longrightarrow \operatorname{im} T \subset Z$, induced by T, is bijective but not bilinear, according to Definition 1.9, if $\operatorname{im} T$ is not a vector space. The induced function $T': V \times W \longrightarrow L(\operatorname{im} T)$, where $L(\operatorname{im} T)$ is the linear subspace of Z spanned (or generated) by $\operatorname{im} T$, is bilinear but not bijective in general. However, we can still write $T = \iota T'$, where $\iota: L(\operatorname{im} T) \longrightarrow Z$ is linear and injective.

<u>Examples</u>. (i) In Definition I, 1.1, the process of defining an operation of addition on a set V can be described as the specification of a function $F: V \times V \longrightarrow V$, where we write $A + B$ for $F(A, B) \in V$. However, an arbitrary function $V \times V \longrightarrow V$ cannot be used for this purpose: the function F must be such that the resulting law of composition satisfies Axioms 1 - 4. For example, Axiom 3 may be described as follows:

for a certain element of the set V, which will be denoted by $\tilde{0}$, the function $F_{\tilde{0}}: V \longrightarrow V$ is the identity function on V. The process of defining the further operation of multiplication by a scalar can be described as the specification of a function $S: R \times V \longrightarrow V$, where we write rA for $S(r, A) \in V$. Again, the function S must satisfy certain conditions relative to the addition already defined in R. For example, (i) and (ii) of Axiom 8 require that $S_0: V \longrightarrow V$ be the constant map $A \longrightarrow \tilde{0}$ and that $S_1: V \longrightarrow V$ be the identity function on V. (ii) Let V be a vector space and let $T: V \times V^* \longrightarrow R$ be defined by $T(A, \omega) = \langle A, \omega \rangle$, $A \in V$, $\omega \in V^*$. Then $T \in L(V, V^*; R)$. (iii) Let V be a vector space. The specification of a particular scalar product on V can be described by a $T \in L(V, V; R)$ where we write $A \cdot B$ for $T(A, B) \in R$, provided that T satisfies certain conditions in order that Axioms S1, S4, and S5 of Definition III, 1.1 be satisfied. (Axioms S2 and S3 are covered by the fact that T is bilinear.)

1.10. <u>Proposition</u>. Let V, W, Z be vector spaces. Then

$$L(V, W; Z) = L(V, L(W, Z)) = L(W, L(V, Z)) \quad .$$

<u>Proof</u>. For each $T \in L(V, W; Z)$ and each $A \in V$, we have $T_A \in L(W, Z)$. Given $T \in L(V, W; Z)$, let $\hat{T}: V \longrightarrow L(W, Z)$ be defined by $A \longrightarrow T_A$. It follows from (6) that $\hat{T}$ is linear, that is, $\hat{T} \in L(V, L(W, Z))$. It is easily verified that the function $L(V, W; Z) \longrightarrow L(V, L(W, Z))$ defined by $T \longrightarrow \hat{T}$ is

linear. To show that this canonical linear transformation is an isomorphism, it is sufficient to exhibit the inverse correspondence. Given $\hat{T} \in L(V, L(W, Z))$, we can define a $T \in L(V, W; Z)$ by

$$(7) \qquad T(X, Y) = (\hat{T}X)Y , \qquad X \in V, Y \in W .$$

In practice, the identification is merely a reinterpretation of the function $T(X, Y)$, first with $X \in V$ fixed as $Y \in W$ varies, as in (7), and then allowing $X \in V$ to vary.

The relation $L(V, W; Z) = L(W, L(V, Z))$ is proved similarly.

Example. The linear transformation $S: V \longrightarrow V^*$ defined by (3) is the interpretation in $L(V, L(V, R))$ of the corresponding scalar product, an element of $L(V, V; R)$.

1.11. Definition. The product of a finite number of sets $D_1, D_2, \ldots, D_n$, denoted by $D_1 \times D_2 \times \ldots \times D_n$, is the set whose elements are sequences $(X_1, X_2, \ldots, X_n)$, $X_i \in D_i$, $i = 1, \ldots, n$.

1.12. Definition. Let $V_1, \ldots, V_n$, Z be vector spaces. An element T of $Z^{V_1 \times V_2 \times \ldots \times V_n}$ is called a multilinear transformation if the function $V_j \longrightarrow Z$, induced by T for each fixed choice of an element in each V_i, $i \neq j$, is linear. The vector space of the multilinear transformations from $V_1 \times V_2 \times \ldots \times V_n$ to Z is denoted by $L(V_1, V_2, \ldots, V_n; Z)$.

1.13. Proposition. Let $V_1, \ldots, V_n$, Z be vector spaces. Then

$$L(V_1, V_2, \ldots, V_n; Z) = L(V_1, V_2, \ldots, V_{n-1}; L(V_n, Z))$$
$$= L(V_1, V_2, \ldots, V_{n-2}; L(V_{n-1}, L(V_n, Z))) ,$$

etc.

The proof of this proposition is left as an exercise.

§2. The tensor product

2.1. Definition. The tensor product of two vector spaces V and W, denoted by $V \otimes W$, is the vector space constructed as follows. (i) Let $F(V, W) = (R^{V \times W})_o$; then the standard basis for $F(V, W)$ is the set of all pairs (A, B), $A \in V$, $B \in W$. (ii) Let $K(V, W)$ be the linear subspace of $F(V, W)$ spanned by all elements of the form

$$(1) \qquad \begin{aligned} &(A + A', B) - (A, B) - (A', B) , \\ &(A, B + B') - (A, B) - (A, B') , \\ &(rA, B) - r(A, B), \quad (A, rB) - r(A, B) , \end{aligned}$$

where $A, A' \in V$, $B, B' \in W$, $r \in R$. (iii) Then $V \otimes W$ is the quotient space $F(V, W)/K(V, W)$, or

$$(2) \qquad \tilde{0} \longrightarrow K(V, W) \longrightarrow F(V, W) \xrightarrow{j} V \otimes W \longrightarrow \tilde{0} .$$

We write $A \otimes B$ for $j(A, B)$, $A \in V$, $B \in W$.

Remarks (cf. Definition II, 11.1). An element of $V \otimes W$ represents an equivalence class of elements of $F(V, W)$, two elements lying in the same class if and only if their difference is an element of the kernel $K(V, W)$. An arbitrary element

f of $F(V, W)$ can be expressed, in terms of the standard basis, as a finite sum $\Sigma_{\alpha=1}^{n} c^{\alpha}(A_{\alpha}, B_{\alpha})$, $A_{\alpha} \in V$, $B_{\alpha} \in W$. Since j is linear, we write $\Sigma_{\alpha=1}^{n} c^{\alpha} A_{\alpha} \otimes B_{\alpha}$ for $j(f)$. Obviously, the symbol used to denote a given element of $V \otimes W$ is not uniquely determined. Nevertheless these symbols can be used for computation in $V \otimes W$, since the non-uniqueness in the symbol used derives from an element of $K(V, W)$, and any element of $K(V, W)$ lies in the class $\tilde{0} = \tilde{0}_{V \otimes W}$. The basic rules for computation in $V \otimes W$ derive from the same fact:

$$
\begin{aligned}
&(A + A') \otimes B = A \otimes B + A' \otimes B \ , \\
(3) \qquad &A \otimes (B + B') = A \otimes B + A \otimes B' \ , \\
&(rA) \otimes B = r(A \otimes B) = A \otimes (rB) \ .
\end{aligned}
$$

Since the elements (A, B), $A \in V$, $B \in W$, form a basis for $F(V, W)$, it is clear that the elements corresponding to the symbols $A \otimes B$, $A \in V$, $B \in W$, span $V \otimes W$. However, the collection of classes so designated do not form an independent set in $V \otimes W$ (unless $V = W = \tilde{0}$) and therefore do not give a basis for $V \otimes W$, but merely a set of generators of $V \otimes W$.

The fundamental theorem (II, 4.1) concerning the existence and uniqueness of linear transformations must be modified if we wish to use a set of generators in place of a basis. The uniqueness part of the proof holds in this case also and we have

2.2. <u>Proposition</u>. A linear transformation is uniquely determined by its values on a set of generators for its domain.

However, there need not exist a linear transformation

having specified values on a set of generators, unless these values satisfy certain conditions (cf. Corollary 2.5) to offset the fact that the expression of an element in terms of a set of generators is not unique as in the case of a basis.

2.3. Proposition. The function

$$\tau: V \times W \longrightarrow V \otimes W$$

defined by $\tau(A, B) = A \otimes B$, that is, $\tau = j\kappa$, is a bilinear transformation.

Proof. The conditions (6) of §1 for τ to be bilinear are exactly the formulas (3).

2.4. Theorem. Let V, W, Z be vector spaces. Then

$$L(V \otimes W, Z) = L(V, W; Z) \quad .$$

Proof. Let

$$\tau^*: L(V \otimes W, Z) \longrightarrow L(V, W; Z)$$

be defined by $\tau^* T = T\tau$, $T \in L(V \otimes W, Z)$. It is left as an exercise to verify that $T\tau: V \times W \longrightarrow Z$ is a bilinear transformation for each $T \in L(V \otimes W, Z)$ and that τ^* is linear. Note that $T\tau(A, B) = T(A \otimes B)$, from which it is clear that τ^* is injective, by Proposition 2.2. To show that τ^* is an isomorphism it is sufficient (Exercise II, 3.2) to construct a function

$$\sigma: L(V, W; Z) \longrightarrow L(V \otimes W, Z)$$

such that $\tau^*\sigma$ is the identity on $L(V, W; Z)$. Given $S \in L(V, W; Z)$, let $\hat{T} \in L(F(V, W), Z)$ be determined as in

Proposition 1.7, with $S = \hat{T}\kappa$. Since S is bilinear, T is zero on $K(V, W)$ and therefore, by Proposition II, 11.3, there is a $T \in L(V \otimes W, Z)$ such that $\hat{T} = Tj$. We define $\sigma S = T$. Then $\tau^*\sigma S = T\tau = Tj\kappa = \hat{T}\kappa = S$.

Example. A scalar product on V may equally well be defined as an element of $L(V \otimes V, R)$, satisfying certain additional conditions, rather than as an element of $L(V, V; R)$.

2.5. Corollary. There exists a $T \in L(V \otimes W, Z)$ having specified values $T(A \otimes B)$ on the generators $A \otimes B$ of $V \otimes W$ if and only if the function $F: V \times W \longrightarrow Z$ defined by $F(A, B) = T(A \otimes B)$ is a bilinear transformation.

Remark. There are more distinct symbols $A \otimes B$ than there are classes represented by these symbols; e.g. $2A \otimes B = A \otimes 2B$. Nevertheless, the values $T(A \otimes B)$, apparently given for the symbols, give a well-defined value on each class if the conditions of Corollary 2.5 are satisfied. Thus, in this context, there is no confusion if we speak of the symbols themselves as generators.

2.6. Corollary. $L(V \otimes W, Z) = L(V, L(W, Z))$.

This result is obtained by combining Theorem 2.4 and Proposition 1.10.

The construction given in the proof of Theorem 2.4 is often stated as follows: any bilinear transformation S on $V \times W$ can be "factored through" $V \otimes W$, that is, expressed in the form $T\tau$, where T is a linear transformation on $V \otimes W$. Note that $T = \sigma S$ is uniquely determined by S, since $\sigma = \tau^{*-1}$ is also

an isomorphism. This "universal factorization property" characterizes the tensor product to within isomorphism.

2.7. <u>Theorem</u>. Let V, W, U_1, U_2 be vector spaces and let

$$\rho_i : V \times W \longrightarrow U_i , \qquad i = 1, 2 ,$$

be bilinear. Suppose further that, for every choice of a vector space Z, each bilinear transformation

$$S : V \times W \longrightarrow Z$$

can be written uniquely in the form $S = T_i \rho_i$, where $T_i \in L(U_i, Z)$, $i = 1, 2$; that is, each couple (U_i, ρ_i) has the <u>universal factorization property</u> for $V \times W$. Then U_1 and U_2 are isomorphic. In particular, U_i is isomorphic to $V \otimes W$.

<u>Remark</u>. The uniqueness condition in the hypothesis of Theorem 2.7 is equivalent to the condition $L(\text{im } \rho_i) = U_i$, $i = 1, 2$. For example, take $i = 1$. Let $\rho_1' : V \times W \longrightarrow L(\text{im } \rho_1)$ be induced by ρ_1, so $\rho_1 = \iota_1 \rho_1'$, where $\iota_1 : L(\text{im } \rho_1) \longrightarrow U_1$. If we apply the factorization hypothesis for $S = \rho_1'$, we get $\rho_1' = T_1 \rho_1$, where $T_1 \in L(U_1, L(\text{im } \rho_1))$. Then

$$\rho_1 = \iota_1 \rho_1' = \iota_1 T_1 \rho_1 ,$$

where $\iota_1 T_1 \in L(U_1, U_1)$. Since $\rho_1 = I_{U_1} \rho_1$, the assumption of uniqueness (applied for $S = \rho_1$) implies $\iota_1 T_1 = I_{U_1}$ and, in particular, that ι_1 is surjective. Conversely, if an

arbitrary bilinear S can be written (not necessarily uniquely) in the form $S = T_1\rho_1$, then

$$S = T_1(\iota_1\rho_1') = (T_1\iota_1)\rho_1' ;$$

that is, the vector space $L(\text{im } \rho_1)$ has the factorization property relative to ρ_1'. But $T_1\iota_1$ is uniquely determined by its values on a set of generators for $L(\text{im } \rho_1) = L(\text{im } \rho_1')$, viz. the elements of $\text{im } \rho_1'$, and these values are uniquely determined by S. If we have $L(\text{im } \rho_1) = U_1$, then it follows that $T_1 = T_1\iota_1$ is uniquely determined by S.

<u>Proof</u>. The factorization hypotheses, applied for $S = \rho_1$ and $S = \rho_2$, give $\rho_1 = T_2\rho_2$, $\rho_2 = T_1\rho_1$, where $T_i \in L(U_i, U_j)$, $i = 1, 2$, $j \neq i$. Then

$$\rho_1 = T_2T_1\rho_1 , \qquad \rho_2 = T_1T_2\rho_2 .$$

The uniqueness hypotheses then give $T_2T_1 = I_{U_1}$ and $T_1T_2 = I_{U_2}$, that is, T_1 and T_2 are isomorphisms.

Thus any two vector spaces having the universal factorization property for $V \times W$ are isomorphic; any such vector space is therefore isomorphic to the vector space $V \otimes W$ constructed in Definition 1.1, which has already been shown to have this property.

If the universal factorization property is taken as the definition of the tensor product, which is possible because of Theorem 2.7, then the construction in Definition 1.1, together with Theorem 2.4, serves as an existence theorem.

2.8. Theorem. For any vector space V, we have

$$R \otimes V = V = V \otimes R \quad .$$

Proof. The values $\mu(r \otimes A) = rA$, $r \in R$, $A \in V$, determine a linear transformation $\mu: R \otimes V \longrightarrow V$. Let $\nu: V \longrightarrow R \otimes V$ be defined by $\nu A = 1 \otimes A$. Then $\mu\nu = I_V$ and $\nu\mu = I_{R \otimes V}$. In fact, $\nu\mu(r \otimes A) = r \otimes A$ (since $1 \otimes rA = r \otimes A$); we then apply Proposition 2.2. Thus μ gives a canonical isomorphism. The identification $V \otimes R \longrightarrow V$ is proved similarly.

2.9. Proposition. For any vector spaces V and W, we have a canonical isomorphism

$$T: V \otimes W \longrightarrow W \otimes V$$

determined by the values $T(A \otimes B) = B \otimes A$.

It is left as an exercise to verify that T is an isomorphism. This isomorphism is never used as an identification, since it is not suitable for computation.

2.10. Theorem. For any vector spaces V, W, Z, we have

$$(V \otimes W) \otimes Z = V \otimes (W \otimes Z) \quad .$$

Proof. Clearly, the identification should be based on a canonical linear transformation $\mu: (V \otimes W) \otimes Z \longrightarrow V \otimes (W \otimes Z)$ such that $\mu((A \otimes B) \otimes C) = A \otimes (B \otimes C)$, $A \in V$, $B \in W$, $C \in Z$, together with the inverse $\nu: V \otimes (W \otimes Z) \longrightarrow (V \otimes W) \otimes Z$ such that $\nu(A \otimes (B \otimes C)) = (A \otimes B) \otimes C$. However, since the symbols

$A \otimes B$ do not represent all classes in $V \otimes W$, for example, we cannot apply Corollary 2.5 directly, but must use the method of proof of Theorem 2.4. The correspondence $\mu_1 : V \times W \times Z \longrightarrow V \otimes (W \otimes Z)$ defined by $\mu_1(A, B, C) = A \otimes (B \otimes C)$ is trilinear, because of the laws of computation in $V \otimes (W \otimes Z)$. For fixed $C \in Z$, the induced function is bilinear, and therefore determines a function $(V \otimes W) \times Z \longrightarrow V \otimes (W \otimes Z)$. This function is bilinear and therefore determines a linear transformation $\mu : (V \otimes W) \otimes Z \longrightarrow V \otimes (W \otimes Z)$ having the desired properties. In the same way we verify that ν is well-defined and that $\nu\mu = I_{(V \otimes W) \otimes Z}$ and $\mu\nu = I_{V \otimes (W \otimes Z)}$.

2.11. <u>Definition</u>. Let V, V', W, W' be vector spaces and let $T \in L(V, W)$, $T' \in L(V', W')$. Then we define $T \otimes T' \in L(V \otimes V', W \otimes W')$ by the values

$$T \otimes T'(A \otimes A') = TA \otimes T'A' , \qquad A \in V,\ A' \in V' .$$

If $V' = W' = Z$, say, and $T' = I_Z$, we shall write T_* for $T \otimes I_Z$.

2.12. <u>Theorem</u>. If $V = V_1 \oplus V_2$, and Z is any vector space, then

$$V_1 \otimes Z \oplus V_2 \otimes Z = V \otimes Z . \tag{4}$$

Analogously,

$$Z \otimes V_1 \oplus Z \otimes V_2 = Z \otimes V . \tag{5}$$

<u>Proof</u>. Since $V = V_1 \oplus V_2$, we have (Proposition II,

12.5) injective linear transformations $\iota_j: V_j \longrightarrow V$, $j = 1, 2$, and surjective linear transformations $P_k: V \longrightarrow V_k$, $k = 1, 2$, such that $P_k \iota_j$ is the identity transformation on V_j if $k = j$ and the zero transformation if $k \neq j$. Let

$$\iota_{j*}: \quad V_j \otimes Z \longrightarrow V \otimes Z , \qquad j = 1, 2 ,$$

$$P_{k*}: \quad V \otimes Z \longrightarrow V_k \otimes Z , \qquad k = 1, 2 .$$

It is easily verified that $P_{k*}\iota_{j*}$ is the identity on $V_j \otimes Z$ if $k = j$ (in particular, ι_{j*} is injective, by Exercise II, 3.2) and is the zero transformation if $k \neq j$.

2.13. <u>Corollary</u>. Let U be a linear subspace of a vector space V and let $\iota: U \longrightarrow V$ be the canonical inclusion. Then $\iota_*: U \otimes Z \longrightarrow V \otimes Z$ is injective, and $U \otimes Z$ may be identified with a subspace of $V \otimes Z$.

<u>Proof</u>. By Theorem II, 11.11, we may write $V = U \oplus W$ for some choice of $W \subset V$. Then ι corresponds to ι_1 in the proof of Theorem 2.11, so ι_* is injective.

2.14. <u>Remarks</u>. In the axioms of I, §1, defining a vector space V, the scalars were assumed to be the real numbers, that is, the elements of the field R, and $V = R$ itself satisfies the axioms (but has additional properties, too). If R is only supposed to be a commutative ring with unit element, the same axioms define an <u>R-module</u> V.

The axioms for the scalars R to form a <u>ring</u> (with unit element 1) are exactly the ones obtained from those of I, §1, by

taking $V = R$, "multiplication by a scalar" being called simply "multiplication". In general, it is not assumed that $xy = yx$ for $x, y \in R$. However, if this is true, then R is called a <u>commutative ring</u>. A commutative ring with unit element differs from a field in that, in general, division is not possible. The most familiar example of a ring is the set of all integers with the usual addition and multiplication; however, this is a rather well-behaved ring, since multiplication is commutative and cancellation, if not division, is possible.

A large part of the definitions and constructions which have been given for vector spaces are equally valid for R-modules. However, those results which depend upon division by scalars may fail (in any case the original method of proof fails). Thus, for example, although the definition (I, 9.1) of a dependent set of elements of V retains its meaning (with "real numbers $x_1, \ldots, x_k$" replaced by "$x_1, \ldots, x_k \in R$"), Proposition I, 9.2 no longer holds. In particular, an R-module need not have a basis. Those which do are called <u>free</u> R-modules. It is still true that <u>if</u> a finite basis exists for an R-module V, then all other bases have the same number of elements, so $\dim V$ is well-defined in this case. Yet, if U is a submodule of V (the analogue of a linear subspace) and both U and V are assumed to be finite dimensional, it need not be true that a basis for U can be completed to a basis for V.

All the definitions and propositions of II, §11 concerning direct sum decompositions are equally valid if V is

assumed to be an R-module and U, W submodules, with the exception of the existence statement of Theorem II, 11.11. Given $U \subset V$, it is not always possible to find a $W \subset V$ such that $V = U \oplus W$, even if U and V are free R-modules. (A sufficient condition is that the R-module V/U be free.)

Again, the tensor product (with respect to R) of two R-modules may be defined as for vector spaces, but Corollary 2.13 does not hold in general, that is, without additional assumptions on the R-modules involved (for example, that U is a direct summand of V, or that Z is a free R-module).

2.15. Lemma (compare Lemma II, 12.10). Let V, W, Z be vector spaces, let $T \in L(V, W)$, and let $T_*: V \otimes Z \longrightarrow W \otimes Z$. Then

(i) $$\operatorname{im} T_* = \operatorname{im} T \otimes Z ,$$

(ii) $$\ker T_* = \ker T \otimes Z .$$

Proof. (i) Let $\iota: \operatorname{im} T \longrightarrow W$ be the canonical inclusion. Then it is easily checked that $\operatorname{im} T_* = \iota_*(\operatorname{im} T \otimes Z) \subset W \otimes Z$. Note that $\iota_*(\operatorname{im} T \otimes Z) = \operatorname{im} \iota_*$ may also be described as the linear subspace of $W \otimes Z$ generated by the elements of the form $B \otimes C$, $B \in \operatorname{im} T$, $C \in Z$. By Corollary 2.13, we can identify $\operatorname{im} T \otimes Z$ with $\iota_*(\operatorname{im} T \otimes Z)$. (ii) Let $\tilde{\iota}: \ker T \longrightarrow V$ be the canonical inclusion. Then it is clear that $\tilde{\iota}_*(\ker T \otimes Z) \subset \ker T_* \subset V \otimes Z$. To prove that $\tilde{\iota}_*(\ker T \otimes Z) = \ker T_*$, we proceed as follows. By Corollary 2.13, we may write

$\ker T \otimes Z$ rather than $\tilde{\iota}_*(\ker T \otimes Z)$ — this step is not essential to the proof — and consider the quotient space $V \otimes Z/\ker T \otimes Z$. Then $\tilde{j}: V \otimes Z \longrightarrow V \otimes Z/\ker T \otimes Z$ is linear and $\ker \tilde{j} = \ker T \otimes Z$. Since T_* is zero on $\ker T \otimes Z \subset \ker T_*$, there is a unique surjective linear transformation $\tilde{T}_*: V \otimes Z/\ker T \otimes Z \longrightarrow \operatorname{im} T_*$ such that $T_* = \tilde{T}_* \tilde{j}$ (Proposition II, 11.3). We shall construct a linear transformation $\tilde{S}: \operatorname{im} T_* \longrightarrow V \otimes Z/\ker T \otimes Z$ such that $\tilde{S}\tilde{T}_*$ is the identity. This implies that $\ker \tilde{T}_* = \tilde{0}$, and therefore $\ker T_* \subset \ker \tilde{j} = \ker T \otimes Z$.

<u>Construction</u>. By Corollary 2.13 we have $\operatorname{im} T_* = \operatorname{im} T \otimes Z$. This is essential, since we actually define $\tilde{S}$ on $\operatorname{im} T \otimes Z$. (Thus, if we are dealing with R-modules rather than vector spaces, an additional assumption is required unless $\operatorname{im} T = W$.)

Now let $S: \operatorname{im} T \longrightarrow V$ be any function such that $TS = I_{\operatorname{im} T}$. By Corollary II, 12.9, it is possible to choose S to be linear, but the non-linearity of an arbitrary S is limited in its effect. For $B, B' \in \operatorname{im} T$, $r \in R$, we have

$$(6) \qquad \begin{aligned} S(B + B') &= S(B) + S(B') + A , \\ S(rB) &= rS(B) + A' , \end{aligned}$$

where $A, A' \in \ker T$. In fact, since T is linear,

$$\begin{aligned} T[S(B + B') - S(B) - S(B')] &= TS(B + B') - TS(B) - TS(B') \\ &= B + B' - B - B' = \tilde{0} ; \end{aligned}$$

that is, $S(B + B') - S(B) - S(B') \in \ker T$. Similarly,

$$T[S(rB) - rS(B)] = TS(rB) - rTS(B) = rB - rB = \tilde{0} \quad .$$

Further, ST is not the identity (unless $\ker T = \tilde{0}$) but, for $A \in V$, we obviously have $STA - A \in \ker T$.

Next, define $S_1 : \operatorname{im} T \times Z \longrightarrow V \otimes Z/\ker T \otimes Z$ by $S_1 = \tilde{j}j\, S$ (where $j : F(V, W) \longrightarrow V \otimes Z$); that is, $S_1(B, C) = \tilde{j}(SB \otimes C)$, $B \in \operatorname{im} T$, $C \in Z$. Then S_1 is bilinear. For fixed $B \in \operatorname{im} T$, the computation uses the laws of computation in $V \otimes Z$ and the fact that $\tilde{j}$ is linear. For fixed $C \in Z$, we need also (6) and the fact that $\ker \tilde{j} = \ker T \otimes Z$. Since S_1 is bilinear, it follows that there is a unique linear transformation $\tilde{S} : \operatorname{im} T \otimes Z \longrightarrow V \otimes Z/\ker T \otimes Z$ with $\tilde{S}(B \otimes C) = S_1(B, C)$.

Now let $\tilde{E} = \tilde{j}(A \otimes C)$, $A \in V$, $C \in Z$. Since $T_* = \tilde{T}_*\tilde{j}$, we have

$$\begin{aligned} \tilde{S}\,\tilde{T}_*\tilde{E} = \tilde{S}T_*(A \otimes C) = \tilde{S}(TA \otimes C) = S_1(TA, C) = \tilde{j}j(STA, C) \\ = \tilde{j}(STA \otimes C) = \tilde{j}(A \otimes C) = \tilde{E} \quad . \end{aligned}$$

Since an arbitrary element of $V \otimes Z/\ker T \otimes Z$ is a linear combination of elements of the type considered above, we conclude that $\tilde{S}\,\tilde{T}_*$ is the identity.

2.16. <u>Corollary</u> (compare Corollary II, 12.11). If T of Lemma 2.15 is surjective, then T_* is surjective; if T is injective, then T_* is injective. If T is an isomorphism, then T_* is an isomorphism.

2.17. <u>Theorem</u> (compare Theorem II, 12.12). Let U, V, W, Z be vector spaces and let

$$\tilde{0} \longrightarrow U \xrightarrow{S} V \xrightarrow{T} W \longrightarrow \tilde{0}$$

be an exact sequence. Then the induced sequence

$$\tilde{0} \longrightarrow U \otimes Z \xrightarrow{S_*} V \otimes Z \xrightarrow{T_*} W \otimes Z \longrightarrow \tilde{0}$$

is also exact.

Proof. The only part not covered by Corollary 2.16 is exactness at $V \otimes Z$, i.e. $\operatorname{im} S_* = \ker T_*$. Since $\operatorname{im} S = \ker T$ by hypothesis, this follows from Lemma 2.15.

2.18. Proposition. Let V and W be finite dimensional and let $A_1, \ldots, A_k$ be a basis for V, and $B_1, \ldots, B_n$ a basis for W. Then the elements $A_j \otimes B_i$, $j = 1, \ldots, k$, $i = 1, \ldots, n$, form a basis for $V \otimes W$. In particular,

$$\dim V \otimes W = \dim V \cdot \dim W \quad . \tag{7}$$

Proof. If $k = n = 1$, it is clear that every element of $V \otimes W$ can be expressed in the form $r(A_1 \otimes B_1)$, $r \in R$, so $\dim V \otimes W$ is 0 or 1. By (3), the surjective function $V \otimes W \longrightarrow R$ defined by $r(A_1 \otimes B_1) \longrightarrow r$ is linear, so that $\dim V \otimes W = 1$. In the general case, let $V_j = L(A_j) \subset V$, $W_i = L(B_i) \subset W$. Then

$$V = V_1 \oplus \ldots \oplus V_k \ , \qquad W = W_1 \oplus \ldots \oplus W_n \ ,$$

and

$$V \otimes W = \Sigma_{i=1}^{n} \ \Sigma_{j=1}^{k} \ V_j \otimes W_i$$

by repeated applications of Theorem 2.12. The conclusion then follows from the fact that the element $A_j \otimes B_i$ is a basis for $V_j \otimes W_i$.

2.19. Proposition. If V and W are finite dimensional vector spaces, then

$$(V \otimes W)^* = V^* \otimes W^* .$$

Proof. Let the canonical linear transformation

$$\mu: V^* \otimes W^* \longrightarrow (V \otimes W)^*$$

be defined as follows: for $\omega \in V^*$, $\varphi \in W^*$, $\mu(\omega \otimes \varphi)$ is the linear form on $V \otimes W$ determined by the values

$$(8) \quad < A \otimes B, \mu(\omega \otimes \varphi) > = < A, \omega > < B, \varphi > , \qquad A \in V, B \in W .$$

It is easily verified that (8) defines a linear form and that μ is linear. To see that μ is an isomorphism, we choose bases $A_1, \ldots, A_k$ and $B_1, \ldots, B_n$ for V and W and let $\omega^1, \ldots, \omega^k$ and $\varphi^1, \ldots, \varphi^n$ be the dual bases for V^* and W^*. By Proposition 2.18, the elements $A_j \otimes B_i$ form a basis for $V \otimes W$. We then verify that the forms $\mu(\omega^\ell \otimes \varphi^m)$ coincide with the basis for $(V \otimes W)^*$ which is dual to the basis $A_j \otimes B_i$ for $V \otimes W$.

2.20. Theorem. Let $V_1, V_2, \ldots, V_n, U_1, U_2$ be vector spaces and let

$$\rho_i: V_1 \times V_2 \times \ldots \times V_n \longrightarrow U_i , \qquad i = 1, 2 ,$$

be multilinear. Suppose further that, for every choice of a

vector space Z, each multilinear transformation

$$S: V_1 \times V_2 \times \ldots \times V_n \longrightarrow Z$$

can be written uniquely in the form $S = T_i \rho_i$, where $T_i \in L(U_i, Z)$, $i = 1, 2$; that is, each couple (U_i, ρ_i) has the <u>universal factorization property</u> for $V_1 \times V_2 \times \ldots \times V_n$. Then U_1 and U_2 are isomorphic.

The proof is the same as for Theorem 2.7.

2.21. <u>Definition</u>. The <u>tensor product</u> of a finite number of vector spaces $V_1, V_2, \ldots, V_n$ is a vector space U together with a multilinear transformation $\tau: V_1 \times V_2 \times \ldots \times V_n \longrightarrow U$ such that the couple (U, τ) has the universal factorization property for $V_1 \times V_2 \times \ldots \times V_n$. We write $V_1 \otimes V_2 \otimes \ldots \otimes V_n$ for U and $A_1 \otimes A_2 \otimes \ldots \otimes A_n$ for $\tau(A_1, A_2, \ldots, A_n)$, $A_k \in V_k$, $k = 1, \ldots, n$.

2.22. <u>Theorem</u>. The tensor product of any finite number n of vector spaces, $n \geqq 2$, exists.

For example, let $n = 3$. The argument used in proving Theorem 2.10 (with $S: V_1 \times V_2 \times V_3 \longrightarrow Z$ in place of $\mu_1: V \times W \times Z \longrightarrow V \otimes (W \otimes Z)$) shows that $(V_1 \otimes V_2) \otimes V_3$ has the desired property. Alternatively, $V_1 \otimes V_2 \otimes V_3$ can be constructed directly as a quotient space of the vector space generated by $V_1 \times V_2 \times V_3$.

§3. <u>Exercises</u>

1. Let $V = R^n$ and define $\omega^1, \ldots, \omega^n \in V^*$ by

$$< (a_1,\ldots,a_n), \omega^i > = a_i - a_{i+1} , \qquad i \leq i < n ,$$
$$< (a_1,\ldots,a_n), \omega^n > = a_n .$$

Show that these linear forms give a basis for V^*. For what basis for R^n is this basis the dual basis?

2. Let V be a finite dimensional vector space and let $A_1,\ldots,A_n$ be a basis for V. Let $\omega^1,\ldots,\omega^n$ be the corresponding dual basis for V^*. Show that the basis for V^{**} which is dual to this basis for V^* will be identified with the given basis for V under the canonical identification of Proposition 1.4.

3. Let V, W, Z be vector spaces. Show that the function

$$L(W, Z) \times L(V, W) \longrightarrow L(V, Z)$$

defined by $(S, T) \longrightarrow ST$ is bilinear.

4. Let V and W be vector spaces, where V is finite dimensional. Let $T \in L(V, W; R)$ be such that

(i) $A \in V$ and $T(A, B) = 0$ for all $B \in W$ implies $A = \tilde{0}$;

(ii) $B \in W$ and $T(A, B) = 0$ for all $A \in V$ implies $B = \tilde{0}$.

Show that T, considered as an element of $L(W, L(V, R))$, gives an isomorphism of W with V^*.

5. Show that, for any vector space V,

$$V \otimes \tilde{0} = \tilde{0} .$$

6. Show that an arbitrary element of $V \otimes W$ can be expressed in the form $\Sigma_{\alpha=1}^{n} A_\alpha \otimes B_\alpha$ for some choice of n and

elements $A_\alpha \in V$, $B_\alpha \in W$, $\alpha = 1,\ldots,n$.

7. Give a proof of Theorem 2.12 by defining explicitly a canonical linear transformation

$$\mu: V_1 \otimes Z \oplus V_2 \otimes Z \longrightarrow V \otimes Z ,$$

where $V = V_1 \oplus V_2$, and showing that μ is injective and surjective.

8. In analogy with Definition 2.1, construct a quotient space of $F(V, W)$ which is isomorphic to $V \oplus W$, and demonstrate the isomorphism.

9. In analogy with Theorem 2.7, show that the vector space $(R^D)_o$ generated by a set D is characterized to within isomorphism by the universal factorization property proved in Proposition 1.7.

10. Let W be the vector space of all real-valued functions of $(x, y) \in R^2$; let V_x denote the linear subspace of W composed of functions which are independent of y, and V_y the linear subspace composed of functions which are independent of x. Let

$$T: V_x \otimes V_y \longrightarrow W$$

be the linear transformation which sends $f \otimes g \in V_x \otimes V_y$ into $fg \in W$. Prove that T is injective.

11. Let Z denote the ring of (rational) integers, Z_2 the Z-module of the integers modulo 2, and $2Z$ the submodule of Z composed of the even integers. Show that $Z \otimes Z_2 = Z_2$

and that the image of $2Z \otimes Z_2$ in $Z \otimes Z_2$ is zero.

12. Let Z denote the ring of (rational) integers. Given a positive integer $n \geq 2$, construct n Z-modules $V_1, V_2, \ldots, V_n$ such that $V_1 \otimes V_2 \otimes \ldots \otimes V_n = \tilde{0}$.

§4. Graded vector spaces

4.1. Definition. A graded vector space V is a sequence $V = \{V_0, V_1, V_2, \ldots\}$ where V_n is a vector space, $n = 0, 1, 2, \ldots$. The index n of V_n is called the degree of V_n. If $V_n = \tilde{0}$ for $n \neq p$, the graded vector space V will be said to be concentrated in degree p.

Remarks. A single vector space V may always be considered as a graded vector space concentrated in some particular degree. All definitions and constructions for graded vector spaces give the usual definitions, etc. for ordinary vector spaces by specializing to the case of graded vector spaces concentrated in degree 0, but additional possibilities arise even in the case of a single vector space if it is considered as a graded vector space concentrated in some degree other than 0. However, the vector space R will always be considered as a graded vector space concentrated in degree 0, that is, $R = \{R, \tilde{0}, \tilde{0}, \ldots\}$.

In some applications, it is convenient to use "dimension" rather than "degree" (see Chapter XII). However, the index n of V_n has no relation to $\dim V_n$ (but refers to the dimension of some other vector space with which V_n is associated).

We shall also use the notation $(V)_n = V_n$.

4.2. Definition. A graded vector space U is a graded linear subspace of the graded vector space V if U_n is a linear subspace of V_n for each n. The graded vector space V/U is defined by $(V/U)_n = V_n/U_n$. The graded vector space V^* is defined by $(V^*)_n = V_n^*$. The graded vector space $V \oplus W$ is defined by

$$(V \oplus W)_n = V_n \oplus W_n \quad .$$

The graded vector space $V \otimes W$ is defined by

$$(V \otimes W)_n = \Sigma_{i+j=n} V_i \otimes W_j \ ,$$

that is, $(V \otimes W)_0 = V_0 \otimes W_0$, $(V \otimes W)_1 = V_0 \otimes W_1 \oplus V_1 \otimes W_0$, etc.

Remark. Note that, if V is concentrated in degree p and if W is concentrated in degree q, then $V \otimes W$ is concentrated in degree $p + q$.

4.3. Definition. A linear transformation of degree r, $T : V \longrightarrow W$, of a graded vector space V into a graded vector space W is a sequence of linear transformations $T_n: V_n \longrightarrow W_{n+r}$, defined for $n \geq \max(0, -r)$. If the degree of a linear transformation is not mentioned, this degree is assumed to be zero.

Remarks. We do not require that T_n be defined for all $n = 0, 1, 2, \ldots$ if $r < 0$. This would require extending the graded vector space W, usually by setting $(W)_{n+r} = \tilde{0}$ if $n + r < 0$. In applications, the chief reason for such an extension

can be eliminated by a different convention, entirely compatible with Definition 4.1, namely, we define $\ker T$ by

$$(\ker T)_n = \begin{cases} V_n , & \text{for } n + r < 0 , \\ \ker T_n , & \text{for } n + r \geqq 0 . \end{cases}$$

Note that $\ker T$ is a graded linear subspace of V (where $T : V \longrightarrow W$). Analogously, to allow for the case $r > 0$, we define $\operatorname{im} T$ by

$$(\operatorname{im} T)_n = \begin{cases} \tilde{0} , & \text{for } n - r < 0 , \\ \operatorname{im} T_{n-r} , & \text{for } n - r \geqq 0 . \end{cases}$$

Then $\operatorname{im} T$ is a graded linear subspace of W.

The set $L_r(V, W)$, consisting of the linear transformations $T : V \longrightarrow W$ of degree r, is an ordinary vector space, with the obvious definitions of addition and multiplication by a scalar. These vector spaces do not give a graded vector space in the sense of Definition 4.1 unless we discard the linear transformations of negative degree.

4.4. <u>Definition</u>. A linear transformation T will be called <u>injective</u> if it is of degree 0 and if $(T)_n$ is injective for each n; T will be called <u>surjective</u> if T is of degree 0 and if $(T)_n$ is surjective for each n; T will be called an <u>isomorphism</u> if T is of degree 0 and if $(T)_n$ is an isomorphism for each n.

If $T : V \longrightarrow W$ is an isomorphism, it is clear that T^{-1}, defined by $(T^{-1})_n = T_n^{-1}$, is an isomorphism of W with

V. The inclusion $\iota: U \longrightarrow V$ of a graded linear subspace U of V in V, defined by $(\iota)_n = \iota_n: U_n \longrightarrow V_n$, is injective, and $j: V \longrightarrow V/U$, defined by $(j)_n = j_n: V_n \longrightarrow V_n/U_n$, is surjective.

4.5. <u>Theorem</u>. Let V, W, Z be graded vector spaces. Then

(1) $$R \otimes V = V = V \otimes R,$$

(2) $$(V \otimes W) \otimes Z = V \otimes (W \otimes Z),$$

(3) $$V \otimes Z \oplus W \otimes Z = (V \oplus W) \otimes Z,$$

(4) $$Z \otimes V \oplus Z \otimes W = Z \otimes (V \oplus W).$$

The isomorphisms on which the above identifications are based are constructed in each degree from those given in Theorems 2.8, 2.10, and 2.12.

4.6. <u>Theorem</u>. For any graded vector spaces V and W, the graded vector spaces $V \otimes W$ and $W \otimes V$ are canonically isomorphic.

<u>Remark</u>. The canonical isomorphism generally used in this case is not the obvious one induced from Theorem 2.9 by setting $T_{p+q}(A \otimes B) = B \otimes A$ for $A \in V_p$, $B \in W_q$, but is defined by $T_{p+q}(A \otimes B) = (-1)^{pq} B \otimes A$ (which agrees with Theorem 2.9 in degree 0). This choice is dictated by applications which are outside the scope of this book.

§5. Graded algebras

The axioms for an algebra were given in Chapter II, §9, when we considered the algebra of endomorphisms of a given vector space. We recall that an algebra (with unit) is first of all a vector space, that is, a set V on which are defined two operations, called addition and multiplication by a scalar, so that Axioms 1 - 8 (Chapter I, §1) are satisfied. However, there is defined a further operation called multiplication, which assigns to each pair A, B of elements of V an element AB in V, called their product, so that Axioms 9 - 12 (Chapter II, §9) are satisfied. Axiom 9 states that the multiplication is associative, that is

$$A(BC) = (AB)C , \qquad \text{for all } A, B, C \in V .$$

Axiom 12 states that there exists a (unique) unit element $I \in V$ such that

$$IA = AI = A , \qquad \text{for all } A \in V .$$

It is left as an exercise to show that the word "unique" in Axiom 12 is redundant. Axioms 10 and 11 can now be restated as follows: the function $V \times V \longrightarrow V$ defined by $(A, B) \longrightarrow AB$ is bilinear. Then this function can equally well be expressed as a linear transformation $V \otimes V \longrightarrow V$. Thus

5.1. Definition. An algebra is a vector space V together with a linear transformation

$$\nu: V \otimes V \longrightarrow V$$

such that the multiplication defined by $AB = \nu(A \otimes B)$ is associative and such that some element I of V is a unit for this multiplication. If $AB = BA$ for all $A, B \in V$, the algebra is called commutative (see, however, Definition 5.11).

5.2. Definition. Let V and W be algebras. A linear transformation $T: V \longrightarrow W$ is called a homomorphism (of algebras) if

$$(1) \qquad T(AB) = (TA)(TB) \qquad \text{for all } A, B \in V$$

and if TI is the unit element of W. That is, T preserves multiplication and units as well as addition and multiplication by scalars.

Remarks. A bijective homomorphism is also called an isomorphism, but it is usually clear from the context whether an isomorphism of vector spaces or of algebras is meant. It is left as an exercise to verify that, if T is an isomorphism of algebras, then T^{-1} is a homomorphism (and therefore an isomorphism).

Examples. The set of endomorphisms of a vector space form an algebra, multiplication being defined by composition of endomorphisms. If the given vector space has finite dimension n, and a fixed basis is chosen in it, then the matrix representation gives an isomorphism of this algebra with the algebra of all $n \times n$ matrices (Proposition II, 8.2 and Exercise II, 10.7).

The real numbers R give a commutative algebra in which multiplication and multiplication by a scalar coincide. The set of all polynomials of finite degree in an indeterminate x, with real coefficients, is a commutative algebra (cf. Exercise I, 11.7).

5.3. Proposition. Let V be an algebra and let D be a set of generators of the vector space V, that is, $V = L(D)$. In order to verify that (1) of Definition 5.2 holds for a linear transformation T on V, it is sufficient to verify (1) for all $A, B \in D$.

The proof of this proposition is left as an exercise.

There are two generalizations of the notion of linear subspace to the case of algebras, depending on whether we generalize the definition (I, 7.1) or the applications (Chapter II).

5.4. Definition. A linear subspace U of an algebra V is a subalgebra of V if $XY \in U$ for all $X, Y \in U$ and if $I \in U$.

5.5. Definition. A linear subspace U of an algebra V is an ideal in V if $AX \in U$ and $XA \in U$ for all $X \in U$ and all $A \in V$.

Remark. An ideal in V is not a subalgebra of V in general: an ideal which contains the unit element of V must be all of V.

5.6. Proposition. If $T: V \longrightarrow W$ is a homomorphism, then $\operatorname{im} T$ is a subalgebra of W and $\ker T$ is an ideal in V.

The proof of this proposition is left as an exercise.

5.7. Proposition. If U is an ideal in an algebra V,

then a unique multiplication can be defined on the quotient V/U of the vector spaces V and U so that V/U is an algebra and $j: V \longrightarrow V/U$ is a homomorphism.

Proof (cf. Definition II, 11.1). If j is a homomorphism, then we must have

$$j(AB) = j(A)j(B) \qquad \text{for all } A, B \in V .$$

If U is an ideal in V, multiplication in V/U is well-defined by this condition since, for $X, Y \in U$, we have

$$j((A + X)(B + Y)) = j(AB + XB + AY + XY) = j(AB) ,$$

using the fact that $XB + AY + XY \in U = \ker j$. It is left as an exercise to verify that the multiplication defined in V/U is associative and that $j(I)$ is a unit element in V/U.

5.8. Definition. The ideal generated by a subset S of an algebra V is the ideal $L(D)$, where D is the subset of V composed of all elements of the form AXB, for $X \in S$, $A, B \in V$.

It is left as an exercise to verify that $L(D)$ is an ideal and is also the smallest ideal of V containing S (cf. Remark following Proposition I, 7.6).

5.9. Definition. A graded algebra is a graded vector space V together with a linear transformation

$$\nu: V \otimes V \longrightarrow V$$

such that the multiplication defined by ν is associative and such that some element of V is a unit for this multiplication.

Remarks. The multiplication in V is defined explicitly as follows. Let $A \in V_p$, $B \in V_q$. Then $A \otimes B \in V_p \otimes V_q \subset (V \otimes V)_{p+q}$ and $AB = \nu_{p+q}(A \otimes B) \in V_{p+q} = (V)_{p+q}$. Thus the product of an element of degree p by an element of degree q is an element of degree $p + q$. Consequently, the vector space V_n is not an algebra in the sense of Definition 5.1 except in the case $n = 0$, since the product of two elements in V_n is an element in V_{2n}. Clearly, the unit element I of V must lie in V_0.

5.10. Proposition. A graded algebra V contains a graded linear subspace isomorphic to R.

Proof. If I is the unit element of V, we can define an injective linear transformation

$$\mu: \ R \longrightarrow V$$

by $(\mu)_0(r) = rI$, $(\mu)_n = \tilde{0}$, $n > 0$.

The linear transformation μ will be taken as an identification, so we have $R \subset V$, and we write 1 for the unit element $I \in (V)_0$. This identification causes no difficulty in computation since

$$rA = r(IA) = (rI)A$$

for all $A \in V_n$, $n = 0, 1, \ldots$.

Example. An arbitrary graded vector space V can be identified with an ideal in a graded algebra Z. The graded

algebra Z is constructed as follows: $Z = R \oplus V$, and the product of an element of R and an element of V is defined by the scalar multiplication in V, the product of two elements of V being defined to be $\tilde{0}$.

5.11. Definition. A graded algebra V is called commutative if

$$AB = (-1)^{pq}BA , \qquad A \in V_p , B \in V_q .$$

Remark. As usual, this definition reduces to the ordinary one in the case that V is concentrated in degree 0, but includes skew-symmetry if V is concentrated in degree 1, etc. The example above is commutative.

5.12. Definition. Let V and W be graded algebras. A linear transformation $T : V \longrightarrow W$ is called a homomorphism if T preserves multiplication and units, that is, if

$$T_{p+q}(AB) = (T_pA)(T_qB) , \qquad A \in V_p, B \in V_q ,$$

and if $T_o(1) = 1$.

Remark. It is clear that a homomorphism T must induce the identity transformation from $R \subset V$ to $R \subset W$.

5.13. Definition. A graded linear subspace U of a graded algebra V is a graded ideal in V if, for each $p = 0, 1, \ldots,$ and $X \in U_p$, the elements AX and XA, for all $A \in V_n$, are in U_{p+n}, $n = 0, 1, \ldots$.

5.14. Proposition. If $T : V \longrightarrow W$ is a homomor-

phism, then $\ker T$ is a graded ideal in V.

5.15. Proposition. If U is a graded ideal in a graded algebra V, then a unique multiplication can be defined on the graded vector space V / U so that V / U is a graded algebra (with unit) and $j: V \longrightarrow V / U$ is a homomorphism.

The proofs of these two propositions are left as exercises.

§6. The graded tensor algebra

6.1. Theorem. Let V be a graded vector space and let U_1, U_2 be graded algebras, and let

$$\rho_i: V \longrightarrow U_i, \qquad i = 1, 2,$$

be linear transformations. Suppose further that, for every graded algebra Z, each linear transformation

$$S : V \longrightarrow Z$$

can be written uniquely in the form $S = T_i \rho_i$, where $T_i : U_i \longrightarrow Z$ is a homomorphism, $i = 1, 2$. That is, each couple (U_i, ρ_i) has the universal factorization property for V. Then U_1 and U_2 are isomorphic.

The proof is the same as for Theorem 2.7.

6.2. Definition. The graded tensor algebra of a graded vector space V is a graded algebra U together with a linear transformation $\rho: V \longrightarrow U$ such that the couple (U, ρ) has the universal factorization property for V. We write

$\otimes V$ for U .

6.3. Proposition. The linear transformation ρ of Definition 6.2 is injective.

Proof. Take Z to be the graded algebra constructed in the Example in §5 on the vector space $R \oplus V$, and take S to be the inclusion of V in Z . By hypothesis, we have $S = T\rho$. Then for each $A \in V_n$, we have

$$A = (T\rho)_n A = T_n \rho_n A \quad ,$$

and this implies that ρ_n is injective, $n = 0, 1, \ldots$.

6.4. Theorem. If V is concentrated in degree 1, that is $V = \{\tilde{0}, V, \tilde{0}, \tilde{0}, \ldots\}$, then

$$\otimes V = \{R, V, V \otimes V, V \otimes V \otimes V, \ldots\}$$

or

$$(\otimes V)_n = \overset{n}{\otimes} V \ ,$$

where $\overset{0}{\otimes} V = R$, $\overset{1}{\otimes} V = V$, and $\overset{n}{\otimes} V$, for $n > 1$, denotes the tensor product $V \otimes V \otimes \ldots \otimes V$ (n factors), with

$$\iota : V \longrightarrow \otimes V$$

defined by $(\iota)_1 = \iota_1 : V \longrightarrow (\otimes V)_1$, $(\iota)_n = \tilde{0}$, $n \neq 1$, and with multiplication in $\otimes V$ defined by

$$(1) \quad (A_1 \otimes \ldots \otimes A_p)(B_1 \otimes \ldots \otimes B_q) = A_1 \otimes \ldots \otimes A_p \otimes B_1 \otimes \ldots \otimes B_q$$

for $A_1, \ldots, A_p, B_1, \ldots, B_q \in V$.

Proof. The formula (1) does determine a multiplication

in $\otimes V$, since these values determine a bilinear transformation

$$(\overset{p}{\otimes} V) \times (\overset{q}{\otimes} V) \longrightarrow \overset{p+q}{\otimes} V$$

and therefore determine a unique linear transformation

$$(\overset{p}{\otimes} V) \otimes (\overset{q}{\otimes} V) \longrightarrow \overset{p+q}{\otimes} V \quad .$$

The set of all these linear transformations gives a linear transformation of graded vector spaces

$$(\otimes V) \otimes (\otimes V) \longrightarrow \otimes V \quad .$$

It is obvious that this multiplication is associative.

Now let Z be any graded algebra and suppose that $S : V \longrightarrow Z$ is linear (then $S_n = \tilde{0}$ unless $n = 1$). Define

(2) $$T_n : \overset{n}{\otimes} V \longrightarrow Z_n$$

by the values

(3) $$T_n(A_1 \otimes A_2 \otimes \ldots \otimes A_n) = (S_1 A_1)(S_1 A_2) \ldots (S_1 A_n) \ , \quad n > 0 \ ,$$

and set $T_o(r) = r \in Z_o$ for $n = 0$. Then $T : \otimes V \longrightarrow Z$ is a homomorphism, determined uniquely by S, and $S = T\iota$.

6.5. <u>Theorem</u>. If V is concentrated in degree $p > 0$, with $(V)_p = V$, then

$$(\otimes V)_o = R \ , \quad (\otimes V)_{np} = \overset{n}{\otimes} V \ , \quad (\otimes V)_m = \tilde{0} \text{ otherwise.}$$

This is proved in the same way as Theorem 6.4. In

factoring $s : V \longrightarrow Z$ we now have $S_m = \tilde{0}$ unless $m = p$, and we define

$$(2') \qquad T_{np} : \overset{n}{\otimes} V \longrightarrow Z_{np}$$

by the values

$$(3') \quad T_{np}(A_1 \otimes A_2 \otimes \ldots \otimes A_n) = (S_p A_1)(S_p A_2)\ldots(S_p A_n) , \qquad n > 0 .$$

Note that these formulas check as to degrees and that the elements of V are identified with elements of degree p in $\otimes V$, and that these elements are a set of generators for $\otimes V$, that is, the smallest subalgebra of $\otimes V$ which contains the elements $\iota(V)$ is all of $\otimes V$.

It is clear that, if we try to construct the graded tensor algebra of a graded vector space concentrated in degree 0, all our products of elements from $\iota(V)$ will accumulate in degree 0. To express this result we must first define the direct sum of an infinite number of ordinary vector spaces.

6.6. <u>Definition</u>. Let $W_0, W_1, \ldots$ be vector spaces. The <u>direct sum</u>

$$\Sigma_{i=0}^{\infty} W_i = W_0 \oplus W_1 \oplus \ldots \oplus W_n \oplus \ldots$$

is the vector space whose elements are sequences $(B_0, B_1, \ldots, B_n, \ldots)$ where $B_n \in W_n$, but $B_n \neq \tilde{0}$ for at most a finite number of values of n. Addition is defined by

$$(B_0, B_1, \ldots, B_n, \ldots) + (B_0', B_1', \ldots, B_n', \ldots)$$
$$= (B_0 + B_0', B_1 + B_1', \ldots, B_n + B_n', \ldots)$$

and scalar multiplication by

$$r(B_o, B_1, \ldots, B_n, \ldots) = (rB_o, rB_1, \ldots, rB_n, \ldots) \quad .$$

<u>Remarks</u>. The above definition is a generalization of Definition II, 11.8 for the direct sum of two vector spaces. As before, $\Sigma_{i=o}^{\infty} W_i$ contains a linear subspace isomorphic to W_n, for each n. This linear subspace is composed of the elements $(B_o, B_1, \ldots)$ with $B_m = \tilde{0}$ for $m \neq n$. If we identify W_n with this subspace of $\Sigma_{i=o}^{\infty} W_i$, then any $B \in \Sigma_{i=o}^{\infty} W_i$ can be expressed as a finite sum

$$B = \Sigma_{i=1}^{k} B_i \ , \qquad\qquad B_i \in W_i \ ,$$

for some choice of k, with $B = \tilde{0}$ if and only if $B_i = \tilde{0}$ for each i.

6.7. <u>Theorem</u>. If V is concentrated in degree o with $(V)_o = V$, then $\otimes V$ is concentrated in degree o and

$$(\otimes V)_o = R \oplus V \oplus V \otimes V \oplus V \otimes V \otimes V \oplus \ldots \quad .$$

6.8. <u>Theorem</u>. If V is any graded vector space, then

$$\otimes V = R \oplus V \oplus V \otimes V \oplus V \otimes V \otimes V \oplus \ldots \quad ,$$

that is,

$$(\otimes V)_n = (R)_n \oplus (V)_n \oplus (V \otimes V)_n \oplus (V \otimes V \otimes V)_n \oplus \ldots \ .$$

In all cases multiplication in $\otimes V$ is defined by the analogue of (1), that is,

$$(A_1 \otimes \ldots \otimes A_p)(B_1 \otimes \ldots \otimes B_q) = A_1 \otimes \ldots \otimes A_p \otimes B_1 \otimes \ldots \otimes B_q$$

where, however, if $A_i \in V_{m_i}$ and $B_j \in V_{n_j}$, this formula is now the product of an element of $(\otimes V)_m$, $m = \Sigma_{i=1}^{p} m_i$, and an element of $(\otimes V)_n$, $n = \Sigma_{j=1}^{q} n_j$, and gives an element in $(\otimes V)_{m+n}$.

Remarks. The result of Theorem 6.7 gives the ordinary tensor algebra of a vector space V (not graded). The formula of Theorem 6.8 includes those of Theorems 6.4, 6.5, and 6.7 by specializing V.

6.9. Theorem. Let V and W be graded vector spaces. Any linear transformation

$$S : V \longrightarrow W$$

may be extended uniquely to a homomorphism

$$\otimes S : \otimes V \longrightarrow \otimes W .$$

Proof. Take $Z = \otimes W$. Since W is identified with a graded linear subspace of $Z = \otimes W$, we may consider $S : V \longrightarrow W$ as a linear transformation $S : V \longrightarrow Z$, and there is a unique homomorphism $T : \otimes V \longrightarrow Z = \otimes W$. We write $T = \otimes S$.

Remark. This extension is computed by

$$(\otimes S)_n(A_1 \otimes \ldots \otimes A_p) = (S A_1) \otimes \ldots \otimes (S A_p) ,$$

where n is the sum of the degrees of the elements $A_i \in V$

so that $A_1 \otimes \ldots \otimes A_p \in (\otimes V)_n$.

§7. The commutative algebra

7.1. Definition. The commutative algebra of a graded vector space V is the commutative graded algebra $Q(V)$ obtained as the quotient of the graded tensor algebra $\otimes V$ by the graded ideal K in $\otimes V$ generated by all elements of the form

$$(1) \qquad AB - (-1)^{pq}BA , \qquad A \in V_p , B \in V_q .$$

Remarks. It follows from Proposition 5.15 that $Q(V)$ is a graded algebra and that

$$j: \otimes V \longrightarrow Q(V)$$

is a homomorphism. It remains only to verify that the multiplication in $Q(V)$ is commutative, that is,

$$(2) \qquad j(A)j(B) = (-1)^{mn}j(B)j(A) , \qquad A \in (\otimes V)_m , B \in (\otimes V)_n .$$

Now any $A \in (\otimes V)_m$ is a linear combination of terms, each of which is an element of V_m or a product of two or more terms of lower degree such that the sum of the degrees of the factors is m. It is clearly sufficient to verify (2) for such products. For example, let $A = A_1 \otimes A_2 = A_1A_2$, $A_1 \in V_{m_1}$, $A_2 \in V_{m_2}$, $m_1 + m_2 = m$, and let $B \in V_n$. Then

$$\begin{aligned} AB - (-1)^{mn}BA &= (A_1A_2)B - (-1)^{mn}B(A_1A_2) \\ &= A_1(A_2B) - (-1)^{nm_2}A_1(BA_2) + (-1)^{nm_2}(A_1B)A_2 \\ &\quad - (-1)^{nm_2}(-1)^{nm_1}(BA_1)A_2 + (-1)^{nm}(BA_1)A_2 - (-1)^{mn}B(A_1A_2) \\ &= A_1[A_2B - (-1)^{m_2n}BA_2] + (-1)^{nm_2}[A_1B - (-1)^{m_1n}BA_1]A_2 . \end{aligned}$$

That is, $AB - (-1)^{mn}BA \in K = \ker j$. Therefore, when we apply j, which is a homomorphism, we obtain (2).

The above method of proof is essentially equivalent to showing that, if (2) is satisfied for the generators of $Q(V)$, then (2) holds for arbitrary elements of $Q(V)$. Here $Q(V)$ is generated by the elements of $j(V)$; for these elements (2) is an immediate consequence of (1).

7.2. <u>Theorem</u>. Let V be a graded vector space and let

$$\kappa : V \longrightarrow Q(V) \tag{3}$$

be defined by $\kappa = j\iota$, where $\iota : V \longrightarrow \otimes V$ and $j : \otimes V \longrightarrow Q(V)$. Then the couple $(Q(V), \kappa)$ has the <u>universal factorization property</u> for V, that is, for any commutative graded algebra Z, any linear transformation

$$S : V \longrightarrow Z$$

can be factored uniquely in the form $S = T\kappa$, where $T : Q(V) \longrightarrow Z$ is a homomorphism.

<u>Proof</u>. The linear transformation S can be factored through $\otimes V$ to give $S = \tilde{T}\iota$ where $\tilde{T} : \otimes V \longrightarrow Z$ is a homomorphism. Since $\tilde{T}$ is zero on $K = \ker j$, there is a unique $T : \otimes V/K \longrightarrow Z$ such that $\tilde{T} = Tj$ (Exercise 8.4). Then $T\kappa = Tj\iota = \tilde{T}\iota = S$. [To see that $\tilde{T}$ is zero on K, note that for $A \in V_p$, $B \in V_q$, we have

$$\tilde{T}(AB - (-1)^{pq}BA) = (SA)(SB) - (-1)^{pq}(SB)(SA) = \tilde{0}$$

since Z is commutative. Any element C of K is a linear combination of terms, each of which has at least one element of the form (1) as factor, so $\tilde{T}\, C = \tilde{\tilde{0}}$ also, since $\tilde{T}$ is a homomorphism.]

7.3. Corollary. The linear transformation $\kappa: V \longrightarrow Q(V)$ is injective.

This is proved in the same way as Proposition 6.3 — the particular choice of Z used there is a commutative algebra.

Remark. Since κ is injective, we can identify V with a graded linear subspace of $Q(V)$. In effect, this means that we omit the symbol j, not only for elements of $Q(V)$ now identified with an element from V, but in general. The use of the symbol A_1A_2 to denote the equivalence class $j(A_1A_2)$ causes no difficulty in computation in $Q(V)$ because any variations in the symbol used give zero in computation, using the commutativity rule (2) which now becomes

$$(4) \qquad AB = (-1)^{mn}BA\,, \qquad A \in (Q(V))_m\,,\ B \in (Q(V))_n\ .$$

7.4. Theorem. Let V be a graded vector space and suppose that the couple (U, ρ), where U is a commutative graded algebra and

$$\rho: V \longrightarrow U$$

is linear, has the universal factorization property for V (relative to commutative graded algebras). Then U is isomorphic to the commutative graded algebra $Q(V)$.

This is proved in the same way as Theorem 2.7.

7.5. Theorem. Let V and W be graded vector spaces. Any linear transformation

$$s : V \longrightarrow W$$

may be extended uniquely to a homomorphism

$$Q(s): Q(V) \longrightarrow Q(W) \quad .$$

This is proved in the same way as Theorem 6.9.

7.6. Theorem. For any graded vector spaces V and W, we have

$$Q(V \oplus W) = Q(V) \otimes Q(W) ,$$

that is,

$$(Q(V \oplus W))_n = \Sigma_{p+q=n} (Q(V))_p \otimes (Q(W))_q ,$$

where $Q(V) \otimes Q(W)$ is the commutative graded algebra constructed on the graded vector space $Q(V) \otimes Q(W)$ by defining

$$(5) \qquad (A_1 \otimes B_1)(A_2 \otimes B_2) = (-1)^{q_1 p_2} A_1 A_2 \otimes B_1 B_2$$

for $A_i \in (Q(V))_{p_i}$, $B_i \in (Q(W))_{q_i}$, $i = 1, 2$.

Proof. Let the linear transformation

$$\rho: V \oplus W \longrightarrow Q(V) \otimes Q(W)$$

be defined by

$$\rho(A + B) = A \otimes 1 + 1 \otimes B$$

for $A \in V_m$; $B \in W_m$. Then it is sufficient to show that the couple $(Q(V) \otimes Q(W), \rho)$ has the universal factorization property for $V \oplus W$. Let

$$S : V \oplus W \longrightarrow Z$$

be given. Since V is a graded linear subspace of $V \oplus W$, the given S induces a linear transformation $V \longrightarrow Z$ which can be factored uniquely through $Q(V)$ to give a homomorphism $T': Q(V) \longrightarrow Z$ such that $T'(A) = S(A)$, $A \in V$. Similarly, there is a unique $T'': Q(W) \longrightarrow Z$ such that $T''(B) = S(B)$, $B \in W$. Let $T : Q(V) \otimes Q(W) \longrightarrow Z$ be defined by the values

$$T(A \otimes B) = (T'A)(T''B) \tag{6}$$

for $A \in (Q(V))_p$, $B \in (Q(W))_q$. Then T is a homomorphism, uniquely determined by S, and

$$\begin{aligned} T\rho(A + B) &= T(A \otimes 1 + 1 \otimes B) \\ &= T'(A) + T''(B) = S(A + B) \quad . \end{aligned}$$

7.7. <u>Corollary</u>. For each $n = 0, 1, \ldots,$ the canonical isomorphism of Theorem 7.6 induces a direct sum decomposition of the vector space $(Q(V \oplus W))_n$ into $n + 1$ summands:

$$\begin{aligned} (Q(V \oplus W))_n = (Q(V \oplus W))_{n,0} &\oplus (Q(V \oplus W))_{n-1,1} \\ &\oplus \ldots \oplus (Q(V \oplus W))_{0,n} \end{aligned} \tag{7}$$

where $(Q(V \oplus W))_{p,q}$ is the linear subspace isomorphic to

$(Q(V))_p \otimes (Q(W))_q$.

Remarks. The isomorphism of $Q(V) \otimes Q(W)$ with $Q(V \oplus W)$ can be constructed directly by factoring $\iota: V \oplus W \longrightarrow Q(V \oplus W)$ through $Q(V) \otimes Q(W)$. Then it is clear from (5), for the case $S = \iota$, that $(Q(V \oplus W))_{p,q}$ may be described as the linear subspace of $(Q(V \oplus W))_{p+q}$ generated by elements of the form AB with $A \in (Q(V))_p$, $B \in (Q(W))_q$. The elements of $(Q(V \oplus W))_{p,q}$ will be said to be of type (p, q).

7.8. Definition. If V is concentrated in degree 1, $(V)_1 = V$, then $Q(V)$ will be called the graded exterior algebra of V. We write $\wedge V$ for $Q(V)$ and $\wedge^p V$ for $(Q(V))_p$ and denote the multiplication by an explicit symbol $\wedge$, that is, we write $A \wedge B$ rather than AB, and refer to the multiplication as the exterior product.

Remarks. If V is concentrated in any odd degree, the corresponding $Q(V)$ may be called an exterior algebra. The characteristic property in all cases is

$$A \wedge A = \tilde{0}, \qquad \text{for all } A \in V .$$

This follows from the fact that, if the degree of A is odd, then $A \wedge A = - A \wedge A$, by (4), and the only vector which equals its negative is $\tilde{0}$.

7.9. Definition. If V is concentrated in degree p, where p is even, then $Q(V)$ will be called a graded symmetric algebra.

Remark. In this case, $(\otimes V)_n$, and therefore $(Q(V))_n$, is $\tilde{0}$ if n is odd. Consequently, for the multiplication defined in $Q(V)$, we have

$$(4') \qquad AB = BA, \qquad A, B \in Q(V) .$$

7.10. Lemma. Let V be concentrated in degree p and suppose that $\dim V = 1$, where $V = (V)_p$. (i) If p is odd, then

$$Q(V) = R \oplus V ;$$

if A is a basis for V, then A generates $Q(V)$ and the elements $1, A$ give a "basis" for $Q(V)$. (ii) If p is even, and A is a basis for V, then the element A generates $Q(V)$ and the elements $1, A, A^2, \ldots$ give a "basis" for $Q(V)$. (If we write x rather than A, the resulting symmetric algebra is called the "polynomial algebra with one generator x of degree p".)

The above statements follow directly from the fact that an arbitrary element of V can be expressed as rA, $r \in R$, together with the fact that

$$(rA)(sA) = rsAA = rsA^2 .$$

(Note: in the case that p is even, the element A^n has degree np.)

7.11. Theorem. Let V be a finite dimensional vector space, $\dim V = n$. Then

$$\wedge^p V = \tilde{0} \ , \qquad p > n \ ,$$

and

$$\dim \wedge^p V = \binom{n}{p} \ , \qquad 0 \leqq p \leqq n \ .$$

Moreover, if $A_1, \ldots, A_n$ is a basis for V, then the elements 1, A_i, $i = 1, \ldots, n, \ldots$, $A_{i_1} \wedge \ldots \wedge A_{i_p}$, $i_1 < \ldots < i_p, \ldots, A_1 \wedge \ldots \wedge A_n$ give a basis for $\wedge V$.

<u>Proof</u>. For any basis $A_1, \ldots, A_n$ for V, let

$$V = V_1 \oplus \ldots \oplus V_n \ ,$$

where $V_i = L(A_i)$. We then apply Theorem 7.6 repeatedly to obtain

$$\wedge V = \wedge V_1 \otimes \wedge V_2 \otimes \ldots \otimes \wedge V_n \tag{8}$$

where, by Lemma 7.10, $\wedge V_i = \{R, V_i, \tilde{0}, \tilde{0}, \ldots\}$. If we write out (8) in each degree, we find that $\wedge^p V$ is a direct sum of vector spaces, each of which is the tensor product of n vector spaces, one from each graded vector space $\wedge V_i$, so that the sum of the degrees of these vector spaces is p. Now any factor is $\tilde{0}$ unless its degree is 0 or 1. If $p > n$, it is clear that every direct summand in the direct sum expressing $\wedge^p V$ has $\tilde{0}$ for at least one factor, so $\wedge^p V = \tilde{0}$. If $p \leqq n$, there are exactly $\binom{n}{p}$ direct summands involving p factors of degree 1 and $n - p$ factors R (of degree 0). These summands are each of dimension 1, by Proposition 2.18. Any other direct summands are $\tilde{0}$. Thus, for $p \leqq n$, $\wedge^p V$ is isomorphic to a direct sum of $\binom{n}{p}$ vector spaces, each of dimension 1. Thus

$$\dim \Lambda^p V = \binom{n}{p} \quad .$$

If we consider the induced direct sum decomposition of $\Lambda^p V$ into 1-dimensional subspaces, as in Corollary 7.7, then a basis for $\Lambda^p V$ is obtained by selecting a non-zero element from each such subspace. The standard selection is given in the statement of the theorem. Note, however, that we could just as well choose $A_2 \wedge A_1$ in place of $A_1 \wedge A_2 = - A_2 \wedge A_1$, etc.

§8. Exercises

1. Let V be a graded vector space. Show that $(V^*)_n = L_{-n}(V, R)$.

2. Show that an element I which satisfies

$$AI = A = IA$$

for all $A \in V$ is uniquely determined by this property.

3. Let V and W be algebras and let $T: V \longrightarrow W$ be a homomorphism. Show that $\ker T$ is an ideal in V.

4. State and prove the analogue of Proposition II, 11.3 for algebras and homomorphisms rather than for vector spaces and linear transformations.

5. Show that the multiplication determined by (5) of §7 satisfies the conditions for $Q(V) \otimes Q(W)$ to be a commutative graded algebra.

6. Let V and Z be vector spaces. A multilinear transformation $T \in L(V,\ldots,V; Z)$ on the product $V \times \ldots \times V$

(p factors, $p \geq 2$) is called alternating if $T(A_1,\ldots,A_p) = \tilde{0}$ whenever $A_k = A_{k+1}$ for some $k = 1,\ldots,p$. Show that an equivalent condition is

$$T(A_1,\ldots,A_p) = -\,T(A_1,\ldots,A_{j-1},\ A_{j+1},\ A_j,\ A_{j+2},\ldots,A_p)\ ,$$

for each $j = 1,\ldots,p$. Show that any alternating linear transformation on $V \times \ldots \times V$ (p factors) can be factored through $\wedge^p V$.

7. Let $T\colon V \longrightarrow W$ and $S\colon W \longrightarrow Z$ be linear. Show that $\wedge(ST) = (\wedge S)(\wedge T)$, where $\wedge T$ denotes the extension of T (Theorem 7.5) to the graded exterior algebras of V and W, etc.

§9. The exterior algebra of a finite dimensional vector space

In this section we shall study further the graded exterior algebra $\wedge V$ of a vector space V in the case that V is finite dimensional. We begin by summarizing for this case the results obtained previously.

The computing rules for the multiplication (the exterior product) in $\wedge V$ are:

(1) $$(X + X') \wedge Y = X \wedge Y + X' \wedge Y\ ,$$

(2) $$X \wedge (Y + Y') = X \wedge Y + X \wedge Y'\ ,$$

(3) $$rX \wedge Y = r(X \wedge Y) = X \wedge rY\ ,$$

(4) $$X \wedge Y = (-1)^{pq} Y \wedge X\ ,$$

where $X, X' \in \wedge^p V$, $Y, Y' \in \wedge^q V$, $r \in R$. In particular,

$$(4') \qquad X \wedge X = \tilde{0}$$

if $X \in \wedge^p V$ where p is odd. The first three rules reflect properties of the graded tensor algebra of V which pass to the quotient algebra.

If V is finite dimensional, say $\dim V = n$, then $\wedge^p V$ is a finite dimensional vector space for each $p = 0, 1, \ldots,$ with

$$\wedge^p V = \tilde{0}\ , \qquad p > n\ ,$$

$$\dim \wedge^p V = \binom{n}{p}\ , \qquad 0 \leqq p \leqq n\ .$$

If $A_1, \ldots, A_n$ is any basis for $V = \wedge^1 V$, then the elements

$$A_{i_1} \wedge A_{i_2} \wedge \ldots \wedge A_{i_p}\ , \qquad i_1 < i_2 < \ldots < i_p\ ,$$

form a basis for $\wedge^p V$, $p = 2, \ldots, n$, while 1 is a basis for $\wedge^0 V = R$.

If $T: V \longrightarrow V$ is linear, then there is a unique extension $\wedge T: \wedge V \longrightarrow \wedge V$, by Theorem 7.5. Now $\wedge^n V$ has dimension 1, and an endomorphism of a 1-dimensional vector space must be of the form $X \longrightarrow \lambda X$ for some $\lambda \in R$, that is,

$$\wedge^n T(X) = \lambda X\ , \qquad \text{for all } X \in \wedge^n V\ .$$

9.1. <u>Definition</u>. If V is a vector space of dimension n and T is an endomorphism of V, then the <u>determinant</u> of the

endomorphism T is the scalar $\det T$ such that

(5) $$\wedge^n T(X) = (\det T)X , \qquad \text{for all } X \in \wedge^n V .$$

9.2. <u>Theorem</u> (cf. Proposition V, 2.3 and Theorem V, 2.4). The determinant has the following properties:

(i) $$\det I_V = 1 ,$$

(ii) $$\det (ST) = (\det S)(\det T) , \qquad S, T \in E(V) ,$$

(iii) for $T \in E(V)$, we have $\det T = 0$ if and only if T is singular, i.e. if and only if $\ker T \neq \tilde{0}$,

(iv) if T is non-singular, that is, if $T \in A(V)$, then

$$\det T^{-1} = 1/\det T .$$

<u>Proof</u>. (i) and (ii) follow from the definition of $\wedge I_V$ and from $\wedge(ST) = (\wedge S)(\wedge T)$ (Exercise 8.7). To prove (iii), we note that the value of $\det T$ may be computed as follows. If $A_1,\ldots,A_n$ is a basis for V, then the element $X = A_1 \wedge \ldots \wedge A_n \neq \tilde{0}$ is a basis for $\wedge^n V$. Moreover, $\wedge^n T(A_1 \wedge \ldots \wedge A_n) = TA_1 \wedge \ldots \wedge TA_n$ by the definition of $\wedge T$. Thus,

(6) $$TA_1 \wedge \ldots \wedge TA_n = (\det T)A_1 \wedge \ldots \wedge A_n .$$

If $\ker T \neq \tilde{0}$, let $A_1 \neq \tilde{0}$ be an element of $\ker T$ and choose a basis for V which includes this element. Then (6) becomes

$$\tilde{0} = (\det T)A_1 \wedge \ldots \wedge A_n ,$$

which implies $\det T = 0$. If T is non-singular, then T^{-1} exists. For $S = T^{-1}$, (i) and (ii) give $(\det T^{-1})(\det T) = 1$, which implies $\det T \neq 0$, and also (iv).

Remark. The particular definition of $\det T$ given in Chapter V, for the case $\dim V = 3$, agrees with Definition 9.1. The procedure used there was to define an alternating trilinear function $V \times V \times V \longrightarrow R$ by $(A, B, C) \longrightarrow [A, B, C]$, where the function depended on a choice of scalar and vector products in V, and then to define $\det T$ by

$$[TA, TB, TC] = (\det T)[A, B, C] .$$

By Exercise 8.6, this alternating trilinear function induces a linear function $\Lambda^3 V \longrightarrow R$ which we shall denote by $X \longrightarrow [X]$, $X \in \Lambda^3 V$, determined by the values $[A \wedge B \wedge C] = [A, B, C]$. Then (5) gives

$$[\Lambda^3 T(X)] = [(\det T)X] = (\det T)[X] ,$$

from which it follows that Definition V, 2.2 gives the same value of $\det T$ as Definition 9.1.

If $T: V \longrightarrow W$ is linear, where $\dim V = \dim W = n$, then we again have that $\Lambda^n T: \Lambda^n V \longrightarrow \Lambda^n W$ is a linear transformation from a vector space of dimension 1 into a vector space of dimension 1, so $\Lambda^n T$ is either the zero transformation or is an isomorphism. The argument used in proving (iii) of Proposition 9.2 can be used in this case to show that $\Lambda^n T$ is the zero transformation if and only if T is singular, but it is not

possible to assign a scalar "det T" canonically.

9.3. Definition. If $T: V \longrightarrow W$ is linear, and $A_1, \ldots, A_n$ is a basis for V and $B_1, \ldots, B_n$ a basis for W, then the determinant of the matrix representation of T with respect to this choice of bases is the scalar λ determined by

$$(7) \qquad TA_1 \wedge \ldots \wedge TA_n = \lambda(B_1 \wedge \ldots \wedge B_n) .$$

Remarks. The value λ is well-defined, since the element $B_1 \wedge \ldots \wedge B_n$ is a basis for $\wedge^n W$. The fact that λ coincides with the usual notion of the determinant of a matrix follows directly from the computing rules for the exterior product. For example, if $n = 2$, we have

$$\begin{aligned} TA_1 \wedge TA_2 &= (a_1{}^1 B_1 + a_1{}^2 B_2) \wedge (a_2{}^1 B_1 + a_2{}^2 B_2) \\ &= a_1{}^1 a_2{}^1 B_1 \wedge B_1 + a_1{}^2 a_2{}^1 B_2 \wedge B_1 + a_1{}^1 a_2{}^2 B_1 \wedge B_2 + a_1{}^2 a_2{}^2 B_2 \wedge B_2 \\ &= (a_1{}^1 a_2{}^2 - a_1{}^2 a_2{}^1) B_1 \wedge B_2 , \end{aligned}$$

since $B_1 \wedge B_1 = B_2 \wedge B_2 = \tilde{0}$ and $B_2 \wedge B_1 = -B_1 \wedge B_2$.

The notion of the determinant of a (square) matrix can be defined by (7), since every matrix determines a linear transformation. Then all the usual properties of determinants follow from various properties of the exterior product. However, we shall regard the algebraic properties of determinants as known, and the determinants themselves as a convenient device for computing the value of certain scalars connected with the exterior algebra.

9.4. Proposition. If $T: V \longrightarrow W$ is linear, where

dim V = dim W, then T is an isomorphism if and only if the determinant of the matrix representation of T, relative to any choice of bases for V and W, is different from zero.

9.5. Proposition. A set $A_1,\ldots,A_p$ of vectors in V is independent if and only if $A_1 \wedge \ldots \wedge A_p \neq \tilde{0}$.

Proof. If the given set of vectors is dependent, then one vector, say A_1, can be expressed as a linear combination $\Sigma_{i=2}^{p} a^i A_i$ of the remaining vectors (Theorem I, 9.2). Then

$$A_1 \wedge A_2 \wedge \ldots \wedge A_p = (\Sigma_{i=2}^{p} a^i A_i) \wedge A_2 \wedge \ldots \wedge A_p = \tilde{0}$$

since any exterior product with a repeated factor of odd degree must vanish (why?). If the vectors $A_1,\ldots,A_p$ are independent, they can be included in a basis for V. Then $A_1 \wedge \ldots \wedge A_p$ is included in the corresponding standard basis for $\wedge^p V$, so cannot be zero.

9.6. Proposition. Let $T: V \longrightarrow W$ be linear, and let r be the largest integer for which $\wedge^r T$ is not zero. Then

$$r = \dim \operatorname{im} T = \dim V - \dim \ker T .$$

Proof. As in the proof of Theorem II, 2.5, a basis for V may be chosen so that $A_1,\ldots,A_p$ is a basis for $\ker T \subset V$ and $TA_{p+1},\ldots,TA_n$ is a basis for $\operatorname{im} T \subset W$. Then $\wedge^q T$ is zero for $q > n - p$, since it is clear that, for $q > n - p$, every basis element for $\wedge^q V$ includes at least one factor from ker T. This implies $r \leqq n - p$. But $\wedge^{n-p} T(A_{p+1} \wedge \ldots \wedge A_n) = TA_{p+1} \wedge \ldots \wedge TA_n \neq \tilde{0}$, so $\wedge^{n-p} T$ is not zero and $r = n - p$.

Remarks. Note that this result includes the statement made in the case $\dim V = \dim W$. If $r = \dim V = \dim W$, then T is an isomorphism. More generally, if $r = \dim V$, then T is injective; if $r = \dim W$, then T is surjective; if $r = 0$, then T is the zero transformation.

The value r is called the rank of T. This value coincides with the usual rank of the matrix, for any matrix representation of T. In fact, the entries in the matrix representation of $\wedge^q T$ are the minors of order q of the matrix representing T. To see this, let $A_1,\ldots,A_n$ be a basis for V, and $B_1,\ldots,B_m$ a basis for W, and let

$$(8). \qquad TA_i = \Sigma_{j=1}^m a_i{}^j B_j , \qquad i = 1,\ldots,n .$$

Then, if $i_1 < \ldots < i_q$,

$$\wedge^q T(A_{i_1} \wedge \ldots \wedge A_{i_q}) = TA_{i_1} \wedge \ldots \wedge TA_{i_q}$$
$$= (\Sigma_{j=1}^m a_{i_1}{}^j B_j) \wedge \ldots \wedge (\Sigma_{j=1}^m a_{i_q}{}^j B_j) .$$

When this is expanded in the same way as (7), terms with a repeated factor B_j vanish, while the coefficient of a basis element $B_{j_1} \wedge \ldots \wedge B_{j_q}$, $j_1 < \ldots < j_q$, of $\wedge^q W$ (after collecting terms with the same factors in degree 1) is exactly the determinant of the $q \times q$ submatrix whose entries are the coefficients $a_{i_\ell}{}^{j_k}$. If we denote the determinant of this submatrix by $a_{i_1 \ldots i_q}{}^{j_1 \ldots j_q}$, then the matrix representation of $\wedge^q T$ is

$$T(A_{i_1}\wedge \ldots \wedge A_{i_q}) = \Sigma_{j_1<\ldots<j_q}\, a_{i_1\ldots i_q}{}^{j_1\ldots j_q}\, B_{j_1}\wedge \ldots \wedge B_{j_q}\,, \qquad i_1 < \ldots < i_q\,.$$

In this notation, $a_{1\ldots n}{}^{1\ldots n} = \det T$ if $m = n$.

9.7. <u>Theorem</u>. If V is a finite dimensional vector space, then

$$\wedge^q V^* = (\wedge^q V)^*\,, \qquad q = 0, 1, \ldots\,.$$

<u>Proof</u>. As in the proof of Proposition 2.19, we define a canonical linear transformation

$$\mu\colon \wedge^q V^* \longrightarrow (\wedge^q V)^*\,,$$

this time by the values

$$(9) \qquad \langle A_1\wedge \ldots \wedge A_q,\ \mu(\omega^1\wedge \ldots \wedge \omega^q)\rangle = \det(\langle A_i, \omega^j\rangle)\,,$$

for any $A_1,\ldots,A_q \in V$ and $\omega^1,\ldots,\omega^q \in V^*$. It is left as an exercise to verify that μ is well-defined. If a basis $A_1,\ldots,A_n$ is selected for V, and $\omega^1,\ldots,\omega^n$ is the dual basis for V^*, then (9) gives, for $i_1 < \ldots < i_q$, $j_1 < \ldots < j_q$,

$$\langle A_{i_1}\wedge \ldots \wedge A_{i_q},\ \mu(\omega^{j_1}\wedge \ldots \wedge \omega^{j_q})\rangle = \begin{cases} 1 & \text{if } i_1 = j_1,\ldots,i_q = j_q\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

That is, μ sends the elements of the standard basis for $\wedge^q V^*$ into the basis of $(\wedge^q V)^*$ which is dual to the standard basis

for $\wedge^q V$, and is therefore an isomorphism.

Remark. The isomorphism μ is always used as an identification, so the action of an element $\omega^1 \wedge \ldots \wedge \omega^q$ in $\wedge^q V^*$ as a linear form on $\wedge^q V$ is expressed by

$$(10) \qquad < A_1 \wedge \ldots \wedge A_q, \omega^1 \wedge \ldots \wedge \omega^q > = \det (< A_i, \omega^j >) .$$

These values determine $< X, \varphi >$ for any $X \in \wedge^q V$, $\varphi \in \wedge^q V^*$, since X and φ are linear combinations of terms of the above type. As computing rules, we have

$$< X + X', \varphi > = < X, \varphi > + < X', \varphi > ,$$

$$< X, \varphi + \varphi' > = < X, \varphi > + < X, \varphi' > ,$$

$$< rX, \varphi > = < X, \varphi > = < X, r\varphi > ,$$

where $X, X' \in \wedge^q V$, $\varphi, \varphi' \in \wedge^q V^*$, $r \in R$; in particular,

$$< A_1 \wedge A_2 \wedge A_3 \wedge \ldots \wedge A_q, \varphi > = - < A_2 \wedge A_1 \wedge A_3 \wedge \ldots \wedge A_q, \varphi > ,$$

etc.

9.8. Definition. The elements of $\wedge^q V^*$ are called q-forms, in view of their interpretation as linear forms on $\wedge^q V$. The elements of $\wedge^q V$ are called q-vectors.

By Corollary II, 8.6, any endomorphism T of V induces an endomorphism T^* of V^*, which in the present notation would be defined as follows: for $\omega \in V^*$, the linear form $T^*\omega$ is determined by the values

$$(11) \qquad < A, T^*\omega > = < TA, \omega > , \qquad \text{for all } A \in V .$$

More generally, any $T \in L(V, W)$ induces $T^* \in L(W^*, V^*)$, defined by the same formula except that $\omega \in W^*$, $T^*\omega \in V^*$. The linear transformation T^* is often called the transpose of T since, if T is represented by a matrix, relative to a given choice of bases for V and W, then the matrix representation of T^*, relative to the dual bases, is the transpose matrix.

9.9. <u>Theorem</u>. Let $T: V \longrightarrow W$ be linear. Then

$$\wedge T^* = (\wedge T)^* \quad .$$

<u>Proof</u>. Let $A_1, \ldots, A_q$ be elements of V and $\omega^1, \ldots, \omega^q$ elements of W^*. Then

$$\begin{aligned}
< A_1 \wedge \ldots \wedge A_q, \ \wedge^q T^*(\omega^1 \wedge \ldots \wedge \omega^q) > &= < A_1 \wedge \ldots \wedge A_q, \ T^*\omega^1 \wedge \ldots \wedge T^*\omega^q > \\
&= \det (< A_i, \ T^*\omega^j >) = \det (< TA_i, \ \omega^j >) \\
&= < TA_1 \wedge \ldots \wedge TA_q, \ \omega^1 \wedge \ldots \wedge \omega^q > \\
&= < \wedge^q T(A_1 \wedge \ldots \wedge A_q), \ \omega^1 \wedge \ldots \wedge \omega^q > \\
&= < A_1 \wedge \ldots \wedge A_q, \ (\wedge^q T)^*(\omega^1 \wedge \ldots \wedge \omega^q) > .
\end{aligned}$$

9.10. <u>Corollary</u>. If $V = W$, that is, if T is an endomorphism, then $\det T^* = \det T$.

We have seen, in §1, that a choice of scalar product in V is equivalent to a linear transformation $S: V \longrightarrow V^*$ such that, for $X \in V$, the linear form $SX \in V^*$ is defined by the values

$$< Y, SX > = Y \cdot X , \qquad \text{for all } Y \in V .$$

9.11. Definition. The scalar product induced in $\wedge^q V$, $q \geqq 2$, by a scalar product in V is defined by

$$Y \cdot X = < Y, (\wedge^q S)X > , \qquad X, Y \in \wedge^q V .$$

Remarks. For $A_1, \ldots, A_q, B_1, \ldots, B_q \in V$, we then have

$$(A_1 \wedge \ldots \wedge A_q) \cdot (B_1 \wedge \ldots \wedge B_q) = < A_1 \wedge \ldots \wedge A_q, SB_1 \wedge \ldots \wedge SB_q >$$

$$= \det (< A_i, SB_j >) = \det (A_i \cdot B_j) .$$

The fact that this definition gives a scalar product on $\wedge^q V$ follows from $A_i \cdot B_j = B_j \cdot A_i$ and the properties of the determinant of a symmetric matrix. (If these properties are not assumed, Axioms S4 and S5 of Definition III, 1.1 can be proved directly by computing in terms of an orthonormal basis for V).

The linear transformation $S: V \longrightarrow V^*$ which defines the scalar product on V induces $S^*: V^* \longrightarrow V^{**}$. By Exercise 10.4, the composition $S^*S: V \longrightarrow V^{**}$ coincides with the function used in Proposition 1.4 to identify V^{**} with V. After the identification, S^* is equivalent to S^{-1}. Thus,

9.12. Definition. The scalar product induced in V^* by a scalar product $S: V \longrightarrow V^*$ in V is defined by $S^{-1}: V^* \longrightarrow V$, that is,

$$\omega \cdot \varphi = < S^{-1}\omega, \varphi > = S^{-1}\omega \cdot S^{-1}\varphi , \qquad \omega, \varphi \in V^* .$$

The scalar product induced in $\wedge^q V^*$, $q \geq 2$, by a scalar product in V is defined by

$$\varphi \cdot \psi = \langle (\wedge^q S^{-1})\varphi, \psi \rangle , \qquad \varphi, \psi \in \wedge^q V^* .$$

<u>Remark</u>. For $\omega^1, \ldots, \omega^q, \varphi^1, \ldots, \varphi^q \in V^*$, we then have

$$(\omega^1 \wedge \ldots \wedge \omega^q) \cdot (\varphi^1 \wedge \ldots \wedge \varphi^q) = \det (\omega^i \cdot \varphi^j) .$$

We may also consider the vector space

$$\wedge V = \wedge^0 V \oplus \wedge^1 V \oplus \ldots \oplus \wedge^n V$$

of dimension $2^n = \Sigma_{p=0}^n \binom{n}{p}$. (With the exterior product of elements induced by the exterior product in the graded exterior algebra, we have an algebra in the sense of Definition 5.1). This point of view will enable us to consider an important transformation which is not a homomorphism of algebras nor even an endomorphism of the graded vector space $\wedge V$, but is an automorphism of the vector space $\wedge V$ defined above.

Since $\binom{n}{p} = \binom{n}{n-p}$, the vector spaces $\wedge^p V$ and $\wedge^{n-p} V$ are isomorphic, but the isomorphism cannot be given canonically in general. However, for each choice of scalar product on V, we can define a standard isomorphism

$$*: \wedge^p V \longrightarrow \wedge^{n-p} V , \qquad p = 0, 1, \ldots, n ,$$

so as to give an automorphism of the vector space $\wedge V$. Actually, there are two such automorphisms for each choice of scalar product (and one is the negative of the other just as, in Chapter IV, there

were two vector products).

A scalar product on V induces a scalar product in $\wedge^p V$, $p = 0, 1, \ldots, n$. In $\wedge^n V$, which is 1-dimensional, there are exactly two elements A such that $A \cdot A = 1$, and the choice to be made is of one of these elements.

9.13. <u>Definition</u>. Let V be a vector space of dimension n, with scalar product, and let $A \in \wedge^n V$ be chosen so that $A \cdot A = 1$. For each $p = 0, 1, \ldots, n$,

$$(12) \qquad *: \wedge^p V \longrightarrow \wedge^{n-p} V$$

is determined, for each $X \in \wedge^p V$, by

$$(13) \qquad *X \cdot Y = (X \wedge Y) \cdot A \qquad \text{for all } Y \in \wedge^{n-p} V .$$

9.14. <u>Proposition</u>. The operator $*: \wedge V \longrightarrow \wedge V$ determined by Definition 9.13 is an automorphism of $\wedge V$.

It is left as an exercise to verify that (13) defines a linear transformation (12). The fact that $*$ is an automorphism follows from the property (Exercise 10.11)

$$**X = (-1)^{np+p} X , \qquad X \in \wedge^p V .$$

Several important properties of this operator will be found in the exercises.

The transpose operator will also be denoted by $*$, that is,

$$*: \wedge V^* \longrightarrow \wedge V^* .$$

This operator may be defined directly as a transpose operator by

(14) $$< X, *\varphi > = < *X, \varphi > , \qquad \varphi \in \wedge^p V^*, X \in \wedge^{n-p} V .$$

Alternatively, we may define $\Theta \in \wedge^n V^*$ by the condition $<A, \Theta > = 1$. Then Θ is the dual element of A and $\Theta \cdot \Theta = 1$. For $\varphi \in \wedge^p V^*$, the (n-p)-form $*\varphi$ is determined by the condition

(15) $$*\varphi \cdot \psi = (\varphi \wedge \psi) \cdot \Theta \qquad \text{for all } \psi \in \wedge^{n-p} V^* .$$

§10. Exercises

1. Let $T: V \longrightarrow W$ be linear. Show that $\wedge T: \wedge V \longrightarrow \wedge W$ is surjective if T is surjective. Show that, if T is injective and V and W are finite dimensional, then $\wedge T$ is injective.

2. If V is finite dimensional, show that $\otimes V^* = (\otimes V)^*$. If $T: V \longrightarrow W$ is linear, where V and W are finite dimensional, show that $\otimes T^* = (\otimes T)^*$.

3. Let $A_1, \ldots, A_n$ be an orthonormal basis for a vector space V with scalar product (Definition III, 5.1). Then the p-vectors $A_{i_1} \wedge \ldots \wedge A_{i_p}$, $i_1 < \ldots < i_p$, form a basis for $\wedge^p V$. Show that these vectors form an orthonormal basis for $\wedge^p V$ with respect to the induced scalar product on $\wedge^p V$.

4. Let V be a finite dimensional vector space with scalar product $S: V \longrightarrow V^*$. Let $S^*: V^* \longrightarrow V^{**}$ be the transpose of S. Show that $S^*S: V \longrightarrow V^{**}$ coincides with $\mu: V \longrightarrow V^{**}$ of Proposition 1.4.

5. Let $A_1, \ldots, A_n$ be an orthonormal basis for a vector

space V with scalar product $S: V \longrightarrow V^*$. Show that the elements $SA_1, \ldots, SA_n$ form an orthonormal basis for V^* and coincide with dual basis of V^* relative to the given basis for V.

6. Let V be a finite dimensional vector space with scalar product. Show that the induced scalar product on V^* has the following property:

$$|\omega| = \max_{|X|=1} < X, \omega > .$$

7. Let V be a finite dimensional vector space with scalar product $S: V \longrightarrow V^*$. Show that the adjoint (Definition V, 5.1) of an endomorphism T of V may be identified with $S^{-1}T^*S$, where T^* is the transpose of T.

8. Let $\dim V = 2$. Show that $*X$ is orthogonal to X, for each $X \in V$, and that $*: V \longrightarrow V$ is a rotation through $\pm 90°$.

9. Let V be a vector space of dimension n, with scalar product, and let $A \in \wedge^n V$ denote the element used in Definition 9.13. Show that $*1 = A$ and that $*A = 1$.

Let $A_1, \ldots, A_n$ be an orthonormal basis for V. Then $A = \lambda A_1 \wedge \ldots \wedge A_n$ (what are the possible values of λ?). Show that, for $i_1 < \ldots < i_p$,

$$*(A_{i_1} \wedge \ldots \wedge A_{i_p}) = \lambda\varepsilon \; A_{j_1} \wedge \ldots \wedge A_{j_{n-p}}, \qquad 1 \leqq p < n ,$$

where ε is the signature of the permutation $i_1, \ldots, i_p$, $j_1, \ldots, j_{n-p}$ of the integers $1, \ldots, n$.

10. Let $X, Z \in \wedge^p V$, $Y \in \wedge^{n-p} V$, $p \leqq n$. Show that

(a) $$*X\cdot Y = (-1)^{np+p}X\cdot *Y \ ;$$

(b) $$*X\wedge Z = *Z\wedge X \ .$$

11. Using a basis, verify that

$$**X = (-1)^{np+p}X \ , \qquad X \in \wedge^p V \ .$$

12. Combining the results of Exercises 10 and 11, show that for $X, Z \in \wedge^p V$, $p \leq n$,

$$*X\cdot *Z = X\cdot Z \ .$$

13. Let $\dim V = 3$. For any $B, C \in V$, define $B \hat{\times} C = *(B\wedge C)$. Verify that $\hat{\times}$ satisfies the axioms 1, 2, 3, 4, 5' of Definition IV, 1.1. Thus, by Corollary IV, 2.6,

$$B \times C = \pm *(B\wedge C) \ .$$

Finally, let $B, C, D \in V$ and let $\mu^2 = (B\wedge C\wedge D)\cdot(B\wedge C\wedge D)$. Then $B\wedge C\wedge D = \pm \mu A$, where $A \in \wedge^3 V$ is the element used in defining $*$. Show that

$$(B\cdot *(C\wedge D))^2 = (B\cdot C \times D)^2 = \mu^2 \ .$$

Thus $|B\wedge C\wedge D|$ represents the volume of the parallelepiped whose edges are B, C, D as in Chapter IV, §2.

14. Let $X, Y \in \wedge^p V$. Then $X\wedge *Y \in \wedge^n V$, so $X\wedge *Y = \mu A$ for some μ. Show that $\mu = X\cdot Y$.

15. Let $A_1,\ldots,A_n$ be an arbitrary basis for V, and let $\omega^1,\ldots,\omega^n$ be the basis of V^* which is dual to the basis $A_1,\ldots,A_n$. Let the scalar product on V^* be determined from the

scalar product $S: V \longrightarrow V^*$ on V. Let $g_{ij} = A_i \cdot A_j$ and $g^{ij} = \omega^i \cdot \omega^j$.

(a) Show that for $B = \Sigma_{i=1}^n b^i A_i \in V$, the element $SB \in V^*$ is given by

$$SB = \Sigma_{j=1}^n \Sigma_{i=1}^n b^i g_{ij} \omega^j .$$

(b) Show that

$$\Sigma_{k=1}^n g_{ik} g^{kj} = \delta_i^j = \begin{cases} 0 & \text{if } i \neq j , \\ 1 & \text{if } i = j . \end{cases}$$

(c) Let

$$(A_1 \wedge \ldots \wedge A_n) \cdot (A_1 \wedge \ldots \wedge A_n) = g > 0 .$$

Show that g is the determinant of the matrix representation of S in terms of the above bases. If $A \in \wedge^n V$ satisfies $A \cdot A = 1$, what are the possible values of λ in the formula $A = \lambda A_1 \wedge \ldots \wedge A_n$? Evaluate $(\omega^1 \wedge \ldots \wedge \omega^n) \cdot (\omega^1 \wedge \ldots \wedge \omega^n)$.

(d) Let X, Y of Exercise 14, $1 \leq p \leq n$, be given by

$$X = \Sigma_{i_1 < \ldots < i_p} X^{i_1 \ldots i_p} A_{i_1} \wedge \ldots \wedge A_{i_p} ,$$

$$Y = \Sigma_{i_1 < \ldots < i_p} Y^{i_1 \ldots i_p} A_{i_1} \wedge \ldots \wedge A_{i_p} .$$

Verify that Y can be expressed in the form

$$Y = \frac{1}{p!} \Sigma_{k_1=1}^n \ldots \Sigma_{k_p=1}^n Y^{k_1 \ldots k_p} A_{k_1} \wedge \ldots \wedge A_{k_p}$$

if the set of coefficients $Y^{k_1 \ldots k_p}$ is extended to arbitrary sets of indices $k_1, \ldots, k_p$ by the requirement that the coefficients be alternating with respect to these indices. Show that μ (Exercise 14) is given by

$$\mu = \Sigma_{i_1 < \ldots < i_p} \Sigma_{k_1 = 1} \cdots \Sigma_{k_p = 1} X^{i_1 \ldots i_p} g_{i_1 k_1} \cdots g_{i_p k_p} Y^{k_1 \ldots k_p} .$$

(e) For $\varphi, \psi \in \Lambda^p V^*$, $1 \leqq p \leqq n$, where

$$\varphi = \Sigma_{i_1 < \ldots < i_p} \varphi_{i_1 \ldots i_p} \omega^{i_1} \wedge \ldots \wedge \omega^{i_p} ,$$

$$\psi = \Sigma_{i_1 < \ldots < i_p} \psi_{i_1 \ldots i_p} \omega^{i_1} \wedge \ldots \wedge \omega^{i_p} ,$$

verify that

$$\begin{aligned} \varphi \wedge *\psi &= \psi \wedge *\varphi \\ &= \Sigma_{i_1 < \ldots < i_p} \Sigma_{k_1 = 1} \cdots \Sigma_{k_p = 1} \pm \sqrt{g}\, \varphi_{i_1 \ldots i_p} g^{i_1 k_1} \cdots g^{i_p k_p} \psi_{k_1 \ldots k_p} \cdot \\ &\qquad \omega^1 \wedge \ldots \wedge \omega^n , \end{aligned}$$

where the correct choice of sign depends only on the original choice of $A \in \Lambda^n V$ and the choice of the basis $A_1, \ldots, A_n$. Explain how to determine the correct sign here.

(f) Show how to compute the coefficients $(*\varphi)_{j_1 \ldots j_{n-p}}$, for $\varphi \in \Lambda^p V^*$, in the expression

$$*\varphi = \Sigma_{j_1 < \ldots < j_{n-p}} (*\varphi)_{j_1 \ldots j_{n-p}} \omega^{j_1} \wedge \ldots \wedge \omega^{j_{n-p}}$$

by using (e) with particular choices of $\psi \in \Lambda^p V^*$.

16. Show that (13) and (14) of §9 are equivalent by computing $*\varphi$ in two ways, for $\varphi = \omega^{i_1} \wedge \ldots \wedge \omega^{i_p}$, where

$\omega^1, \ldots \omega^n$ is a basis for V^* which is dual to an orthonormal basis $A^1, \ldots, A^n$ for V.

17. (Continuation of Exercise 3.10.) Show that any $f \in \operatorname{im} T \subset W$ has the following property: there exists an n_o (depending on f) such that for $n > n_o$ and any sequence $a_1, \ldots, a_n$ of real numbers, the endomorphism S of R^n whose matrix representation is $(f(a_i, a_j))$ satisfies $\Lambda^n S = \tilde{\tilde{0}}$. Show that $T: V_x \otimes V_y \longrightarrow W$ is not surjective by exhibiting a function $f \in W$ which does not have the above property.

1

Introduction to Clifford Algebras

Pertti Lounesto

1.9 The exterior algebra and contractions

The exterior algebra $\bigwedge V$ of a linear space V, with basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$, has a basis consisting of

$$\begin{array}{rl} 1 & \text{scalar} \\ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n & \text{vectors} \\ \mathbf{e}_1 \wedge \mathbf{e}_2,\ \mathbf{e}_1 \wedge \mathbf{e}_3,\ \ldots,\ \mathbf{e}_{n-1} \wedge \mathbf{e}_n & \text{bivectors} \\ \vdots & \vdots \\ \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \ldots \wedge \mathbf{e}_n & \text{volume element.} \end{array}$$

The multiplication rules are

$$\mathbf{e}_i \wedge \mathbf{e}_j = -\mathbf{e}_j \wedge \mathbf{e}_i$$

together with associativity and the unity 1. A scalar product on V can be extended to the homogeneous parts $\bigwedge^k V$ by

$$<\mathbf{x}_1 \wedge \mathbf{x}_2 \wedge \ldots \wedge \mathbf{x}_k, \mathbf{y}_1 \wedge \mathbf{y}_2 \wedge \ldots \wedge \mathbf{y}_k> = \det(\mathbf{x}_i \cdot \mathbf{y}_j)$$

and further by orthogonality to all of $\bigwedge V$. This scalar valued product can be used to define the **contraction** $\bigwedge V \times \bigwedge V \to \bigwedge V$, $(u, v) \to u \lrcorner v$ by

$$<u \lrcorner v, w> = <v, \tilde{u} \wedge w> \quad \text{for all} \quad w \in \bigwedge V.$$

The contraction could also be defined by its characteristic properties

$$\begin{aligned} &\mathbf{x} \lrcorner \mathbf{y} = \mathbf{x} \cdot \mathbf{y}, \\ &\mathbf{x} \lrcorner (u \wedge v) = (\mathbf{x} \lrcorner u) \wedge v + \hat{u} \wedge (\mathbf{x} \wedge v), \\ &(u \wedge v) \lrcorner w = u \lrcorner (v \lrcorner w) \end{aligned}$$

which hold for all $\mathbf{x}, \mathbf{y} \in V$ and $u, v, w \in \bigwedge V$.

The contraction could also be introduced via the Clifford product as

$$u \lrcorner v = (u \wedge (v\mathbf{e}_{12\ldots n}))\mathbf{e}_{12\ldots n}^{-1}.$$

This can be proved by observing $<u\mathbf{e}_{12\ldots n}, w> = <\mathbf{e}_{12\ldots n}, \tilde{u} \wedge w>$ and computing

2

Mathematical Structure of Clifford Algebras

Ian Porteous

ABSTRACT The first part of this chapter is mainly concerned with the construction of Clifford algebras for real and complex nondegenerate quadratic spaces of arbitrary rank and signature, these being presented as matrix algebras over $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, ${}^2\mathbb{R}$, ${}^2\mathbb{C}$ or ${}^2\mathbb{H}$. In each case the algebra has an anti-involution known as conjugation, and the second part is concerned with determining products, or equivalently correlations, on the spinor space for which the induced adjoint anti-involution on the matrix algebra is conjugation. Applications of the classification are to the description of the Spin groups and conformal groups for quadratic spaces of low dimension.

2.1 Clifford algebras

2.1.1 Construction of the algebras $C\ell_{p,q}$

We have seen in the previous lecture how well adapted the algebra of quaternions $\mathbb{H}$ is to the study of the groups $SO(3)$ and $SO(4)$. The center of interest is a finite-dimensional vector space X over the real field $\mathbb{R}$, furnished with a quadratic form, in the one case $\mathbb{R}^3$ and in the other case $\mathbb{R}^4$. In either case the real associative algebra of quaternions $\mathbb{H}$ contains both $\mathbb{R}$ and X as linear subspaces, there being an anti-involution, namely conjugation, of the algebra, such that, for all $x \in X$,

$$\overline{x}\,x = x^{(2)} = x \cdot x.$$

In the former case, when $\mathbb{R}^3$ is identified with the subspace of pure quaternions, this formula can also be written in the simpler form

$$x^2 = -x^{(2)} = -x \cdot x.$$

In an analogous, but more elementary way, the real algebra of complex numbers $\mathbb{C}$ may be used in the study of the group $SO(2)$.

AMS Subject Classification: 15A66, 17B37, 20C30, 81R25.

Keywords: Clifford algebra, spinor, Radon–Hurwitz number, correlation, conjugation, conformal transformation, Vahlen matrix.

Our aim is to put these rather special cases into a wider context. To keep the algebra simple, the emphasis is laid at first on generalizing the second of the two displayed formulae.

Let X be a finite-dimensional real *quadratic space*, that is a finite-dimensional vector space assigned a possibly degenerate quadratic form. As follows from one of the definitions in Lecture 1, the *Clifford Algebra* for X is a real associative algebra, $C\ell(X)$, with unit element 1, containing isomorphic copies of $\mathbb{R}$ and X as linear subspaces, and generated as a real associative algebra by them in such a way that, for all $x \in X$, $x^2 = -x^{(2)}$. The dimension of the algebra is $2^{\dim X}$.

Superficially, it would seem to be simpler to arrange things so that for all $x \in X$, $x^2 = x^{(2)}$, as was the case in Lecture 1. But there is a good reason for adopting the 'negative' convention, as we shall see in a moment. To simplify notations, in practice $\mathbb{R}$ and X are identified with their copies in $C\ell(X)$.

The Clifford algebra has an anti-involution, known as *conjugation*

$$C\ell(X) \to C\ell(X);\ a \mapsto a^-$$

such that, for all $x \in X$, $x^- = -x$. Then, for all $x \in X$, $x \cdot x = x^- x$.

The vector space $\mathbb{R} \oplus X$ is called the space of *paravectors* of the Clifford algebra. It becomes a real quadratic space on being assigned the quadratic form $\lambda + x \mapsto \lambda^2 + x \cdot x = \lambda^2 - x^2 = (\lambda - x)(\lambda + x) = (\lambda + x)^-(\lambda + x)$. Here is where we get a positive payoff for adhering to the negative convention! In particular, if X is a positive-definite quadratic space of dimension n then the space of paravectors $\mathbb{R} \oplus X$ is a positive-definite quadratic space of dimension $n + 1$.

One almost invariably works with an orthonormal basis for the vector space X. We denote by $\mathbb{R}^{p,q,r}$ the real vector space $\mathbb{R}^{p+q+r}$ such that the first p vectors of the standard basis have scalar square -1, the next q have scalar square $+1$, and the final r have scalar square 0, these basis vectors being mutually orthogonal. The Clifford algebra generated by these basis vectors is then denoted by $C\ell_{p,q,r}$, the first p basis vectors having square $+1$, the next q vectors having square -1 and the final r basis vectors having square 0. In the Clifford algebra the basis vectors *anticommute*, the equation

$$0 = x \cdot y = \tfrac{1}{2}((x + y) \cdot (x + y) - x \cdot x - y \cdot y)$$

becoming in the Clifford algebra

$$0 = -xy - yx = -(x + y)^2 + x^2 + y^2.$$

The quadratic space $\mathbb{R}^{p,q,0}$ is also denoted by $\mathbb{R}^{p,q}$, the associated Clifford algebra being denoted by $C\ell_{p,q}$. In the same spirit, the quadratic space $\mathbb{R}^{0,n}$ is also denoted by $\mathbb{R}^n$, the associated Clifford algebra being denoted by $C\ell_{0,n}$ (not $C\ell_n$, to avoid confusion with the conventions of Lecture 1). A quadratic space is said to be *neutral* if it is nondegenerate, that is $r = 0$ and the number of positive squares is equal to the number of negative squares, that is $p = q$.

It is generally helpful to have an explicit construction of the Clifford algebras for quadratic spaces of arbitrary rank, signature and nullity.

Proposition 1 *Let W be a linear subspace of a finite-dimensional real quadratic space X, assigned the induced quadratic form with Clifford algebra $C\ell(X)$. Then the subalgebra of $C\ell(X)$ generated by W may be identified with $C\ell(W)$.*

By this proposition the existence of a Clifford algebra for an arbitrary n-dimensional quadratic space X is implied by the existence of a Clifford algebra for the neutral nondegenerate space $\mathbb{R}^{n,n}$. Such an algebra is constructed below.

The starting point for all our work is the observation that the three matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

in $\mathbb{R}(2)$ are mutually anticommutative and satisfy the equations

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^2 = 1, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}^2 = 1 \quad \text{and} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = -1,$$

the product of any two of them being plus or minus the third.

We select the second and third of these to represent an orthonormal set of vectors of the quadratic space $\mathbb{R}^{1,1}$, embedded as a linear subspace of the matrix algebra $\mathbb{R}(2)$ of 2×2 matrices over the field of real numbers $\mathbb{R}$. These vectors generate the algebra, the identity matrix and the three matrices together forming a basis for the algebra as a four-dimensional linear space. With these choices, the algebra represents the real Clifford algebra $C\ell_{1,1}$ of the quadratic space $\mathbb{R}^{1,1}$.

It is now very easy to construct a representation of the Clifford algebra $C\ell_{2,2}$ of the quadratic space $\mathbb{R}^{2,2}$. Let a and b be the generators of $C\ell_{1,1}$. The algebra we require is none other than $\mathbb{R}(4)$, thought of as $\mathbb{R}(2)(2)$, taking as a set of four mutually anticommuting matrices the matrices

$$\begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}, \quad \begin{pmatrix} b & 0 \\ 0 & -b \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

with 0 the 2×2 matrix $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and 1 the 2×2 matrix $\left(\begin{smallmatrix} 1 & \\ 0 & 1 \end{smallmatrix}\right)$. Clearly two of these have square -1 and two have square $+1$, and together they generate the sixteen-dimensional real algebra $\mathbb{R}(4)$. Continuing in the same way, one proves by induction that the real algebra $\mathbb{R}(2^n)$ represents the real Clifford algebra $C\ell_{n,n}$ of the real quadratic space $\mathbb{R}^{n,n}$.

Now consider the case of the quadratic space $\mathbb{R}^{p,q}$ with an orthonormal basis, the first p vectors of which have scalar square -1, so Clifford square $+1$, while the remaining q have scalar square 1, so Clifford square -1. Clearly its Clifford algebra $C\ell_{p,q}$ is representable as a subalgebra of the real algebra $C\ell_{n,n}$ where n is the larger of p and q.

We can even construct in this way Clifford algebras for degenerate quadratic spaces, where some of the generating vectors have square 0. For in any of the algebras constructed so far if one chooses one basis vector a, say of square -1, and another b, say of square $+1$, then because a and b anticommute, $a + b$ and $a - b$ both have square 0. In particular the *Grassmann algebra* of the vector space $\mathbb{R}^n$, all of whose basis vectors anticommute and have square 0, is representable as a real subalgebra of the real algebra $C\ell_{n,n} \cong \mathbb{R}(2^n)$.

The main theorem of this section is that for any p, q the Clifford algebra $C\ell_{p,q}$ of the *nondegenerate* real quadratic space $\mathbb{R}^{p,q}$ is representable as a full matrix algebra with entries in one of the real algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, ${}^2\mathbb{R}$ or ${}^2\mathbb{H}$. ${}^2\mathbb{R}$ is shorthand notation for the real algebra of diagonal 2×2 real matrices, ${}^2\mathbb{H}$ is shorthand notation for the real algebra of diagonal 2×2 quaternionic matrices, and ${}^2\mathbb{C}$ is a shorthand notation for the algebra of diagonal 2×2 complex matrices, regarded either as a real or as a complex algebra according to the context. For example $C\ell_{0,1}$ $\mathbb{C}$, with generator i, $C\ell_{0,2}$ $\cong \mathbb{H}$, with generators i and k, while $C\ell_{0,3}$ ${}^2\mathbb{H}$, with generators $\left(\begin{smallmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} \mathrm{j} & 0 \\ 0 & -\mathrm{j} \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} \mathrm{k} & 0 \\ 0 & -\mathrm{k} \end{smallmatrix}\right)$

This last algebra is to be preferred to the algebra $\mathbb{H}$ as a Clifford algebra for $\mathbb{R}^{0,3}$, since although i, j and k mutually anticommute, and each has square -1, the product ij $=$ k, and they generate only the four-dimensional real algebra $\mathbb{H}$, and not an eight-dimensional algebra. (In Porteous 1995 $\mathbb{H}$ is referred to as a *nonuniversal* Clifford algebra for $\mathbb{R}^{0,3}$.)

One then obtains $\mathbb{H}(2)$ as a representation of the Clifford algebra $\mathbb{R}_{0,4}$ by taking as the fourth generator the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, which clearly anticommutes with the previous three and also has square -1.

The main result of this section is the following theorem:

Theorem 1 *For $0 \leq p,\ q < 7$, matrix representations of the Clifford algebras $C\ell_{p,q}$ are exhibited in the following table.*

Table 1.1

$p \downarrow$ \ $q \rightarrow$	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	${}^2\mathbb{H}$	$\mathbb{H}(2)$	$\mathbb{C}(4)$	$\mathbb{R}(8)$	${}^2\mathbb{R}(8)$
	${}^2\mathbb{R}$	$\mathbb{R}(2)$	$\mathbb{C}(2)$	$\mathbb{H}(2)$	${}^2\mathbb{H}(2)$	$\mathbb{H}(4)$	$\mathbb{C}(8)$	$\mathbb{R}(16)$
	$\mathbb{R}(2)$	${}^2\mathbb{R}(2)$	$\mathbb{R}(4)$	$\mathbb{C}(4)$	$\mathbb{H}(4)$	${}^2\mathbb{H}(4)$	$\mathbb{H}(8)$	$\mathbb{C}(16)$
	$\mathbb{C}(2)$	$\mathbb{R}(4)$	${}^2\mathbb{R}(4)$	$\mathbb{R}(8)$	$\mathbb{C}(8)$	$\mathbb{H}(8)$	${}^2\mathbb{H}(8)$	$\mathbb{H}(16)$
	$\mathbb{H}(2)$	$\mathbb{C}(4)$	$\mathbb{R}(8)$	${}^2\mathbb{R}(8)$	$\mathbb{R}(16)$	$\mathbb{C}(16)$	$\mathbb{H}(16)$	${}^2\mathbb{H}(16)$
	${}^2\mathbb{H}(2)$	$\mathbb{H}(4)$	$\mathbb{C}(8)$	$\mathbb{R}(16)$	${}^2\mathbb{R}(16)$	$\mathbb{R}(32)$	$\mathbb{C}(32)$	$\mathbb{H}(32)$
	$\mathbb{H}(4)$	${}^2\mathbb{H}(4)$	$\mathbb{H}(8)$	$\mathbb{C}(16)$	$\mathbb{R}(32)$	${}^2\mathbb{R}(32)$	$\mathbb{R}(64)$	$\mathbb{C}(64)$
	$\mathbb{C}(8)$	$\mathbb{H}(8)$	${}^2\mathbb{H}(8)$	$\mathbb{H}(16)$	$\mathbb{C}(32)$	$\mathbb{R}(64)$	${}^2\mathbb{R}(64)$	$\mathbb{R}(128)$

the table extending to higher p, q with 'period' 8.

Most of this follows at once from what we have already done. For example, for any p, q, $C\ell_{p+1,q+1} \cong C\ell_{p,q}(2)$. Also, clearly, $C\ell_{1,0} \cong {}^2\mathbb{R}$. What is missing are the remarks, first that, if S is an orthonormal subset of type $(p+1,\ q)$, generating a real associative algebra A, then, for any $a \in S$ with $a^2 = 1$, the set

$$\{b\,a : b \in S \backslash \{a\}\} \cup \{a\}$$

is an orthonormal subset of type $(q+1,\ p)$ generating A. From this it follows at once that the Clifford algebras $\mathbb{R}_{p+1,q}$ and $\mathbb{R}_{q+1,p}$ are isomorphic. The second remark requires first that we introduce the concept of the *tensor product* of two real algebras.

A tensor product decomposition of a real algebra is somewhat analogous to a direct sum decomposition of a vector space, but involves the multiplicative structure rather than the additive structure. Suppose that B and C are subalgebras of a finite-dimensional algebra A over $\mathbb{K}$, the algebra being associative and with unit element, such that (i) for any $b \in B$, $c \in C$, $c\,b = b\,c$, (ii) A is generated as an algebra by B and C, and (iii) $\dim A = \dim B \dim C$.

Then we say that A is the *tensor product* $B \otimes_{\mathbb{K}} C$ over $\mathbb{K}$, the abbreviation $B \otimes C$ being used when the field $\mathbb{K}$ is not in doubt.

Let B and C be subalgebras of a finite-dimensional algebra A over $\mathbb{K}$, such that $A = B \otimes C$, the algebra A being associative and with unit element. Then $B \cap C = \mathbb{K}$ (the field $\mathbb{K}$ is identified with the set of scalar multiples of the unit element $1_{(A)}$) .

(The periodicity theorem.) *For all finite p, q,*

$$\mathbf{R}_{p,q+8} \cong \mathbf{R}_{p,q} \otimes \mathbf{R}(16).$$

For our present purposes the important facts are that

$$\mathbb{R} \otimes \mathbb{R} = \mathbb{R}, \quad \mathbb{C} \otimes \mathbb{R} = \mathbb{C}, \quad \mathbb{H} \otimes \mathbb{R} = \mathbb{H},$$
$$\mathbb{C} \otimes \mathbb{C} \cong {}^2\mathbb{C}, \quad \mathbb{H} \otimes \mathbb{C} \cong \mathbb{C}(2) \quad \mathbb{H} \otimes \mathbb{H} \cong \mathbb{R}(4).$$

The first three of these statements are obvious. The two mutually commuting copies of $\mathbb{C}$ in ${}^2\mathbb{C}$ are multiples of the identity matrix and the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. The mutually commuting copies of $\mathbb{C}$ and $\mathbb{H}$ in $\mathbb{C}(2)$ are multiples of the identity and the standard representation of quaternions as 2×2 complex matrices. The two mutually commuting copies of $\mathbb{H}$ in $\mathbb{R}(4)$ are the representations of quaternions, first by themselves multiplying other quaternions on the left, and secondly by their conjugates multiplying other quaternions on the right, the commutativity in this case following from the associativity of the quaternions. Moreover, for any m, n, $\mathbb{R}(m) \otimes \mathbb{R}(n) \cong \mathbb{R}(m\ n)$. We leave it as an exercise to work out why.

The other result that we need is that, for all p, q,

$$C\ell_{p,q+4} \cong C\ell_{p,q} \otimes C\ell_{0,4}$$

bearing in mind that $C\ell_{0,4} \cong \mathbb{H}(2)$.

To prove this, let a, b, c, d be the last four generators of $C\ell_{p,q+4}$, and let the others be e_i. Then $(abcd)^2 = 1$. Let $f_i = abcde_i$, for each i. It is easily verified that the f_i anticommute, and that $f_i^2 = e_i^2$, for all i. The f_i then generate a copy of $C\ell_{p,q}$ in $C\ell_{p,q+4}$ that commutes with the copy of $C\ell_{0,4}$, generated by a, b, c, d. Hence the result.

For example

$$\begin{array}{lclcl} C\ell_{0,5} & \cong & \mathbb{C} \otimes \mathbb{H}(2) & \cong & \mathbb{C}(4), \\ C\ell_{0,6} & \cong & \mathbb{H} \otimes \mathbb{H}(2) & \cong & \mathbb{R}(8), \\ C\ell_{0,7} & \cong & {}^2\mathbb{H} \otimes \mathbb{H}(2) & \cong & {}^2\mathbb{R}(8), \\ C\ell_{0,8} & \cong & \mathbb{H}(2) \otimes \mathbb{H}(2) & \cong & \mathbb{R}(16). \end{array}$$

and

The *negative* of a real quadratic space is that obtained by replacing the quadratic form by the same form multiplied by -1. In particular the negative of the quadratic space $\mathbb{R}^{p,q}$ is isomorphic to the quadratic space $\mathbb{R}^{q,p}$. An important, and perhaps surprising feature of Table 1.1 is that the Clifford algebras of a quadratic vector space and of its negative are quite different.

2.1.2 The even Clifford algebras $C\ell^0_{p,q}$

A map from a vector space or algebra to itself that respects the vector space or algebra structure is said to be an *involution* if its square is the identity. Such a map from an algebra to itself is an *anti-involution* if the order of products is reversed. For example, conjugation on the real algebra of quaternions $\mathbb{H}$ is an anti-involution, since, for any $q, q' \in \mathbb{H}$, $qq' = q'\bar{q}$.

Let $C\ell(X)$ be the Clifford algebra of a finite-dimensional quadratic space X. We have already mentioned *conjugation* as the unique anti-involution of $C\ell(X)$ that sends each vector to its negative. In fact, the linear involution $X \to X : x \mapsto -x$ also extends uniquely to an involution of $C\ell(X)$, known as the *main involution* of the algebra, while the identity on X extends to an anti-involution of $C\ell(X)$, known as *reversion*. The main involution will be denoted by $a \mapsto \hat{a}$, while reversion will be denoted by $a \mapsto \bar{a}$. The two anti-involutions and the involution commute with each other, each the composite of the other two.

It is not difficult to prove that the main involution leaves invariant a subalgebra of $C\ell(X)$, spanned by all even products of basis elements of X, and known as the *even Clifford algebra* of X. The even subalgebra of the Clifford algebra $C\ell_{p,q}$ will be denoted here by $C\ell^0_{p,q}$

Proposition 2 *For any finite* p, q,

$$C\ell^0_{p,q+1} \cong C\ell_{p,q}, \qquad C\ell^0_{p+1,q} \cong C\ell_{q,p}.$$

We give the proof in the particular case that $p = 0$, $q = n + 1$. Let $e_1, \ldots, e_{n+1}$ be the standard orthonormal basis for $\mathbb{R}^{n+1}$ Each has square equal to -1 in $C\ell_{0,n+1}$, and, of course, they anticommute. Then the n elements $e_1e_{n+1}, \ldots, e_ne_{n+1}$ each have square -1, and these also anticommute, while together with $\mathbb{R}$ they generate $C\ell^0_{0,n+1}$ as a real algebra, thus being an algebra isomorphic to $C\ell_{0,n}$.

Similar arguments hold in the other cases.

It follows, in particular, that the table of the even Clifford algebras $C\ell^0_{p,q}$ is the same as Table 1.1, except that there is an additional line of entries down the left-hand side matching the existing line of entries across the top row. The symmetry about the main diagonal in the table of even Clifford algebras expresses the fact that the *even* Clifford algebras of a finite-dimensional nondegenerate quadratic space, and of its negative, are mutually isomorphic.

2.1.3 Complex Clifford algebras

So far we have only considered real Clifford algebras of *real* quadratic spaces. Everything that we have done before can be carried through for the complex quadratic spaces $\mathbb{C}^n$. We denote this Clifford algebra by $C\ell_n(\mathbb{C})$. This algebra may be considered as the tensor product over $\mathbb{R}$ or $\mathbb{C}$ and any of the real Clifford algebras $C\ell_{p,q}$, where $p + q = n$.

Theorem 2 *For any finite k,*

$$\mathbb{C} \otimes C\ell_{0,2k} \cong \mathbb{C}(2^k) \text{ and } \mathbb{C} \otimes C\ell_{0,2k+1} \cong {}^2\mathbb{C}(2^k).$$

In defining the conjugation anti-involution for the algebra $\mathbb{C} \otimes C\ell_{p,q}$, we have to decide whether or not to conjugate the complex *coefficients*. It turns out, as we shall see in a later section, that if we do not conjugate the coefficients, then the only invariant is the rank n of the relevant real quadratic space, but that if we conjugate the coefficients, then signature remains important. We also will have occasion to consider briefly the tensor products of $C\ell_{p,q}$ by ${}^2\mathbb{R}^\sigma$ and of $\mathbb{C}_n$ by ${}^2\mathbb{C}^\sigma$, where σ denotes the involution that consists in *swapping* the components of ${}^2\mathbb{R}$ and ${}^2\mathbb{C}$, respectively, this swap to be invoked when conjugation is being defined.

2.1.4 Spinors

Table 1.1 exhibits each of the Clifford algebras $C\ell_{p,q}$ as the real algebra of endomorphisms of a right $\mathbb{A}$-linear space of the form $\mathbb{A}^m$, where $\mathbb{A} = \mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, ${}^2\mathbb{R}$ or ${}^2\mathbb{H}$. This space is called the *(real) spinor space* or *space of (real) spinors* of the quadratic space $\mathbb{R}^{p,q}$. It is identifiable with a minimal left ideal of the algebra, namely the space of matrices with every column except the first nonzero. However as a minimal left ideal it is nonunique.

Physicists concerned with space-time have to choose between the

Clifford algebras $C\ell_{1,3} \cong \mathbb{H}(2)$ and $C\ell_{3,1} \cong \mathbb{R}(4)$, with $C\ell^0_{1,3} \cong C\ell^0_{3,1} \cong \mathbb{C}(2)$. Roughly speaking, in the former case the spinor space $\mathbb{H}^2$ is known as the space of *Dirac spinors*, though these, as originally defined, consisted of quadruplets of complex numbers rather than pairs of quaternions, while in the latter case the spinor space $\mathbb{R}^4$ is known as the space of *Majorana spinors*. The complexification of each of these algebras is isomorphic to $\mathbb{C}(4)$. This algebra is known as the *Weyl algebra*, the complex spinor space $\mathbb{C}^4$ being known as the space of *Weyl spinors*.

Minimal left ideals of a matrix algebra are generated by *primitive idempotents*. An idempotent of an algebra is an element y such that $y^2 = y$. It is *primitive* if it cannot be expressed as the sum of two idempotents whose product is zero. The simplest example in a matrix algebra is the matrix consisting entirely of zeros, except for a single entry of 1 somewhere in the main diagonal. The minimal ideal generated by such an idempotent then consists of matrices all of whose columns except one consist of zeros. The easiest idempotents to construct are of the form $\frac{1}{2}(1 + x)$ where $x^2 = 1$, but *not* $x^2 = -1$. Of course, $\frac{1}{2}(1 - x)$ also is an idempotent, so that spinor spaces constructed in this way come naturally in pairs. However, these are not necessarily primitive. They are when the matrix algebra consists of 2×2 matrices over $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, but in the case of 4×4 matrix algebras, the primitive idempotents are products of commuting pairs of such idempotents.

Although as minimal left ideals of matrix algebras any two spinor spaces are equivalent, they may behave differently when the Clifford algebra structure of the matrix algebra is taken into account, and so may have possibly different physical interpretations. Therefore, in applications the terms Majorana, Dirac and Weyl spinors may be reserved for specific minimal left ideals of the Clifford algebra.

For example, suppose that the algebra $C\ell_{1,3} \cong \mathbb{H}(2)$ is generated by mutually anti commuting vectors γ_0, γ_1, γ_2, γ_3, where $\gamma_0^2 = 1, \gamma_1^2 = \gamma_2^2 = \gamma_3^2 = -1$. Then the spinor space generated by the primitive idempotent $\frac{1}{2}(1 + \gamma_0\gamma_1)$ consists of Dirac spinors, while the spinor space of $\mathbb{C} \otimes \mathbb{H}(2) \cong \mathbb{C}(4)$ generated by the primitive idempotent $\frac{1}{4}(1 + \gamma_0\gamma_1)(1 + i\gamma_0\gamma_1\gamma_2\gamma_3)$ consists of Weyl spinors.

Likewise, suppose that the algebra $C\ell_{3,1} \cong \mathbb{R}(4)$ is generated by mutually anti-commuting vectors e_0, e_1, e_2, e_3, where $e_0^2 = -1$, $e_1^2 = e_2^2 = e_3^2 = 1$. Then the spinor space generated by the primitive idempotent

$(1 + e_1)(1 + e_0e_2)$ consists of Majorana spinors.

2.1.5 Groups of motions

The Clifford algebra $C\ell_{0,n}$ plays a central role in studying the group of *isometries* or *motions* of $\mathbb{R}^n$, not only the rotations and anti-rotations (orientation-reversing) of $\mathbb{R}^n$, but also translations of $\mathbb{R}^n$. Indeed, we can go further and include conformal transformations of $\mathbb{R}^n$ in our study, though we defer that discussion to a later section.

The starting point of all this is the observation that if v is any nonzero (and therefore invertible) vector of $\mathbb{R}^n$, then the map

$$\rho_v : \mathbb{R}^n \to \mathbb{R}^n; x \mapsto vx\hat{v}^{-1} = -vxv^{-1}$$

is reflection of $\mathbb{R}^n$ in the orthogonal complement in $\mathbb{R}^n$ of the line through the origin spanned by the vector v since any real multiple of v maps to minus itself, while any vector anti-commuting in the algebra with v maps to itself. Now any rotation of $\mathbb{R}^n$ is representable as the composite of an *even* number of hyperplane reflections and so is representable as conjugation of $\mathbb{R}^n$ by that composite, that is by an element of the *even* Clifford algebra $C\ell^0_{0\ n}$. Explicitly one has the following proposition.

Proposition 3 *Let g be a product of nonzero vectors in $C\ell_{0,n}$. Then the map*

$$\rho_g : x \mapsto gx\hat{g}^{-1}$$

is an orthogonal automorphism of $\mathbb{R}^n$, being a rotation if the number of factors is even, in which case $\hat{g} = g$.

The group of products of invertible vectors in $C\ell_{0,n}$ is known as the *Clifford* or *Lipschitz* group $\Gamma(n)$ of $\mathbb{R}^n$. The subgroup, each of whose elements is the product of an even number of nonzero vectors, is the even Clifford group, and is denoted by $\Gamma^0(n)$.

The *quadratic norm* of any element $g \in \Gamma(n)$ is defined to be the necessarily nonnegative real number g^-g, since each vector in $\mathbb{R}^n$ has quadratic norm equal to $v \cdot v$. The subgroup of $\Gamma^0(n)$, consisting of all elements of quadratic norm equal to 1, is known as the group $\mathbf{Spin}(n)$. Clearly any element g of $\mathbf{Spin}(n)$ induces the same rotation as its negative, and it turns out that this is the only ambiguity – the group $\mathbf{Spin}(n)$ *doubly covers* the *special orthogonal group* $SO(n)$ of *rotations*

of $\mathbb{R}^n$.

Of course, the original space $\mathbb{R}^n$ does not belong to the even Clifford algebra $_{0,n}$. We can get around this by multiplying the elements of $\mathbb{R}^n$ by the last basis $C\ell^0$ vector, the one that we used to identify $C\ell^0_n$ with $C\ell_{0,n-1}$. Then the space $\mathbb{R}^n$ maps to the space of *paravectors* in $C\ell_{0,n-1}$.

Proposition 3 therefore implies:

Proposition 4 *Let g be an element of* $\mathbf{Spin}(n)$. *Then, since $\hat{g} = g$, the map*

$$\rho_g : y \mapsto gyg \ ,$$

is a rotation of $\mathbb{R}^n$ and any rotation of that space may be so induced, the only ambiguity being one of the sign of g.

If we work with paravectors this is replaced by

Proposition 5 *Let g be an element of* $\mathbf{Spin}(n)$, *regarded as a subset of $C\ell_{0,n-1}$. Then the map*

$$\rho_g : y \mapsto gyg \ ,$$

where $y = \lambda + x$, with $\lambda \in \mathbb{R}$, and $x \in \mathbb{R}^{n-1}$, is a rotation *of the space of paravectors $\mathbb{R} \oplus \mathbb{R}^{n-1}$, and any rotation of that space may be so induced, the only ambiguity being one of the sign of g.*

Still remaining with the positive-definite case, we have

Theorem 3 *Let $\left(\begin{smallmatrix} a & \\ b & a \end{smallmatrix}\right)$ in $C\ell_{0,n}$ (2) represent an element of the even Clifford group $\Gamma^0(n+1)$ in $C\ell_{0,n+1}$ with $a \in \Gamma^0(n)$. Then the map $\mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto axa^{-1} + ba^{-1}$ is an orientation-preserving isometry of $\mathbb{R}^n$ and any isometry of $\mathbb{R}^n$ may be so represented, the representation being unique up to nonzero real multiples of a and b.*

Strictly speaking what is involved here is the subgroup of $C\ell_{0,n}(2)$ consisting of all matrices of the form $\left(\begin{smallmatrix} a & \\ 0 & a \end{smallmatrix}\right)$, with $a \in \Gamma^0(n)$ and $b = a$ p, where $p \in \mathbb{R}^n$. In the particular case that $n = 3$, $\mathbf{Spin}(4)$ is most frequently identified with the group $S^3 \times S^3 \subset {}^2\mathbb{H}$. An alternative to ${}^2\mathbb{H}$ consists of the matrices of $\mathbb{H}(2)$ of the form $\left(\begin{smallmatrix} a & \\ b & a \end{smallmatrix}\right)$. The subalgebra of $\mathbb{H}(2)$ consisting of all matrices of the form $\left(\begin{smallmatrix} a & \\ 0 & a \end{smallmatrix}\right)$ is known as Clifford's algebra (1873) of *biquaternions*. Elements of it are all of the

form $a + b\epsilon$, where a and b are quaternions and $\epsilon^2 = 0$, ϵ being represented in $\mathbb{H}(2)$ by the matrix $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. This is $C\ell_0^0$

Much of what we have said about isometries of the positive-definite spaces $\mathbb{R}^n$ extends to the indefinite spaces $\mathbb{R}^{p,q}$, especially the important cases for physics where either $p = 3$, $q = 1$, or $p = 1$, $q = 3$. We return to these in the next section, where we also show how Clifford algebras may be used to handle conformal transformations efficiently.

2.2 Conjugation

2.2.1 Symmetric and skew products and anti-involutions

In order to have a clear understanding of the various **Spin** groups, we begin by classifying the conjugation anti-involutions of Clifford algebras of nondegenerate real or complex quadratic spaces. According to Theorem 1 and Theorem 2, any such Clifford algebra is representable as a matrix algebra $\mathbb{A}(m)$, for some number m, where $\mathbb{A} = \mathbb{K}$ or ${}^2\mathbb{K}$ and $\mathbb{K} = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$.

A *correlation* on an n-dimensional vector space over $\mathbb{K} = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$ is just a *real* linear map from the vector space to its dual, this being *nondegenerate* if its kernel is zero, in which case it is a linear isomorphism. Let $\xi : v \mapsto v^\xi$ be such a nondegenerate correlation. It induces a $\mathbb{K}$-valued bilinear product $(a, b) \mapsto a^\xi b$.

It is an important fact, which here we take without proof, that any anti-involution of the matrix algebra $\mathbb{K}(n)$ is induced by either a symmetric or skew correlation on the vector space $\mathbb{K}^n$, in the sense discussed below.

The most familiar example is the map from $\mathbb{R}^n$ to its dual, which sends a column vector to the row vector which is its transpose. The induced product on $\mathbb{R}^n$ is then the standard positive-definite scalar product, *symmetric* since, for any $x, y \in \mathbb{R}^n$, $x \cdot y = y \cdot x$. The induced anti-involution on the real algebra $\mathbb{R}(n)$ of linear maps from $\mathbb{R}^n$ to itself, known as the *adjoint* anti-involution, is in this case a transposition. Let us denote the transpose of an element a of this algebra by a^τ. Then the elements a such that $a^\tau a = 1$ form the *orthogonal group* $O(n)$.

The product just discussed is *positive-definite*, since the square $x^{(2)} = x \cdot x$ of any element x of the vector space $\mathbb{R}^n$ is greater than or equal to zero, being equal to zero only when $x = 0$. Besides this, one has products of signature (p, q), where p of the basis vectors have scalar square -1 and q have scalar square $+1$. The associated groups are the groups $O(p, q)$, with $O(0, n) = O(n)$. The determinant of any element of $O(p, q)$ is

equal to $+1$ or -1. The elements with determinant equal to $+1$ form the *special orthogonal group* $SO(p, q)$.

Another example of a correlation, this time on $\mathbb{R}^2$, is the map $(x, y) \mapsto (y, -x)^\tau$ inducing a *skew* product

$$(x', y') \cdot (x, y) = y'x - x'y,$$

skew, since $(x, y) \cdot (x', y') = -(x', y') \cdot (x, y)$.

The induced anti-involution of $\mathbb{R}(2)$ is the map

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \mapsto \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

This product is known as the *standard symplectic product* on $\mathbb{R}^2$, the induced group consisting of all 2×2 real matrices with determinant 1. The notation used here for this group is $\mathrm{Sp}(2; \mathbb{R})$, or when it is necessary to save space in tables, by $\mathrm{Sp}_2(\mathbb{R})$.

An analogous nondegenerate symplectic product may be defined on each even-dimensional real vector space. The associated classical groups are the *real symplectic groups* $\mathrm{Sp}(2n; \mathbb{R})$. There are complex analogues of these products and groups, the associated classical groups being $O(n; \mathbb{C})$ and $\mathrm{Sp}(2n; \mathbb{C})$.

Elements of $O(n; \mathbb{C})$ have determinant equal to $+1$ or to -1. Those of determinant equal to $+1$ form the *special complex orthogonal group* $SO(n; \mathbb{C})$. All elements of both $\mathrm{Sp}(2n; \mathbb{R})$ and $\mathrm{Sp}(2n; \mathbb{C})$ have determinant equal to $+1$.

Also of importance is the real linear, but complex *semilinear* map sending a column vector of $\mathbb{C}^n$ to its *conjugate transpose*. This real linear map induces a *sesquilinear* (that is, one-and-a-half times linear) product on $\mathbb{C}^n$ which is symmetric in the sense that if one forms the product of two vectors in reverse order, then the resulting product is the *conjugate* of the product of the elements in the given order. The product that is simply the same one multiplied by the scalar i is then a *skew* sesquilinear product. Both give rise to the same anti-involution of the algebra $\mathbb{C}(n)$, namely conjugate transposition, the group of matrices whose inverse is the conjugate transpose forming the *unitary group* $U(n)$. Here again one has analogous products with signature (p, q), and associated groups $U(p, q)$, with $U(0, n) = U(n)$. Elements of $U(p, q)$ have determinant a complex number of modulus 1. The subgroup of those with determinant $+1$ is denoted by $SU(p, q)$.

For quaternionic spaces one has to be careful, since $\mathbb{H}$ is not commutative. The usual convention is to regard $\mathbb{H}^n$ as a *right-vector* space, this implying that the scalars multiply vectors on the *right*. The

dual space is then a left-vector space, and any correlation between the two has to be semilinear, involving one of the anti-involutions of $\mathbb{H}$. One of these is of course conjugation. The others correspond to reflections of the three-dimensional space of pure quaternions in a plane, the typical one that we choose being that in which only the quaternion j changes sign. Symmetric correlations of either type turn out to be equivalent to skew correlations of the other type. The groups that arise are denoted by $\mathrm{Sp}(n)$ for the symmetric conjugation case, equivalently for the skew symmetric case, and $O(n;\ \mathbb{H})$, or, more briefly, $O_n\ (\mathbb{H})$ for the opposite case. In the quaternionic symplectic case, one again has the possibility of other signatures, the groups being $\mathrm{Sp}(p,\ q)$, with $\mathrm{Sp}(0,\ n) = \mathrm{Sp}(n)$.

Finally, for each of the spaces ${}^2\mathbb{K}^{\ n}$, where $\mathbb{K} = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, there are symmetric or equivalent skew correlations involving the involution swap. It is of some interest that in each of these cases the group that arises is isomorphic in an obvious way to the *general linear group* $GL(n;\ \mathbb{K})$, elements of the group consisting of 2×2 matrices with entries in $\mathbb{K}(n)$, the top left entry being any invertible matrix, the bottom right entry some transform of this, depending on the details of the product, the other two off-diagonal entries being zero. The subgroup of $GL(n;\ \mathbb{R})$, consisting of those elements with determinant $+1$, forms the *special linear group* $SL(n;\ \mathbb{R})$. The subgroup of $GL(n;\ \mathbb{C})$, consisting of those elements with determinant $+1$, forms the *special linear group* $SL(n;\ \mathbb{C})$.

The *determinant* of an $n \times n$ quaternionic matrix is defined to be the square root of the determinant of the matrix regarded as a $2n \times 2n$ complex matrix, the latter necessarily having as determinant a nonnegative real number. The subgroup of $GL(n;\ \mathbb{H})$ consisting of those elements with determinant $+1$ forms the *special linear group* $SL(n;\ \mathbb{H})$.

There are altogether, up to fairly obvious equivalences, *ten* families of symmetric or skew correlations to choose from.

Theorem 4 *Any symmetric or skew-symmetric correlation on a right $\mathcal{A}$-linear space of finite dimension > 1 belongs to one of the following ten types, which are mutually exclusive.*

0	*a symmetric* $\mathbb{R}$*-correlation;*
1	*a symmetric, or equivalently, a skew* ${}^2\mathbb{R}^\sigma$*-correlation;*
2	*a skew* $\mathbb{R}$*-correlation;*
3	*a skew* $\mathbb{C}$*-correlation;*
4	*a skew* $\overline{\mathbb{H}}$*- or equivalently a symmetric* $\overline{\mathbb{H}}$*-correlation;*
5	*a skew, or equivalently a symmetric,* ${}^2\overline{\mathbb{H}}^\sigma$*-correlation;*
6	*a symmetric,* $\widetilde{\mathbb{H}}$*- or equivalently a skew* $\overline{\mathbb{H}}$*-correlation;*
7	*a symmetric* $\mathbb{C}$*-correlation;*
8	*a symmetric, or equivalently a skew,* $\overline{\mathbb{C}}$*-correlation;*
9	*a symmetric, or equivalently a skew,* ${}^2\overline{\mathbb{C}}^\sigma$*-correlation.*

The logic behind the numbering of these ten types derives from the order in which most of the cases appear in Table 6.1 below.

The job of identification in any particular case is made easier by the fact that an anti-involution of an algebra is uniquely determined by its restriction to any subset that generates the algebra. This means that in classifying the conjugation anti-involution of a Clifford algebra $C\ell(X)$, all one has to do is to examine the representatives in the algebra of an orthonormal basis for the quadratic space X.

What has to be noted is that among all the products that arise, signature is an invariant for just three, symmetric real products on $\mathbb{R}^n$, unitary products on $\mathbb{C}^n$, and the quaternionic symplectic products on $\mathbb{H}^n$. The corresponding real quadratic spaces are denoted by $\mathbb{R}^{p,q}$, $\overline{\mathbb{C}}^{p,q}$ and $\overline{\mathbb{H}}^{p,q}$. In each of these cases, products of signature $(p,\ p)$ will be said to be *neutral*.

It is not possible to give all the details here. The interested reader is referred to Porteous (1995) for the full story. Table 6.1 first appeared explicitly in Hampson (1969), was published in Porteous (1969), but was implicit in Wall (1968).

One case that is easily verified is that of the positive-definite quadratic spaces. One has

Theorem 5 *Conjugation on the Clifford algebra* $C\ell_{0,n}$ *is conjugate transposition of the matrices representing the elements of the algebra.*

We begin by reminding the reader that the Clifford algebras $C\ell_{p,q}$ for $0 \leq p,\ q < 7$ and $\mathbb{C} \otimes C\ell_{0,n}$ for $0 \leq n < 7$ are as in Table 1.1 and Theorem 2. First, we recall

Table 1.1

p	± 1	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{H}$	${}^2\mathbb{H}$	$\mathbb{H}(2)$	$\mathbb{C}(4)$	$\mathbb{R}(8)$	${}^2\mathbb{R}(8)$
$\downarrow$	$\mathbb{R}$	${}^2\mathbb{R}$	$\mathbb{R}(2)$	$\mathbb{C}(2)$	$\mathbb{H}(2)$	${}^2\mathbb{H}(2)$	$\mathbb{H}(4)$	$\mathbb{C}(8)$	$\mathbb{R}(16)$
	$\mathbb{C}$	$\mathbb{R}(2)$	${}^2\mathbb{R}(2)$	$\mathbb{R}(4)$	$\mathbb{C}(4)$	$\mathbb{H}(4)$	${}^2\mathbb{H}(4)$	$\mathbb{H}(8)$	$\mathbb{C}(16)$
	$\mathbb{H}$	$\mathbb{C}(2)$	$\mathbb{R}(4)$	${}^2\mathbb{R}(4)$	$\mathbb{R}(8)$	$\mathbb{C}(8)$	$\mathbb{H}(8)$	${}^2\mathbb{H}(8)$	$\mathbb{H}(16)$
	${}^2\mathbb{H}$	$\mathbb{H}(2)$	$\mathbb{C}(4)$	$\mathbb{R}(8)$	${}^2\mathbb{R}(8)$	$\mathbb{R}(16)$	$\mathbb{C}(16)$	$\mathbb{H}(16)$	${}^2\mathbb{H}(16)$
	$\mathbb{H}(2)$	${}^2\mathbb{H}(2)$	$\mathbb{H}(4)$	$\mathbb{C}(8)$	$\mathbb{R}(16)$	${}^2\mathbb{R}(16)$	$\mathbb{R}(32)$	$\mathbb{C}(32)$	$\mathbb{H}(32)$
	$\mathbb{C}(4)$	$\mathbb{H}(4)$	${}^2\mathbb{H}(4)$	$\mathbb{H}(8)$	$\mathbb{C}(16)$	$\mathbb{R}(32)$	${}^2\mathbb{R}(32)$	$\mathbb{R}(64)$	$\mathbb{C}(64)$
	$\mathbb{R}(8)$	$\mathbb{C}(8)$	$\mathbb{H}(8)$	${}^2\mathbb{H}(8)$	$\mathbb{H}(16)$	$\mathbb{C}(32)$	$\mathbb{R}(64)$	${}^2\mathbb{R}(64)$	$\mathbb{R}(128)$

extending indefinitely either way with period 8.

Here, and in the tables that follow, the left-hand column is to be added in when the even subalgebras are under consideration.

Theorem 6 *Conjugation types for the algebras $C\ell_{p,q}$ are to be overlaid on Table 1.1.*

Table 6.1

p mod 8	^{20}	0	8	4	^{24}	4	8	0	^{20}
$\downarrow$	0	1	2	3	4	5	6	7	0
	8	2	^{22}	2	8	6	^{26}	6	8
	4	3	2	1	0	7	6	5	4
	^{24}	4	8	0	^{20}	0	8	4	^{24}
	4	5	6	7	0	1	2	3	4
	8	6	^{26}	6	8	2	^{22}	2	8
	0	7	6	5	4	3	2	1	0

q mod 8 $\rightarrow$

For the complexifications of $C\ell_{p,q}$ we have two results:

Theorem 7 *Table 7.2 classifies conjugation for the complexification of the algebras $C\ell_{p,q}$ when the complex coefficients are not conjugated. Here signature is not important, the result depending only on $n = p + q$. The table of algebras is extending indefinitely with period 8.*

The classification then follows by overlaying the following table on it.

Table 7.1

	$n \bmod 8 \rightarrow$								
± 1	|	$\mathbb{C}$	${}^2\mathbb{C}$	$\mathbb{C}(2)$	${}^2\mathbb{C}(2)$	$\mathbb{C}(4)$	${}^2\mathbb{C}(4)$	$\mathbb{C}(8)$	${}^2\mathbb{C}(8)$

Table 7.2

	$n \bmod 8 \rightarrow$								
^{27}	|	7	9	3	^{23}	3	9	7	^{27}

Theorem 8 *Table 8.2 classifies conjugation for the complexification of the algebras $C\ell_{p,q}$ when the complex coefficients are conjugated. Here signature remains important.*

The table of algebras is

Table 8.1

		$q \rightarrow$	
p	± 1	$\mathbb{C}$	${}^2\mathbb{C}$
$\downarrow$	$\mathbb{C}$	${}^2\mathbb{C}$	$\mathbb{C}(2)$

extending indefinitely with period 2.

The classification then follows by overlaying the following table on it.

Table 8.2

		$q \bmod 2 \rightarrow$		
$p \bmod 2$	^{28}	|	8	^{28}
$\downarrow$	8	|	8	8

For the codes 0, 4 and 8 in Tables 6.1 and 7.2, there is a further classification by signature. The choice along the top row or down the extra column on the left is the positive-definite one. Elsewhere, the choice is the neutral one.

For completeness we also have the following two results:

Theorem 9 *For the tensor product of $C\ell_{p,q}$ by ${}^2\mathbb{R}$, with conjugation swapping the components of the coefficients, we have*

Table 9.1

$-p+q \bmod 8 \rightarrow$								
${}^2 1$	1	9	5	${}^2 5$	5	9	1	${}^2 1$

Theorem 10 *For the tensor product of $C\ell_{p,q}$ by ${}^2\mathbb{C}$, with conjugation swapping the components of the coefficients, we have, with $n = p + q$,*

Table 10.1

$n \bmod 2 \rightarrow$		
${}^2 9$	9	${}^2 9$

Tables 6.1, 7.2, 8.2, 9.1 and 10.1 may be more appreciated if the various code numbers are replaced by the classical groups that preserve the sesquilinear forms on the spinor spaces. We give them here for $0 \leq p + q < 8$ in Tables 6.1′ to 10.1′. To save space, we abbreviate the notations slightly in obvious ways.

Table 6.1′

	$q \rightarrow$							
p	O_1	U_1	Sp_1	${}^2Sp_1$	Sp_2	U_4	O_8	${}^2O_8$
$\downarrow$	$GL_1(\mathbb{R})$	$Sp_2(\mathbb{R})$	$Sp_2(\mathbb{C})$	$Sp_{1,1}$	$GL_2(\mathbb{H})$	$O_4(\mathbb{H})$	$O_8(\mathbb{C})$	
	$Sp_2(\mathbb{R})$	${}^2Sp_2(\mathbb{R})$	$Sp_4(\mathbb{R})$	$U_{2,2}$	$O_4(\mathbb{H})$	${}^2O_4(\mathbb{H})$		
	$Sp_2(\mathbb{C})$	$Sp_4(\mathbb{R})$	$GL_4(\mathbb{R})$	$O_{4,4}$	$O_8(\mathbb{C})$			
	$Sp_{1,1}$	$U_{2,2}$	$O_{4,4}$	${}^2O_{4,4}$				
	$GL_2(\mathbb{H})$	$O_4(\mathbb{H})$	$O_8(\mathbb{C})$					
	$O_4(\mathbb{H})$	${}^2O_4(\mathbb{H})$						
	$O_8(\mathbb{C})$							

Table 7.1′

$n \rightarrow$							
$O_1(\mathbb{C})$	$GL_1(\mathbb{C})$	$Sp_2(\mathbb{C})$	${}^2Sp_2(\mathbb{C})$	$Sp_4(\mathbb{C})$	$GL_4(\mathbb{C})$	$O_8(\mathbb{C})$	${}^2 0_8(\mathbb{C})$

Table 8.1′

$p \downarrow$	$q \rightarrow$							
	U_1	${}^2U_1$	U_2	${}^2U_2$	U_4	${}^2U_4$	U_8	${}^2U_8$
$GL_1(\mathbb{C})$	U_2	$GL_2(\mathbb{C})$	U_4	$GL_4(\mathbb{C})$	U_8	$GL_8(\mathbb{C})$		
	U_2	${}^2U_2$	U_4	${}^2U_4$	U_8	${}^2U_8$		
$GL_2(\mathbb{C})$	U_4	$GL_4(\mathbb{C})$	U_8	$GL_8(\mathbb{C})$				
	U_4	${}^2U_4$	U_8	${}^2U_8$				
$GL_4(\mathbb{C})$	U_8	$GL_8(\mathbb{C})$						
	U_8	${}^2U_8$						
$GL_8(\mathbb{C})$								

Table 9.1′

$p \downarrow$	$q \rightarrow$							
	$GL_1(\mathbb{R})$	$GL_1(\mathbb{C})$	$GL_1(\mathbb{H})$	${}^2GL_1(\mathbb{H})$	$GL_2(\mathbb{H})$	$GL_4(\mathbb{C})$	$GL_8(\mathbb{R})$	${}^2GL_8(\mathbb{R})$
${}^2GL_1(\mathbb{R})$	$GL_2(\mathbb{R})$	$GL_2(\mathbb{C})$	$GL_2(\mathbb{H})$	${}^2GL_2(\mathbb{H})$	$GL_4(\mathbb{H})$	$GL_8(\mathbb{C})$	$GL_{16}(\mathbb{R})$	
etc.								

Table 10.1′

$n \rightarrow$							
$GL_1(\mathbb{C})$	${}^2GL_1(\mathbb{C})$	$GL_2(\mathbb{C})$	${}^2GL_2(\mathbb{C})$	$GL_4(\mathbb{C})$	${}^2GL_4(\mathbb{C})$	$GL_8(\mathbb{C})$	${}^2GL_8(\mathbb{C})$

The *dimension* of the classical group consisting of matrices g such that $g^{-}g = 1$ is the dimension of its *Lie algebra*, the real vector space of matrices g such that $g + g^{-} = 0$.

The dimensions of the groups in Table 6.1′ are shown in Table 6.1″.

Table 6.1″

0	1	3	6	10	16	28	56
1	3	6	10	16	28	56	
3	6	10	16	28	56		
6	10	16	28	56			
10	16	28	56				
16	28	56					
28	56						
56							

These depend only on the rank $n = p + q$ and not on the index p, q. The dimensions of the groups in Table 7.1′ are twice those of the groups in Table 6.1′.

2.2.2 *Tables of spin groups*

For each n, the quadratic norm $g^{-}g$ of any element g of the group $\mathbf{Spin}(n)$ is equal to $+1$, the group being a subgroup of the *even* classical group associated in Table 6.1′ to the signature $(0, n)$ or $(n, 0)$. That group, being the part of the classical group that lies in the even Clifford algebra for the given signature, lies in the table either in the position $p = 0$, $q = n - 1$ or in the position $p = n$, $q = -1$ (in an extra column on the left that matches the first row). For any p, q, with neither p nor q equal to 0, the quadratic norm of an element of $\mathbf{Spin}(p, q)$ may be equal either to $+1$ or to -1. It is then the subgroup $\mathbf{Spin}^{+}(p, q)$, consisting of those elements of $\mathbf{Spin}(p, q)$ with quadratic norm $+1$, that is a subgroup of the even classical group for the signature, namely the classical group in the position p, $q - 1$. The group $SO^{+}(p, q)$, with neither p nor q equal to 0, which it covers twice, is the group of *orthochronous* isometries of $\mathbb{R}^{p,q}$ that preserve not only the orientations of $\mathbb{R}^{p,q}$ but also its *semi-orientations*, the important case for physics being when $p, q = 3, 1$ or $1, 3$.

The group $\mathbf{Spin}(n)$ or $\mathbf{Spin}^{+}(p, q)$, with $n = p+q$, has dimension $\frac{1}{2}$ $n(n-1)$. It is the whole group in Table 6.1′, for n $p + q \leq 5$, but is of dimension one less than this, namely of dimension 15, rather than 16, for $n = p + q = 6$. In this case, each algebra has a real-valued determinant, and lowering the dimension by 1 corresponds to taking the determinant equal to 1.

Theorem 11 *The groups* $\mathbf{Spin}(n)$ *for* $n \leq 6$, *as well as the groups* $\mathbf{Spin}^{+}$ (p, q) *for* $p + q \leq 6$, *where both* p *and* q *are nonzero, are shown in Table 1 1.1.*

Table 11.1

$p \downarrow$	$q \rightarrow$ ± 1	O_1	U_1	Sp_1	${}^2Sp_1$	Sp_2	SU_4
	O_1	$GL_1(\mathbb{R})$	$Sp_2(\mathbb{R})$	$Sp_2(\mathbb{C})$	$Sp_{1,1}$	$SL_2(\mathbb{H})$	
	U_1	$Sp_2(\mathbb{R})$	${}^2Sp_2(\mathbb{R})$	$Sp_4(\mathbb{R})$	$SU_{2,2}$		
	Sp_1	$Sp_2(\mathbb{C})$	$Sp_4(\mathbb{R})$	$SL_4(\mathbb{R})$			
	${}^2Sp_1$	$Sp_{1,1}$	$SU_{2,2}$				
	Sp_2	$SL_2(\mathbb{H})$					
	SU_4						

The groups $\mathbf{Spin}(n;\ \mathbb{C})$, *for* $n \leq 6$, *are shown in Table 11.2.*

Table 11.2

$n \rightarrow$ ± 1	$O_1(\mathbb{C})$	$GL_1(\mathbb{C})$	$Sp_2(\mathbb{C})$	${}^2Sp_2(\mathbb{C})$	$Sp_4(\mathbb{C})$	$SL_4(\mathbb{C})$

As a matter of fact it is enough to prove that $\mathbf{Spin}(6;\ \mathbb{C}) \cong SL(4;\ \mathbb{C})$. All the real cases then follow by restriction. Many writers do not give the entirety of these tables and some are in error. Note in particular that $\mathbf{Spin}^+(3,\ 3) \cong SL(4;\ \mathbb{R})$. Of course, for physics, an important case is

$$\mathbf{Spin}^+(3,1) \cong \mathbf{Spin}^+(1,3) \cong Sp_2(\mathbb{C}),$$

lying in the even Clifford algebra

$$C\ell^0_{3,1} \cong C\ell^0_{1,3} \cong \mathbb{C}(2),$$

this group just being the six-dimensional group of 2×2 complex matrices of determinant 1.

2.2.3 *The Radon–Hurwitz numbers*

Because of the overarching position of the positive-definite quadratic spaces, the top row of Table 1.1 is especially important.

One application is to the construction of linear subspaces of the groups $GL(s;\ \mathbb{R})$, for finite s, a *linear subspace* of $GL(s;\ \mathbb{R})$ being, by definition, a linear subspace of the real matrix algebra $\mathbb{R}(s)$ all of whose elements, with the exception of the origin, are invertible.

For example, the standard copy of $\mathbb{C}$ in $\mathbb{R}(2)$ is a linear subspace of $GL(2;\ \mathbb{R})$ of dimension 2, while either of the standard copies of $\mathbb{H}$ in $\mathbb{R}(4)$ is a linear subspace of $GL(4;\ \mathbb{R})$ of dimension 4. On the other hand, when s is odd, there is no linear subspace of $GL(s;\ \mathbb{R})$ of dimension greater than 1. For if there were such a space of dimension greater than 1, then there would exist linearly independent elements a and b of $GL(s;\ \mathbb{R})$ such that, for all $\lambda \in \mathbb{R}$, $a + \lambda b \in GL(s;\ \mathbb{R})$ and therefore such that $c + \lambda\, 1 \in GL(s;\ \mathbb{R})$, where $c = b^{-1}a$. However, by the fundamental theorem of algebra, there is a real number λ such that $\det(c + \lambda\, 1) = 0$, $\mathbb{R} \to \mathbb{R}$; $\lambda \mapsto \det(c + \lambda\, 1)$ being a polynomial map of odd degree. This provides a contradiction.

Proposition 6 *Let $\mathbb{K}(m)$ be a possibly nonuniversal Clifford algebra for the positive-definite quadratic space $\mathbb{R}^n$, for any positive integer n. Then $\mathbb{R} \oplus \mathbb{R}^n$ is a linear subspace of $GL(m;\ \mathbb{K})$ and therefore of one of the groups $GL(m;\ \mathbb{R})$, $GL(2m;\ \mathbb{R})$ or $GL(4m;\ \mathbb{R})$, according as $\mathbb{K} = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. Moreover, the conjugate of any element of $\mathbb{R} \oplus \mathbb{R}^n$ is the conjugate transpose of the representative in $GL(m;\ \mathbb{K})$ or, equivalently, the transpose of its representative in $GL(m;\ \mathbb{R})$, $GL(2m;\ \mathbb{R})$ or $GL(4m;\ \mathbb{R})$.*

The following follows from the top line of Table 1.1.

Proposition 7 *Let $\{\chi(k)\}$ be the sequence of positive integers defined by $\chi(8p + q) = 4p + j$, where $j = 0$ for $q = 0$, 1 for $q = 1$, 2 for $q = 2$ or 3 and 3 for $q = 4$, 5, 6 or 7. Then if $2^{\chi(k)}$ divides s, there exists a k-dimensional linear subspace X of $GL(s;\ \mathbb{R})$ such that*

1. *for each $x \in X$, $x^\tau = -x$, $x^\tau x = -x^2$, being a nonnegative real multiple of ${}^s1$, and zero only if $x = 0$,*
2. *$\mathbb{R} \oplus X$ is a $(k + 1)$-dimensional linear subspace of $GL(s;\ \mathbb{R})$.*

The sequence χ is called the *Radon–Hurwitz sequence* (Radon (1923) and Hurwitz (1923)). It can be proved that there is no linear subspace of $GL(s;\ \mathbb{R})$) of dimension greater than that asserted here.

As a particular case, there is an eight-dimensional linear subspace of $GL(8;\ \mathbb{R})$, since $\mathbb{R}(8)$ is a (nonuniversal) Clifford algebra for $\mathbb{R}^7$. This remark provides a route into the study of the algebra of *Cayley numbers*, also known as the *octonion* algebra.

2.2.4 *Vahlen matrices and conformal transformations*

Consider the positive-definite quadratic space $\mathbb{R}^n$, with Clifford algebra $C\ell_{0,n}$ and Clifford group Γ. It follows from earlier work that the real

algebra $C\ell_{0,n}(2)$ of 2×2-matrices with entries in $C\ell_{0,n}$ is isomorphic to $C\ell_{1,n+1}$, where elements of the vector space $\mathbb{R}^{1,n+1}$ are represented by matrices of the form $\left(\begin{smallmatrix} x & \nu \\ \mu & -x \end{smallmatrix}\right)$, where $x \in \mathbb{R}^n$ and μ, $\nu \in \mathbb{R}$, such matrices being referred to below as *vectors* in $C\ell_{0,n}(2)$. Let $\Gamma(2)$ denote the Clifford group of $C\ell_{0,n}(2)$. For many applications, one would like to characterize the elements of $\Gamma(2)$ in terms of $\mathbb{R}^n$ and Γ. Such a characterization was given by Vahlen (1902), extending the work of Clifford on biquaternions, and his work was re-presented in a series of papers by Ahlfors in the early 80s, for example (1985), (1986). The reader is referred to a whole series of papers that have appeared during the last 15 years, for example, the paper of Jan Cnops (1994), developed from earlier work of Maks (1989) and Fillmore and Springer (1990). For a parallel account, involving paravectors, see Elstrodt, Grunewald and Mennicke (1987). See also Waterman (1993), Porteous (1995) and Pozo and Sobczyk (2002).

Here we limit ourselves to discussing the positive-definite case. We begin by describing conjugation and reversion on $C\ell_{0,n}(2)$.

Proposition 8 *For any element of* $C\ell_{0,n}(2)$

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}^{-} = \begin{pmatrix} d^{-} & -c^{-} \\ -b^{-} & a^{-} \end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} a & c \\ b & d \end{pmatrix}^{\sim} = \begin{pmatrix} d^{\sim} & c^{\sim} \\ b^{\sim} & a^{\sim} \end{pmatrix}$$

These hold for vectors in $C\ell_{0,n}(2)$ and so for the whole of $C\ell_{0,n}(2)$.

The next theorem describes the Clifford group $\Gamma(2)$ of $C\ell_{0,n}(2)$.

Theorem 12 *Let* Γ *be the Clifford group of* $\mathbb{R}^n$, *and let* G *be the set of all matrices* $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ *of* $C\ell_{0,n}(2)$ *such that*

$$(a)\ a,\ b,\ c,\ d \in \Gamma; \quad (b)\ a\,b^{-},\ c\,d^{-},\ a^{\sim}c,\ b^{\sim}d \in \mathbb{R}^n; \quad (c)\ \Delta = a\,d^{-} - c\,b^{-} \in \mathbb{R}^{*}.$$

Then G *is the Clifford group* $\Gamma(2)$.

The number Δ is known as the *pseudo-determinant* of the matrix.

The indefinite case is somewhat trickier to handle. One important difference is that the four entries in a Vahlen matrix need not belong to Γ. Each must be a finite product of vectors, but some of these may be null. Of course, the pseudo-determinant must still be a nonzero real number.

Conformal maps are maps that preserve angle. One's first encounter with conformality is probably in a course on functions of one complex variable, where it is proved that any holomorphic map is conformal. However, it was proved long ago by Liouville (1850) that conformality for transformations of $\mathbb{R}^3$ is much more restrictive. He proved that the image of any plane or sphere must be either a plane or sphere. It then follows from a theorem of Möbius that any such map is representable as the composite of a finite number of orthogonal maps, translations, or inversions of $\mathbb{R}^3$ in spheres. The simplest such inversion is the inversion in the sphere, with centre the origin, namely the map $\mathbb{R}^3 \to \mathbb{R}^3 : x \mapsto x/|x|^2$, defined everywhere except at the origin. The obvious analogue of this theorem holds for positive-definite quadratic spaces of any finite dimension greater than 3. The analogous statement for indefinite quadratic spaces is also true by a theorem of Haantjes (1938).

We show here how Clifford algebras may be used to handle such Möbius maps. We limit ourselves to the case of positive-definite quadratic spaces, explicitly $\mathbb{R}^n$, for any positive n.

The trick is first to map the space $\mathbb{R}^n$ to the unit sphere S^n in $\mathbb{R}^{n+1}$ by stereographic projection from the South Pole $(0, \ldots, 0, 1)$, but factoring this through the null-cone in $\mathbb{R}^{1,n+1}$, as follows:

$$x \to (\tfrac{1}{2}(1 + x \cdot x),\ \tfrac{1}{2}(1 - x \cdot x),\ x) \to (\frac{1 - x \cdot x}{1 + x \cdot x},\ \frac{2x}{1 + x \cdot x}).$$

It is not difficult to prove that any orthogonal transformation of $\mathbb{R}^{1,n+1}$ not only preserves the null-cone, but also induces a Möbius transformation of $\mathbb{R}^n$, and moreover any Möbius transformation of $\mathbb{R}^n$ may be so induced.

The *Möbius group* $M(0, n)$ is the connected component of the identity of the group of Möbius transformations of $\mathbb{R}^n$. One can save a dimension by identifying $\mathbb{R}^n$ with the space of paravectors in $\mathbb{R}^{0,n-1}$, and then representing such a paravector x by the matrix

$$\tfrac{1}{2}(1 + x \cdot x)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \tfrac{1}{2}(1 - x \cdot x)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} x & 0 \\ 0 & x^- \end{pmatrix}$$

$$= \begin{pmatrix} x & x x^- \\ 1 & x^- \end{pmatrix} = \begin{pmatrix} x \\ 1 \end{pmatrix} \begin{pmatrix} 1 & x^- \end{pmatrix}.$$

Consider now an element of $\mathbf{Spin}^+(1, n+1)$ represented by an element of $C\ell_{0,\,n-1}(2)$ of the form

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

By Proposition 5 and Proposition 8, it maps the paravector representing the vector x to

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} x & x x^- \\ 1 & x^- \end{pmatrix} \begin{pmatrix} d^- & c^- \\ b^- & a^- \end{pmatrix} = \lambda \begin{pmatrix} x' & x' x'^- \\ 1 & x' \end{pmatrix},$$

where $x' = (ax + c)(bx + d)^{-1}$ and λ is the real number $(bx + d)(bx + d)^-$.

For example, the *translation* $x \mapsto x + c$ is represented by the matrix

$$\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$$

and *inflation* by the positive scalar ρ by the matrix

$$\begin{pmatrix} \sqrt{\rho} & 0 \\ 0 & \sqrt{\rho}^{-1} \end{pmatrix},$$

while inversion in the unit quasi-sphere composed with the hyperplane reflection $x \mapsto -x^-$ is represented by the matrix

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

Representations of the Möbius groups $M(0, n) = M(\mathbb{R}^n)$ for $n \leq 4$ are given in the following theorem.

Theorem 13

$$\begin{aligned} M(0,1) &\cong Sp(2,\mathbb{R})/\{1,-1\} \\ M(0,2) &\cong Sp(2,\mathbb{C})/\{1,-1\} \\ M(0,3) &\cong Sp(1,1)/\{1,-1\} \\ M(0,4) &\cong SL(2,\mathbb{H})/\{1,-1\} \end{aligned}$$

The indefinite cases follow the same route, except for some detail. In particular, for pq odd, the group $\mathbf{Spin}^+(p+1, q+1)$ covers the Möbius group $M(p, q)$ four times and not twice. In particular, for the case $p, q = 1, 3$, important for physics,

$$M(1,3) = SU(2,2)/\{1, \mathrm{i}, -1, -\mathrm{i}\}.$$

2.3 References

[1] Ahlfors, L. (1985), Möbius transformations and Clifford numbers, in I.

Chavel, H.M. Parkas (eds) *Differential Geometry and Complex Analysis*. Dedicated to H.E. Rauch, Springer-Verlag, Berlin, pp. 65–73.

[2] Ahlfors, L. (1986), Mòbius transformations in $\mathbb{R}^n$ expressed through 2×2 matrices of Clifford numbers, *Complex Variables* **5**, 215–224.

[3] Clifford, W.K. (1876), Preliminary sketch of Biquaternions, *Proc. London Math. Soc.* **4**, 381–395.

[4] Cnops, J. (1994), *Hurwitz Pairs and Applications of Mòbius Transformations*, Thesis, Universiteit Gent.

[5] Elstrodt, J., Grunewald, F. and Mennicke, J. (1987), Vahlen's group of Clifford matrices and spin groups, *Math. Z.* **196**, 369–390.

[6] Fillmore, J. and Springer, A. (1990), Möbius groups over general fields using Clifford algebras associated with spheres, *Int. J. Theo. Phys.* **29**, 225–246.

[7] Grassmann, H. (1844), *Die Wissenschaft der extensiven Grösse oder die Aus- dehnungslehre, eine neue mathematishcen Disclipin*, Leipzig.

[8] Haantjes, J. (1937), Conformal representations of an n-dimensional euclidean space with a nondefinite fundamental form on itself, *Proc. Ned. Akad. Wet. (Math)* **40**, 700–705.

[9] Hamilton, W.R. (1844), On quaternions, or on a new system of imaginaries in algebra, *Phil. Mag.* **25**, 489–495, reprinted in *The Mathematical papers of Sir William Rowan Hamilton*, Vol. III, *Algebra*, Cambridge University Press, London (1967), pp. 106–110.

[10] Hampson, A. (1969), *Clifford Algebras*, MSc. Thesis, University of Liverpool.

[11] Hurwitz, A. (1923), Uber die Komposition der quadratischen Formen, *Math. Ann.* **88**, 294–298.

[12] Liouville, J. (1850), Appendix to Monge, G. *Application de l'analyse à la geométrie*. 5 éd. par Liouville.

[13] Lipschitz, R. (1884), *Üntersuchungen über die Summen von Quadraten*, Bonn.

[14] Maks, J. (1989), *Modulo* (1, 1) *periodicity of Clifford algebras and the generalized (anti-)Möbius transformations*, Thesis, Technische Universiteit Delft.

[15] Maks, J. (1992), Clifford algebras and Möbius transformations, in A. Micali et al. (eds) *Clifford Algebras and their Applications in Mathematical Physics*, Kluwer Acad. Publ., Dordrecht.

[16] Porteous, I.R. (1969), *Topological Geometry*, 1st Edition, Van Nostrand Reinhold.

[17] Porteous, I.R. (1995), *Clifford Algebras and the Classical Groups*, Cambridge University Press, Cambridge.

[18] Pozo, J.M. and Sobczyk, G (2002), Geometric algebra in linear algebra and geometry, *Acta Applicandae Mathematicae* **71**, 207–244.

[19] Radon, J. (1923), Lineare Scharen orthogonaler Matrizen, *Abh. math. Semin. Univ. Hamburg* 1, 1–14.

[20] Vahlen, K.Th. (1902), Uber Bewegungen und complexe Zahlen, *Math. Ann.* **55**, 585593.

[21] Wall, C.T.C. (1968), Graded algebras, anti involutions, simple groups and symmetric spaces, *Bull. Am. Math. Soc.* **74**, 198–202.

[22] Waterman, P.L. (1993), Möbius transformations in several dimensions, *Adv. in Math.* **101**, 87–113.

Ian R. Porteous
Department of Mathematical Sciences
Division of Pure Mathematics
University of Liverpool
Liverpool L69 3BX, U.K.
E-mail: porteous@liverpool.ac.uk

Received: January 31, 2002; Revised: March 15, 2003.

This lecture was presented at "Lecture Series on Clifford Algebras and their Applications", May 18 and 19, 2002, as part of the 6th International Conference on Clifford Algebras and their Applications in Mathematical Physics, Cookeville, TN, May 20–25, 2002.

AMS Subject Classification: 30G35.

Keywords: Cauchy–Riemann operator, monogenic functions, Cauchy's theorem, Cauchy's integral, Plemelj formula, Hardy space.

X. TOPOLOGY AND ANALYSIS

§1. Topological spaces

In Chapters VI, VII, and VIII, the following definitions were used. Let V, W be vector spaces with scalar product.

(A) A subset $D \subset V$ is <u>open</u> if for each $X \in D$ there exists a positive number δ such that $|Y - X| < \delta$ implies $Y \in D$ (Definition VI, 5.1). [Note that any δ' with $0 < \delta' < \delta$ may be used in place of δ. Note also that any ball

$$B_r(A) = \{X \mid X \in V \text{ and } |X - A| < r\} ,$$

with center A and "radius" $r > 0$, is an open subset of V according to this definition (why?) and that the empty set $\emptyset$ is open (why?). A convenient rephrasing of the above definition is: $D \subset V$ is open if for each $X \in D$ there is a ball $B_\delta(X) \subset D$.]

(B) If $F: D \longrightarrow W$ and $A \in V$, then F has the <u>limit</u> $B \in W$ as X tends to A ($\lim_A F = B$) if, for each positive number ε, there exists a positive number δ such that $X \in D$ and $0 < |X - A| < \delta$ imply $|F(X) - B| < \varepsilon$ (Definition VI, 1.1). Here it is assumed that the set $\{X \mid X \in D$ and $0 < |X - A| < \delta\}$ is not the empty set. [This is no problem if D is open and $A \in D$; in general the assumption is that A is a limit point of D (see Definition 2.5).] In particular, $D \neq \emptyset$; the formula $F: D \longrightarrow W$ will always assume

$D \neq \emptyset$.

(C) If $F: D \longrightarrow W$ and if $A \in D$, then F is <u>continuous at A</u> if $\lim_A F$ exists and is $F(A)$ (Definition VI, 1.2). [When the condition of Definition (B) is applied for $B = F(A)$, the condition that F be continuous at A becomes: for each positive number ε there exists a positive number δ such that $X \in D$ and $|X - A| < \delta$ imply $|F(X) - F(A)| < \varepsilon$. If D is open, then the condition is equivalent to: for each $B_\varepsilon(F(A))$ there is a $B_\delta(A)$ such that $F(B_\delta(A)) \subset B_\varepsilon(F(A))$.] The function F is <u>continuous in D</u> if it is continuous at each $A \in D$.

Among the theorems proved was the following: if $T: V \longrightarrow W$ is an element of $L(V, W)$, and if V, W are finite dimensional, then T is continuous in V (Proposition VI, 1.7). To prove this it was sufficient to verify the proposition for the case $W = R$, i.e. $T \in V^*$.

1.1. <u>Proposition</u>. If $F: D \longrightarrow W$ where D is open and if $A \in D$, then F is continuous at A if and only if, for each open set $E' \subset W$ with $F(A) \in E'$, there exists an open set $E \subset D$ with $A \in E$ such that $F(E) \subset E'$. The function F is continuous in D if and only if $F^{-1}(E')$ is open for each open set $E' \subset W$.

<u>Remark</u>. The condition of Proposition 1.1 is stated without explicit mention of scalar products and may be taken as the definition of continuity in more general situations, provided the notion of open sets in V and W is well-defined (see below).

The proposition then states that the new definition is equivalent to the old definition in the case that $D \subset V$ is open.

Proof. Suppose $F: D \longrightarrow W$ is continuous at $A \in D$ according to Definition (C), and let $E' \subset W$ be any open set such that $F(A) \in E'$. By Definition (A), applied to W, there is a $B_\varepsilon(F(A)) \subset E'$. By (C), there is a $B_\delta(A)$ — an open set — such that $F(B_\delta(A)) \subset B_\varepsilon(F(A)) \subset E'$. The condition of the proposition is therefore satisfied by taking $E = B_\delta(A)$. Conversely, if F satisfies the condition of the proposition for every choice of an open set E' with $F(A) \in E'$ and in particular for $E' = B_\varepsilon(F(A))$, there is an open set $E \subset V$ with $A \in E$ and $F(E) \subset E'$. Since E is open, there is a $B_\delta(X) \subset E$. Then $F(B_\delta(X)) \subset F(E) \subset E' = B_\varepsilon(F(A))$. Thus F is continuous at $A \in D$.

Let F be continuous in D and let E' be open in W. If $F^{-1}(E') = \emptyset$, then it is open. If $E = F^{-1}(E') \neq \emptyset$, let A be any point of E. Then $F(A) \in E'$ and there is a $B_\delta(A) \subset D$ with $F(B_\delta(A)) \subset E'$; that is $B_\delta(A) \subset F^{-1}(E') = E$, so E is open. On the other hand, if $E = F^{-1}(E') \neq \emptyset$ is open, then for any $A \in E$ there is a $B_\delta(A) \subset E$; then $F(B_\delta(A)) \subset F(E) = E'$. Thus, if $F^{-1}(E')$ is open for every open E' in W, it is clear that F satisfies the conditions to be continuous at A for any $A \in D$; that is, F is continuous in D.

1.2. Proposition (Properties of open sets). Let V be a vector space with scalar product and let the open subsets of V be defined by (A). Then

(i) the union of any number of open sets of V is open;

(ii) the intersection of a <u>finite</u> number of open sets of V is open;

(iii) the empty subset $\emptyset$ is open;

(iv) the set V itself is open.

<u>Proof</u>. Let $\mathfrak{D}$ denote a collection of open sets D of V. The union of $\mathfrak{D}$ is the set

$$\cup \mathfrak{D} = \{X \mid X \in D \text{ for some } D \in \mathfrak{D}\} .$$

Thus, if $X \in \cup \mathfrak{D}$, then $X \in D$ for some $D \in \mathfrak{D}$. Since D is open, there is a $B_\delta(X)$ such that $Y \in B_\delta(X)$ implies $Y \in D$ and therefore $Y \in \cup \mathfrak{D}$; that is, $B_\delta(X) \subset \cup \mathfrak{D}$. Thus (i) $\cup \mathfrak{D}$ is open. The intersection of $\mathfrak{D}$ is the set

$$\cap \mathfrak{D} = \{X \quad X \in D \text{ for every } D \in \mathfrak{D}\} .$$

If $\cap \mathfrak{D} = \emptyset$, as is surely the case if some $D = \emptyset$, then (ii) follows from (iii). If $X \in \cap \mathfrak{D} \neq \emptyset$, there is a $B_{\delta_D}(X) \subset D$ for each $D \in \mathfrak{D}$. If the number of sets in $\mathfrak{D}$ is finite, it is always possible to choose a δ so that $0 < \delta \leqq \delta_D$ for every $D \in \mathfrak{D}$; for example, take δ to be the smallest of the δ_D's. Then $B_\delta(X) \subset B_{\delta_D}(X) \subset D$ for every $D \in \mathfrak{D}$; that is, $B_\delta(X) \subset \cap \mathfrak{D}$. Thus (ii) $\cap \mathfrak{D}$ is open. Statements (iii) and (iv) follow directly from Definition (A).

<u>Remark</u>. To see that (ii) may fail if the restriction "finite" is omitted, consider the following example. Let $\mathfrak{D}$ be the collection of all balls $B_r(\tilde{0})$ with $r > 1$. Then

$$\cap \mathfrak{D} = \{X \mid |X| \leq 1\} .$$

In fact, if $|X| > 1$, then X cannot be an element of $B_r(\tilde{0})$ for any r with $1 < r \leq |X|$, so X cannot be an element of $\cap \mathfrak{D}$. Now let X be any vector such that $|X| = 1$. Then $X \in \cap \mathfrak{D}$ but it is clear that <u>every</u> ball $B_\delta(X)$ contains points Y with $|Y| > 1$, e.g. $Y = (1 + \delta/2)X$.

1.3. <u>Definition</u>. Let V be any set. A <u>topology</u> on V is defined by a collection $\mathfrak{J}$ of subsets of V which are called "open" sets, provided that the conditions (i) - (iv) of Proposition 1.2, now considered as axioms, are satisfied, viz.

(O1) the union of any number of sets of $\mathfrak{J}$ is a set of $\mathfrak{J}$;

(O2) the intersection of a finite number of sets of $\mathfrak{J}$ is a set of $\mathfrak{J}$;

(O3) $\emptyset \in \mathfrak{J}$;

(O4) $V \in \mathfrak{J}$.

A set V together with an assigned topology is a <u>topological space</u>.

<u>Remarks</u>. If we use a more general definition of union and intersection to include the union and intersection of any empty collection of subsets of V, then (O3) is included under (O1) and (O4) under (O2).

A vector space V with scalar product becomes a topological space if the open sets of V, that is, the elements of $\mathfrak{J}$, are defined by (A). We shall show that, if V is finite dimensional, this topology is independent of the particular choice

of scalar product. This could be done by showing that any two scalar products determine exactly the same open sets by way of Definition (A). Instead we shall give (Definition 1.6) a topology $\mathfrak{J}^*$ on any vector space V, which does not mention any scalar product explicitly, but which coincides, if V is finite dimensional, with the topology $\mathfrak{J}$ on V defined by means of a scalar product on V. Of course, if the topology is given without mention of scalar product, then the definition of continuity must be taken from the condition of Proposition 1.1 (stated explicitly in Definition 1.22) rather than from Definition (C). Thus continuity is a property of functions from one topological space to another topological space. Such functions are often called "maps", and the elements of a topological space are called "points".

In order to apply the new definition of continuity in the case of a function $F: D \longrightarrow W$, $D \subset V$, the set D must be considered as a topological space. The reason for requiring $D \subset V$ in Proposition 1.1 to be open is that in this case a topology on D can be defined by taking subsets of D to be open if and only if they are open sets of V. Then D is itself a topological space with the "induced topology", and $F: D \longrightarrow W$ is a map. If D is not open, it is still possible to define an induced topology on D, by means of a more elaborate definition (Definition 1.20). Then Proposition 1.1 is true without the assumption that D is open, provided that the "open" sets in D are those determined by the induced topology on D.

To see the effects of varying the choice of the topology on V, consider the following two extreme choices of $\mathfrak{J}$: (i) $\mathfrak{J}$ is as large as possible, that is, every subset of V is open; (ii) $\mathfrak{J}$ is as small as possible and consists of the two sets $\emptyset$ and V. In case (i), each subset consisting of a single point is open, and $\mathfrak{J}$ defines the <u>discrete topology</u> on V. If W is any topological space, then every map $F: V \longrightarrow W$ is continuous. This concept of continuity violates the intuitive notion of continuity exactly insofar as the concept of open set in V departs from the intuitive notion of open set as illustrated by Definition (A). On the other hand, if W has the topology corresponding to case (ii) and if V is any topological space, it is again true that every map $F: V \longrightarrow W$ is continuous. If W has more than one element, intuition fails in this case also, and the failure lies in the fact that there are no "small" open sets in W. In general, the finer the topology on V (the larger the choice of $\mathfrak{J}$ for V), the more functions $F: V \longrightarrow W$ are continuous (relative to the given $\mathfrak{J}$); the coarser the topology on W (the smaller the choice of $\mathfrak{J}$ for W), the more functions $F: V \longrightarrow W$ are continuous.

1.4. <u>Definition</u>. Let V be any set and let $\mathfrak{J}_0$ be a collection of subsets of V satisfying (O3) and (O4). Then $\mathfrak{J}_0$ may be enlarged to a collection $\mathfrak{J}$ of subsets of V satisfying (O1) - (O4) by first enlarging $\mathfrak{J}_0$ to a collection $\mathfrak{J}'$ by adjoining all intersections of a finite number of elements of

$\mathfrak{J}_0$ and then enlarging $\mathfrak{J}'$ to a collection $\mathfrak{J}$ by adjoining all unions of elements of $\mathfrak{J}'$. The resulting topology $\mathfrak{J}$ is called the <u>topology generated by</u> $\mathfrak{J}_0$, and is the coarsest topology containing $\mathfrak{J}_0$ (why?).

<u>Proof</u>. The only point to be checked in this construction is that no additional intersections need be added after the enlargement from $\mathfrak{J}'$ to $\mathfrak{J}$. For example, let $\mathfrak{D}$ and $\mathfrak{E}$ be collections of elements of $\mathfrak{J}'$; then $\cup\,\mathfrak{D}$ and $\cup\,\mathfrak{E}$ are in $\mathfrak{J}$, and for each $D \in \mathfrak{D}$ we have that D is the intersection of a finite number of elements of $\mathfrak{J}_0$ and that $E \in \mathfrak{E}$ is also the intersection of a finite number of elements of $\mathfrak{J}_0$. Now a general formula relating unions and intersections of subsets of V (proof omitted) is

$$(\cup\,\mathfrak{D}) \cap (\cup\,\mathfrak{E}) = \cup\,\mathfrak{F} ,$$

where $\mathfrak{F}$ is the collection of all elements of the form $D \cap E$, $D \in \mathfrak{D}$, $E \in \mathfrak{F}$. But $D \cap E$ is also the intersection of a finite number of elements of $\mathfrak{J}_0$, so $\cup\,\mathfrak{F}$ is already in $\mathfrak{J}$.

1.5. <u>Example</u>. If V is a vector space with scalar product, we may take $\mathfrak{J}_0$ to be the collection including $\emptyset$, V, and all balls $B_r(X)$ for $X \in V$ and for positive numbers r. Then every open set $D \subset V$ is the union of elements of $\mathfrak{J}_0$. For, if $D \neq \emptyset$, there is a ball $B_\delta(X) \subset D$ for each $X \in D$. Since $X \in B_\delta(X)$, we have

$$D \subset \cup_{X \in D} B_\delta(X) \subset D , \quad \text{or} \quad D = \cup_{X \in D} B_\delta(X) .$$

Further, any set D which can be obtained from $\mathfrak{J}_0$ by taking finite intersections or arbitrary unions of elements of $\mathfrak{J}_0$ is an open set, by Proposition 1.2. Thus the topology $\mathfrak{J}$ generated by the above $\mathfrak{J}_0$ coincides with the topology on V determined by Definition (A). Actually, $\mathfrak{J}_0$ may be taken even smaller: for example, $\mathfrak{J}_0$ includes $\emptyset$, V, and all balls $B_r(X)$ for $X \in V$ and for positive <u>rational</u> numbers r.

1.6. <u>Definition</u>. Let V be a vector space. We denote by $\mathfrak{J}^*$ the topology induced on V by the requirement that all elements of $V^* = L(V, R)$ be continuous maps. (Here the topology of R is assumed given and to be the topology determined by taking ordinary multiplication as scalar product.) That is, $\mathfrak{J}^*$ is the topology generated by taking $\mathfrak{J}_0^*$ to be the collection of all subsets of V of the form $E = T^{-1}(E')$ where $T \in V^*$ and E' is an open set in R. This topology will be called the <u>standard topology</u> on a vector space.

<u>Remark</u>. $\mathfrak{J}_0^*$ includes $\emptyset$, V since for any $T \in V^*$, we have $T^{-1}(R) = V$, $T^{-1}(\emptyset) = \emptyset$.

1.7. <u>Theorem</u>. If V is a finite dimensional vector space with scalar product, then the standard topology $\mathfrak{J}^*$ coincides with the topology $\mathfrak{J}$ determined on V by the scalar product and Definition (A).

<u>Proof</u>. Since $\mathfrak{J}^*$ is the smallest (coarsest) topology on V for which the elements of V^* are continuous and since the elements of V^* are continuous relative to the topology $\mathfrak{J}$ (Proposition VI, 1.7), it is clear that $\mathfrak{J} \supset \mathfrak{J}^*$. However, we

shall give an explicit proof, showing first that $\mathfrak{J}_o^* \subset \mathfrak{J}$. This is essentially the same as the proof of Proposition VI, 1.7. If $E \in \mathfrak{J}_o^*$, then $E = T^{-1}(E')$ for some $T \in V^*$ and for some open E' in R. Now there is a unique $B \in V$ such that $T: V \longrightarrow R$ is defined by $X \longrightarrow B \cdot X$. If $B = \tilde{0}$, then $E = T^{-1}(E')$ is V or $\emptyset$, according as E' contains the real number 0 or not, so $E \in \mathfrak{J}$. If $B \neq \tilde{0}$, we use the inequality

$$|TY - TX| \leq |B| \; |Y - X|$$

to show that for any $X \in E$ (therefore $TX \in E'$), the ball $B_\delta(X)$ is contained in E if $\delta < \varepsilon / |B|$, where ε is selected so that $B_\varepsilon(TX) \subset E'$. Thus $E \in \mathfrak{J}_o^*$ implies $E \in \mathfrak{J}$. Then any set which can be obtained from the elements of $\mathfrak{J}_o^*$ by taking finite intersections or arbitrary unions is an element of $\mathfrak{J}$, by Proposition 1.2; that is, $\mathfrak{J}^* \subset \mathfrak{J}$.

To show that $\mathfrak{J} \subset \mathfrak{J}^*$, which then implies $\mathfrak{J} = \mathfrak{J}^*$, it is clearly sufficient to show that $\mathfrak{J}_o \subset \mathfrak{J}^*$, where $\mathfrak{J}_o$ is the collection described in Example 1.5. We already have $\emptyset, V \in \mathfrak{J}^*$. The remaining elements of $\mathfrak{J}_o$ are of the form $B_r(A)$, for some $A \in V$ and some $r > 0$. For each $X \in B_r(A)$ there is a $B_\delta(X)$ with $B_\delta(X) \subset B_r(A)$. We shall show there is a set $E \in \mathfrak{J}^*$ with $X \in E$, $E \subset B_\delta(X) \subset B_r(A)$. Then $B_r(A)$ is the union of the E's constructed for $X \in B_r(A)$; that is, $B_r(A) \in \mathfrak{J}^*$.

Let $A_1, \ldots, A_n$ be an orthonormal basis for V, where $n = \dim V$, and let $T_i: X \longrightarrow A_i \cdot X$, $i = 1, \ldots, n$. Let

$S_{i\varepsilon}(\lambda) \in \mathfrak{J}_o^*$ denote the set $T_i^{-1}(B_\varepsilon(\lambda))$, $\lambda \in R$, $\varepsilon > 0$; that is, $S_{i\varepsilon}(\lambda) = \{Y | Y \in V$ and $T_iY \in B_\varepsilon(\lambda)\}$. Given $B_\delta(X)$, let E be the intersection of the n sets $S_{i\varepsilon}(T_iX)$, $i = 1,\ldots,n$, where $\varepsilon = \delta/n$. Then $E \in \mathfrak{J}^*$ and $X \in E$ (why?) and $E \subset B_\delta(X)$. In fact, if $Y \in E$, then $T_iY \in B_\varepsilon(T_iX)$, $i = 1,\ldots,n$, or $|T_iY - T_iX| < \varepsilon$. By Proposition III, 5.2,

$$Y = \Sigma_{i=1}^n (T_iY)A_i \ , \quad X = \Sigma_{i=1}^n (T_iX)A_i \ .$$

Then, by (v), (iv) of Theorem III, 3.2,

$$|Y - X| = |\Sigma_{i=1}^n T_i(Y - X)A_i| \leqq \Sigma_{i=1}^n |T_i(Y - X)A_i|$$

$$\leqq \Sigma_{i=1}^n |T_i(Y - X)| \cdot |A_i| < \Sigma_{i=1}^n \varepsilon \cdot 1 = n\varepsilon = \delta \ ;$$

that is, $Y \in E$ implies $Y \in B_\delta(X)$, or $E \subset B_\delta(X)$.

<u>Remarks</u>. As a result of Definition 1.6 and Theorem 1.7, we have a standard topology $\mathfrak{J}^*$ for all vector spaces V and, in the case of finite dimensional vector spaces, this topology coincides with the topology defined by any scalar product on V. In particular, if V and W are finite dimensional, continuity of maps $F: D \longrightarrow W$, D an open set in V, may be expressed in terms of any convenient scalar products on V and W.

An examination of the proof of Theorem 1.7 shows that a sufficient condition for the equivalence of two topologies $\mathfrak{J}_1$ and $\mathfrak{J}_2$ on a given set V is: for each $X \in V$ and any $E_i \in \mathfrak{J}_i$ such that $X \in E_i$, $i = 1$ or 2, there is a set $E_j \in \mathfrak{J}_j$, $j \neq i$, such that $X \in E_j$ and $E_j \subset E_i$. The particular

properties of the case considered in Theorem 1.7 were used only to demonstrate the existence of the required sets E_j.

In Definition II, 11.8 we constructed the direct sum of two vector spaces U and W considering the set $V = U \times W$ whose elements are pairs of vectors from U and W, and then defining the operations of addition and scalar multiplication on V so that V is also a vector space and so that the obvious projections $P_U: V \longrightarrow U$ and $P_W: V \longrightarrow W$ are linear transformations.

Analogously, if U and W are topological spaces we assign a topology to $U \times W$ such that the obvious projections are continuous maps, viz.

1.8. <u>Definition</u>. Let U and W be topological spaces, and let $U \times W$ denote the set whose elements are pairs (A, C), $A \in U$, $C \in W$. The <u>product topology</u> on $U \times W$ is the topology induced by the requirement that the projections $P_U: U \times W \longrightarrow U$ and $P_W: U \times W \longrightarrow W$, defined by $P_U(A, C) = A$, $P_W(A, C) = C$, be continuous maps. That is, $\mathfrak{J}$ is the topology generated by taking $\mathfrak{J}_o$ to be the collection of all subsets of $U \times W$ of the form $E = P_U^{-1}(E')$ where E' is an open set in U or of the form $E = P_W^{-1}(E'')$ where E'' is open in W.

<u>Remark</u>. To get "small" open sets in $U \times W$, one takes the intersection of two sets of $\mathfrak{J}_o$, one of each type.

1.9. <u>Examples</u>. If V is any vector space, the operations of addition and scalar multiplication define functions $V \times V \longrightarrow V$, $V \longrightarrow V$, and $R \times V \longrightarrow V$ by $(A, B) \longrightarrow A + B$, $A \longrightarrow -A$, and $(r, A) \longrightarrow rA$ respectively. If V and R

are topological spaces, and if $V \times V$ and $R \times V$ are assigned the product topology, then these functions are maps. A criterion for a "good" topology on a set for which algebraic operations are defined is that the maps induced by these operations should be continuous. It is left as an exercise to verify that the standard topology of Definition 1.6 for a vector space V makes the above maps continuous, and that for the usual topology on R the maps induced by ordinary addition and multiplication are continuous.

If U and W are vector spaces, then $U \oplus W$ may be assigned a topology in two ways: (i) the product topology induced from the standard topologies on U and W, and (ii) the standard topology on $U \oplus W$ determined by the fact that $U \oplus W$ is itself a vector space. Are these two topologies the same?

1.10. <u>Definition</u>. Let V and W be topological spaces. A bijective continuous map $F: V \longrightarrow W$ is called a <u>homeomorphism</u> if and only if the map F^{-1} (which is surely defined) is continuous.

<u>Example</u>. If V and W are vector spaces (with the standard topology) and if $T \in L(V, W)$ is an isomorphism, then T is also a homeomorphism (why?)

1.11. <u>Definition</u>. A map $F: V \longrightarrow W$ is called <u>open</u> if $F(E)$ is open in W for each open E in V.

<u>Example</u>. The projection $P_U: U \times W \longrightarrow U$ of Definition 1.8 is open as well as continuous (but not bijective unless W consists of a single point).

1.12. <u>Theorem</u>. A bijective continuous map $F: V \longrightarrow W$

is a homeomorphism if and only if it is open.

Proof. If $G: W \longrightarrow V$ denotes the map F^{-1}, then G is continuous if and only if $G^{-1}(E)$ is open in W for every open E in V, that is, if and only if $F(E)$ is open for every open E in V.

Remarks. Theorem 1.12 characterizes a homeomorphism $F: V \longrightarrow W$ as giving a one-one correspondence not only between points of V and W but also between the topologies of V and W; that is, as topological spaces V and W may be considered to be identical. If $F: V \longrightarrow W$ is merely bijective and continuous, but not a homeomorphism, then the points of V and W are in one-one correspondence, but the topology of V is finer (more open sets) than that of W.

1.13. Definition. On any topological space, the complement of an open set is called a closed set.

If V is any set, we denote the complement of a subset E of V by $E^{\#}$, that is

$$E^{\#} = \{X \mid X \in V \text{ and } X \notin E\} \quad .$$

We have $E \cup E^{\#} = V$ and $E \cap E^{\#} = \emptyset$, and the formulas $(E \cap F)^{\#} = E^{\#} \cup F^{\#}$, $(E \cup F)^{\#} = E^{\#} \cap F^{\#}$, $\emptyset^{\#} = V$, $V^{\#} = \emptyset$. These may be used with Definition 1.3 to give

1.14. Proposition. If $\mathfrak{J}$ defines a topology on V and if $\mathfrak{J}^{\#}$ denotes the collection consisting of the complements of the elements of $\mathfrak{J}$, that is, the closed sets on V, then

(C1) the intersection of any number of sets of $\mathfrak{J}^{\#}$ is a set of $\mathfrak{J}^{\#}$;

(C2) the union of a finite number of sets of $\mathfrak{J}^{\#}$ is a set of $\mathfrak{J}^{\#}$;

(C3) $V \in \mathfrak{J}^{\#}$;

(C4) $\emptyset \in \mathfrak{J}^{\#}$

Remarks. On any topological space V, the sets $\emptyset$ and V are both open and closed. If V has the discrete topology, every set is both open and closed. (It should be noted that an arbitrary subset of a topological space may be neither open nor closed.) Since $(E^{\#})^{\#} = E$, the complement of a closed set is open. In particular, any collection $\mathfrak{J}^{\#}$ of subsets of a set V which satisfies (C1) - (C4), taken as axioms, determines a topology $\mathfrak{J}$ on V by taking complements, and it is possible to define continuity, etc., entirely in terms of closed sets.

1.15. Definition. Let S be any subset of a topological space V, and let $A \in V$. Then A is called a contact point of S if and only if $S \cap D \neq \emptyset$ for every open set D in V with $A \in D$.

Remark. If $A \in S$, then A is surely a contact point of S since $A \in S \cap D$ for every open D with $A \in D$.

1.16. Proposition (Characterization of closed sets). A set S is closed if and only if it contains its contact points; that is,

(i) if S is closed, and if A is a contact point of S, then $A \in S$;

(ii) if there exists a contact point A of S with $A \notin S$, then S is not closed.

Proof. (i) Suppose that $A \notin S$, where S is closed. Then $A \in S^{\#}$ and $S^{\#}$ is an open set. But $S \cap S^{\#} = \emptyset$, so A cannot be a contact point of S. (ii) Suppose that A is a contact point of S with $A \notin S$. Then $A \in S^{\#}$. By hypothesis every open set D with $A \in D$ satisfies $S \cap D \neq \emptyset$, so no open set D with $A \in D$ can satisfy $D \subset S^{\#}$; that is, $S^{\#}$ is not open, so S is not closed.

1.17. Definition. Let S be any subset of a topological space V. The closure of S, denoted by $\overline{S}$, is the subset of V consisting of all contact points of S.

Remarks. Then $S \subset \overline{S}$ (why?) and $\overline{S}$ is closed. In fact, if D is an open set with $\overline{S} \cap D \neq \emptyset$, then $S \cap D \neq \emptyset$. (For, if $S \cap D = \emptyset$, then any point of $\overline{S}$ in $\overline{S} \cap D$ fails to be a contact point of S, contrary to the construction of $\overline{S}$.) That is, every contact point of $\overline{S}$ must be also a contact point of S, and therefore in $\overline{S}$. Thus, $\overline{S}$ is closed, by Proposition 1.16.

1.18. Proposition. The closure $\overline{S}$ of a subset S of a topological space V is the intersection of all closed sets E of V which satisfy $S \subset E$. Consequently, if $\mathcal{S}$ is any collection of subsets of V, we have $\overline{\cap \mathcal{S}} \subset \cap \overline{\mathcal{S}}$ where $\overline{\mathcal{S}}$ denotes the collection of the sets $\overline{S}$ for $S \in \mathcal{S}$.

The proof of these statements is left as an exercise.

1.19. Definition. A point which is a contact point of both S and $S^{\#}$ is called a boundary point of S. The boundary

of S, denoted by $b(S)$, is the subset of V consisting of all boundary points of S; that is,

$$b(S) = \overline{S} \cap \overline{S^{\#}} = b(S^{\#}) \quad .$$

[Clearly, $b(S)$ is closed.]

Examples. Let V be a finite dimensional vector space with scalar product, and let $D_1 = \{X \mid |X| < r\}$ where $r > 0$, $D_2 = \{X \mid |X| > r\}$. Then $(D_2)^{\#} = \{X \mid |X| \leqq r\} = \overline{D}_1$, $(D_1)^{\#} = \{X \mid |X| \geqq r\} = \overline{D}_2$, and $b(D) = \{X \mid |X| = r\}$ for $D = D_1, D_2, D_1^{\#}$, or $D_2^{\#}$.

If V has the topology in which the only open sets are $\emptyset$ and V, then $\overline{S} = V$ for any $S \neq \emptyset$, $b(S) = V$ if $S \neq \emptyset, V$. If V has the discrete topology, then $\overline{S} = S$, $b(S) = \emptyset$, for every subset S of V.

1.20. Definition. If S is any (non-empty) subset of a topological space V, the induced topology on S is the topology determined by the condition that the inclusion $\iota : S \longrightarrow V$ be continuous; that is, the open sets of S are the sets $S \cap D$ where D is an open set of V. Then S, with the induced topology, is called a subspace of V.

Remark. The closed sets of S must be the complements (in S) of the open sets of S, and are of the form $S \cap E$ where E is a closed set of V. This follows from the formulas

$$\emptyset = S \cap (E \cap E^{\#}) = (S \cap E) \cap (S \cap E^{\#}) \ ,$$

$$S = S \cap (E \cup E^{\#}) = (S \cap E) \cup (S \cap E^{\#}) \quad .$$

1.21. Proposition. A set D which is open in S is open in V if and only if S is open in V. A set E which is closed in S is closed in V if and only if S is closed in V.

Proof. The set $D = S$ must be open (and closed) in S, so cannot be open (or closed) in V unless S itself is open (or closed) in V. Conversely, if S is open in V, then $S \cap D$ is the intersection of two open sets of V, so is open in V. If S is closed in V, then $S \cap E$ is the intersection of two closed sets of V, so is closed in V.

1.22. Definition. Let S be any (non-empty) subset of a topological space V and let $F: S \longrightarrow W$ where W is a topological space. Then F is continuous in S if F is continuous relative to the induced topology on S, that is, if $F^{-1}(E')$ is an open subset of the subspace S for each open set $E' \subset W$. The function F is continuous at a point $A \in S$ if, for each open set $E' \subset W$ with $F(A) \in E'$, there exists an open set $E \subset S$ with $A \in E$ and $F(E) \subset E'$.

Remarks. Note that the case $S = V$ is not excluded in the above definition.

In Proposition 1.1, we demonstrated the equivalence of Definition (C) and Definition 1.22 $(D = S)$ in the case that the topological spaces V and W are vector spaces, with the topologies defined by means of a choice of scalar products on V and W, for the case that D is an open set in V. It is clear that this restriction on D can be dropped, provided D is considered as a subspace of V with the induced topology. In fact,

for $A \in D$ with D arbitrary, the subset $\{X | X \in D$ and $|X - A| < \delta\}$ appearing in Definition (C) is $D \cap B_\delta(A)$ rather than $B_\delta(A)$ (for sufficiently small δ) as in the case of an open D. But $D \cap B_\delta(A)$ is an open set in D for the induced topology on D, and the same proof holds, once we have verified that the appropriate analogue of the property appearing in Definition (A) can be used to characterize the open sets for the induced topology on D (cf. Exercise 5.2).

If V and W are topological spaces, and S is a subset of V, particular care must be taken to distinguish between a function $F: V \longrightarrow W$ and the function on S, usually denoted also by F, obtained by restricting F to S. (The restriction can be more accurately described as the composite function $F\iota$, where $\iota : S \longrightarrow V$.) If S is not open, the function F obtained by restriction may very well be continuous in S when the given function F (on V) is not continuous at some point of S. For example, if V and W are vector spaces, it can happen that for some $A \in S \subset V$ and some open set $E' \subset W$, the condition $F(S \cap B_\delta(A)) \subset E'$ can be satisfied for some $B_\delta(A)$, with $F(B_\delta(A))$ not contained in E' for any choice of $\delta > 0$ (see Exercise 5.7).

§2. Hausdorff spaces

The notion of contact point is intuitively related to the notion of limit point, but is not sufficiently discriminating for a usable definition of limit. In the first place, the condition

of Definition 1.15 (for $A \in V$ to be a contact point of $S \subset V$) is trivial for points $A \in S$, and is satisfied even if $S \cap D$ reduces to the single point A (in which case A is called an <u>isolated point</u> of S). Further, if A is a contact point of S, two different possibilities exist: either (i) $S \cap D \neq \emptyset$ is a finite subset of S for <u>some</u> open D with $A \in D$, in which case $S \cap D'$ is also finite for all open D' with $D' \subset D$; or (ii) $S \cap D \neq \emptyset$ is an infinite set for <u>every</u> open D with $A \in D$. (Case (ii) cannot occur if S is itself a finite set.)

If we wish to exclude the trivial case $S \cap D = A$ for some open set D, we may consider the sets $A^{\#} \cap S \cap D$ in place of $S \cap D$. We then have the following possibilities: (i_1) $A^{\#} \cap S \cap D = \emptyset$ for some open D with $A \in D$; (i_2) $A^{\#} \cap S \cap D$ is a finite subset of S for some open D with $A \in D$, but $A^{\#} \cap S \cap D' \neq \emptyset$ for every open $D' \subset D$ with $A \in D'$; or (ii) $A^{\#} \cap S \cap D$ is an infinite set for every open D with $A \in D$. In case (i_1) we say that A is not a limit point of S. (If $A \in V$ fails even to be a contact point of S, then A satisfies (i_1).) The possibility (i_2) is a permanent obstacle to a decent theory of limits, but cannot be eliminated if V is an arbitrary topological space; that is, there exist topological spaces, as defined in Definition 1.3, for which the possibility (i_2) can be realized. We shall consider topological spaces V for which the topology $\mathfrak{I}$ satisfies an additional axiom, viz.

(H) If X and Y are distinct points of V, there exist open sets D and D' with $X \in D$, $Y \in D'$, and $D \cap D' = \emptyset$;

that is, any two points of V can be "separated" by open sets.

2.1. Definition. A topological space for which the axiom (H) is satisfied is called a Hausdorff space.

2.2. Proposition. Any finite dimensional vector space V, with the standard topology of Definition 1.6, is a Hausdorff space.

Proof. Because of Theorem 1.7, the topology of V is computable in terms of any scalar product on V. If we choose a scalar product on V, then $|X - Y| > 0$ for any given distinct points X and Y, by (ii) of Theorem III, 3.2, say $|X - Y| = 3\delta > 0$. Then $B_\delta(X) \cap B_\delta(Y) = \emptyset$, using Corollary III, 3.3 (iii).

Remark. A vector space which is not finite dimensional, whose topology is defined by means of a scalar product, is also a Hausdorff space.

2.3. Proposition. Any subspace of a Hausdorff space is also a Hausdorff space.

This follows directly from the definition of the open sets in a subspace.

2.4. Proposition. If V is a Hausdorff space, and $A \in V$, then

(a) the set consisting of the point A alone is closed;

(b) $\bigcap \mathcal{E} = A$, where $\mathcal{E}$ is the collection of all closed sets E with $A \in E$;

(c) $\bigcap \mathcal{D} = A$, where $\mathcal{D}$ is the collection of all open sets D with $A \in D$;

(d) $\cap \bar{\mathfrak{D}} = A$, where $\bar{\mathfrak{D}}$ is the collection of the sets $\bar{D}$, for $D \in \mathfrak{D}$ above.

Proof. (a) Let X be any point of $A^{\#}$. Then (H) implies that there is an open set D with $X \in D$, $A \notin D$; that is, $D \subset A^{\#}$. Then $A^{\#}$ is open (Exercise 5.1), so A is closed. (b) Since A is closed, $A = \bar{A}$; but $\bar{A} = \cap \mathcal{E}$ by Proposition 1.18. (c) If $X \neq A$, then by (H) there is an open set D with $A \in D$, $X \notin D$; that is, $X \notin \cap \mathfrak{D}$. (d) If $X \neq A$, then by (H) there are open D, D' with $A \in D$, $X \in D'$ and $D \cap D' = \emptyset$. Then $X \notin \bar{D}$ (why?); that is, $X \notin \cap \bar{\mathfrak{D}}$.

Remarks. The collection $\bar{\mathfrak{D}}$ is a subcollection of $\mathcal{E}$. Properties (a), (b), (c) are weaker than the axiom (H); that is, these properties cannot be used as axioms in place of (H). Property (d) is equivalent to (H). In fact, suppose (d) holds for any $A \in V$, and let $A \neq B$ be distinct points of V. Then $B \notin \cap \bar{\mathfrak{D}}$ so there is an open D such that $A \in D$, $B \notin \bar{D}$. Then $B \in \bar{D}^{\#}$ which is an open set, and $D \cap \bar{D}^{\#} \subset \bar{D} \cap \bar{D}^{\#} = \emptyset$.

Returning to the discussion of limit points, we see that case (i_2) cannot occur if V is a Hausdorff space. [Actually, property (c) of Proposition 2.4 is sufficient to eliminate (i_2), but the full strength of (H) will be needed, at least for the range of a function, to prove the uniqueness of "the" limit in Definition 2.7.] With (i_2) eliminated, case (ii) can be restated as (ii') $A^{\#} \cap S \cap D \neq \emptyset$ for every open set D with $A \in D$; for if $A^{\#} \cap S \cap D$ is a finite set for some D, then there is a $D' \subset D$ with $A \in D'$, $A^{\#} \cap S \cap D = \emptyset$. Thus,

2.5. <u>Definition</u>. Let V be a Hausdorff space and let S be a subset of V. A point $A \in V$ is called a <u>limit point</u> of S if and only if $A^{\#} \cap S \cap D \neq \emptyset$ for every open D in V with $A \in D$.

<u>Remarks</u>. By comparing the definition of contact point and limit point, we see that any contact point of S which is not a limit point of S must be an isolated point of S and therefore an element of S. Thus Proposition 1.16 may be stated as: a subset S of a Hausdorff space is closed if and only if it contains its limit points.

If $\mathfrak{D}$ is the collection of open sets D with $A \in D$, we have seen that $\cap \mathfrak{D} = A$. Let $\dot{\mathfrak{D}}$ be the collection of sets of the form $A^{\#} \cap D$, with $D \in \mathfrak{D}$; then the sets of $\dot{\mathfrak{D}}$ are open, and $\cap \dot{\mathfrak{D}} = \emptyset$. Thus if A is any point of V, then $S \cap (\cap \dot{\mathfrak{D}}) = \emptyset$; but if A is a limit point of S, then $S \cap (\cap \dot{\mathfrak{D}}') \neq \emptyset$ for any finite subcollection $\dot{\mathfrak{D}}'$ of $\dot{\mathfrak{D}}$, since $\cap \dot{\mathfrak{D}}'$ is also of the form $A^{\#} \cap D$ for some open D with $A \in D$, by (O2) of Definition 1.3. A stronger statement will be needed later, viz.

2.6. <u>Proposition</u>. Let V be a Hausdorff space, $S \subset V$. If $A \in \overline{S}$, $A \notin S$, then $S \cap (\cap \overline{\mathfrak{D}}) = \emptyset$ (where $\overline{\mathfrak{D}}$ denotes the sets $\overline{D}$ for $D \in \mathfrak{D}$, and $\mathfrak{D}$ is the collection of all open sets D with $A \in D$), but $S \cap (\cap \overline{\mathfrak{D}}') \neq \emptyset$ for any finite subcollection $\overline{\mathfrak{D}}'$ of $\overline{\mathfrak{D}}$.

<u>Proof</u>. $A \notin S$ implies $S \cap A = \emptyset$, and (d) of Proposition 2.4 implies $\cap \overline{\mathfrak{D}} = A$. Thus $S \cap (\cap \overline{\mathfrak{D}}) = \emptyset$. On the other

hand, $A \in \overline{S}$ implies $S \cap (\bigcap \mathfrak{D}') \neq \emptyset$ for any finite subcollection $\mathfrak{D}'$ of $\mathfrak{D}$. But $\overline{D} \supset D$, so $S \cap (\bigcap \overline{\mathfrak{D}'}) \supset S \cap (\bigcap \mathfrak{D}') \neq \emptyset$ for the corresponding subcollection $\overline{\mathfrak{D}'}$ of $\overline{\mathfrak{D}}$.

2.7. <u>Definition</u>. Let V be a Hausdorff space and let S be a subset of V. Let $F: S \longrightarrow W$ where W is a Hausdorff space. Let A be a limit point of S, with $A \notin S$, and let $B \in W$. Then F has the <u>limit</u> B as $X \in S$ tends to A ($\lim_A F = B$) if, for each open $\tilde{D}$ in W with $B \in \tilde{D}$, there is an open D in V with $A \in D$ such that $F(A^{\#} \cap S \cap D) \subset \tilde{D}$.

<u>Remarks</u>. Since $A \notin S$, the set $A^{\#} \cap S \cap D = S \cap D$. The condition "$A$ is a limit point of S" ensures that the sets $A^{\#} \cap S \cap D$ are not empty; otherwise the condition of Definition 2.7 could be satisfied trivially without implying anything about the given function F. If F has the limit B, the value B must satisfy $B \in \overline{F(S)}$: the condition of Definition 2.7 cannot possibly be satisfied if there is a $\tilde{D}$ with $\tilde{D} \cap F(S) = \emptyset$, $B \in \tilde{D}$. However, B can be an isolated point of $F(S)$, e.g. if F is the constant function. If V, W are finite dimensional vector spaces with the standard topology or, more generally, are vector spaces with topologies defined by means of scalar products, the above definition coincides with Definition (B) quoted at the beginning of §1. For if scalar products are given on V and W, then $B_\delta(A)$ is an open set in V, and $A^{\#} \cap S \cap B_\delta(A) = \{X \mid X \in S$ and $0 < |X - A| < \delta\}$; moreover, every open $\tilde{D}$ in W with $B \in \tilde{D}$ contains an open set $\tilde{B}_\varepsilon(B)$ for some choice of $\varepsilon > 0$.

It is not necessary to assume that V is Hausdorff in

Definitions 2.5 and 2.7; property (c) of Proposition 2.4 would be sufficient (cf. Exercise 5.13). However, it is essential that the range W of F have the Hausdorff property in order to have

2.8. Proposition. The limit in Definition 2.7 is unique.

Proof. If $B_1 \neq B_2$, then there are open sets $\tilde{D}_1$ and $\tilde{D}_2$ in W with $B \in \tilde{D}_1$, $B_2 \in \tilde{D}_2$ and $\tilde{D}_1 \cap \tilde{D}_2 = \emptyset$. If we suppose that the condition of Definition 2.7 is satisfied for $B = B_1$ and for $B = B_2$, there are open sets D_1 and D_2 of V with $A \in D_1 \cap D_2$ such that $F(A^{\#} \cap S \cap D_1) \subset \tilde{D}_1$ and $F(A^{\#} \cap S \cap D_2) \subset \tilde{D}_2$. But then $F(A^{\#} \cap S \cap (D_1 \cap D_2)) \subset \tilde{D}_1 \cap \tilde{D}_2 = \emptyset$, which is impossible since $A^{\#} \cap S \cap (D_1 \cap D_2) \neq \emptyset$ by hypothesis.

2.9. Proposition. Let $F: S \longrightarrow W$ be continuous. Let A be a limit point of S, $A \notin S$, and suppose that $\lim_A F$ exists. Then the function $\overline{F}: S \cup A \longrightarrow W$, defined by $\overline{F}(X) = F(X)$ for $X \in S$ and $\overline{F}(A) = \lim_A F$, is continuous.

2.10. Proposition. Let $F: S \longrightarrow W$. Then F is continuous (Definition 1.22) at $A \in S$ if and only if the limit of F, as $X \in A^{\#} \cap S$ tends to A, exists and equals $F(A)$. (Here A is assumed not to be an isolated point of S.)

The proofs of the above two propositions are left as exercises.

2.11. Definition. A collection $\mathfrak{U}$ of subsets of a set V is called a covering of a subset S of V if $S \subset \cup \mathfrak{U}$. If V is a topological space, the covering is called open (or closed) if all the sets in $\mathfrak{U}$ are open (or closed).

Remarks. A covering of $S = V$ satisfies $V = \cup \mathfrak{U}$.

The above definition does not exclude the possibility that $\emptyset \in \mathfrak{U}$. In applications, it is usual to consider "a covering by non-empty open sets", etc.

2.12. Definition. A Hausdorff space V is called compact if it has the following property:

(K) For every open covering of V, there is a finite subcovering of V; that is, if $V = \bigcup \mathfrak{D}$ where $\mathfrak{D}$ is a collection of open sets, then there is a finite subcollection $\mathfrak{D}'$ of $\mathfrak{D}$ such that $V = \bigcup \mathfrak{D}'$.

By taking complements we see that an equivalent property is:

($K^{\#}$) For every collection $\mathcal{E}$ of closed sets of V such that $\bigcap \mathcal{E} = \emptyset$, there is a finite subcollection $\mathcal{E}'$ of $\mathcal{E}$ such that $\bigcap \mathcal{E}' = \emptyset$.

2.13. Definition. A subset S of a Hausdorff space V is called a compact set if the subspace S is compact.

Remark. If $S = \bigcup \mathfrak{D}_S$ is a covering of S by open sets in S, then each $D_S \in \mathfrak{D}_S$ is of the form $S \cap D$ where D is an open set in V, and the sets D form a collection $\mathfrak{D}$ with $S \subset \bigcup \mathfrak{D}$. If $S = \bigcup \mathfrak{D}_S'$ for a finite subcollection $\mathfrak{D}_S'$ of $\mathfrak{D}_S$, then $S \subset \bigcup \mathfrak{D}'$ for the corresponding subcollection $\mathfrak{D}'$ of $\mathfrak{D}$. Thus the condition that a subspace S of V be compact is:

(K_S) For every collection $\mathfrak{D}$ of open sets of V satisfying $S \subset \bigcup \mathfrak{D}$, there is a finite subcollection $\mathfrak{D}'$ of $\mathfrak{D}$ such that $S \subset \bigcup \mathfrak{D}'$,

or, taking complements in S and using Proposition 1.21,

($K_S^{\#}$) For every collection $\mathcal{E}$ of closed sets of V satisfying $S \cap (\bigcap \mathcal{E}) = \emptyset$, there is a finite subcollection $\mathcal{E}'$ of $\mathcal{E}$ such that $S \cap (\bigcap \mathcal{E}') = \emptyset$.

2.14. Proposition. A compact set is closed.

Proof. Let S be a compact set. If there were a point $A \in \overline{S}$ with $A \notin S$, then the collection $\mathcal{E} = \overline{\mathcal{D}}$ constructed in Proposition 2.6 would contradict $K_S^{\#}$.

2.15. Proposition. Any infinite subset of a compact set has at least one limit point.

Proof. If P is any subset of a compact set S, then $\overline{P} \subset S$ by Propositions 1.18 and 2.14. If P has no limit point, then each $X \in S$ fails to be a limit point; that is, for each $X \in S$ there is an open set D_X with $X \in D_X$ such that $X^{\#} \cap P \cap D_X = \emptyset$. (Then $P \cap D_X$ is either X or $\emptyset$.) Since $X \in D_X$, we have $S \subset \bigcup_{X \in S} D_X$ and therefore, since S is compact, $S \subset \bigcup_{X \in Q} D_X$ where Q is a finite subset of S. Then $P \subset Q$ and P is a finite subset of S.

2.16. Definition. Let V be a finite dimensional vector space. A subset S of V is called bounded if, for any choice of scalar product on V, there exists a number M such that $|X| \leq M$ for all $X \in S$.

Remarks. The actual value of M depends on the choice of the scalar product, but the existence of such a number M (for a given set S) does not. For suppose there is an M, such that

$|X| \leqq M$ for $X \in S$, for a particular choice of scalar product. Let $A_1, \ldots, A_n$ be an orthonormal basis for V, $n = \dim V$. For any $X \in V$ we have $X = \Sigma_{i=1}^n x^i A_i$, $|X| = \sqrt{\Sigma_{i=1}^n (x^i)^2}$. For $X \in S$ we have $|X| \leqq M$ and therefore $(x^i)^2 \leqq M^2$, $i = 1, \ldots, n$, or $|x^i| \leqq M$. If we compute in terms of another scalar product, we have, for $X \in S$,

$$|X| \leqq \Sigma_{i=1}^n |x^i| \, |A_i| \leqq M \, \Sigma_{i=1}^n |A_i| = M\alpha ,$$

where $\alpha = \Sigma_{i=1}^n |A_i|$, and $|A_i|$ is the length of the basis vector A_i in terms of the new scalar product. That is, if M satisfied the condition of Definition 2.16 for the first choice of scalar product, then $M\alpha$ (or any larger number) will provide a bound for S for the new choice of scalar product.

<u>Examples</u>. Any ball $B_r(\dot{X})$ defined in terms of some choice of scalar product on V is bounded. Any finite subset of V is bounded. Any subset of a bounded set is bounded.

From here on, through Theorem 2.24, it will be assumed without explicit statement that, if V is a finite dimensional vector space, then (i) $\dim V = n$, (ii) the topology on V is the standard topology of Definition 1.6, (iii) that a scalar product has been chosen on V. The real numbers R will also be assumed to have the standard topology.

2.17. <u>Proposition</u>. A compact set S of a finite dimensional vector space V is bounded.

<u>Proof</u>. Let δ be any positive number. Then $S \subset \cup_{X \in S} B_\delta(X)$. If S is compact, there is a finite subset Q

of S such that $S \subset \cup_{X \in Q} B_\delta(X)$. Let $\beta = \max |X|$ for $X \in Q$. Then $|X| \leqq \delta + \beta$ for all $X \in S$; that is, S is bounded. In fact, $Y \in B_\delta(X)$ implies $|Y| \leqq \delta + |X|$, by III, 3.3 (iii).

By combining Propositions 2.14 and 2.17, we see that a compact set of a finite dimensional vector space must be both closed and bounded. These properties characterize the compact sets of a finite dimensional vector space; that is,

2.18. <u>Theorem</u>. A subset of a finite dimensional vector space is compact if and only if it is both closed and bounded.

The necessity of these properties has already been shown; the sufficiency will be shown below (Theorem 2.24) and depends essentially on (i) the fact that the topology can be defined in terms of a distance function and Definition (A) of §1, and (ii) the following property of the real numbers R :

2.19. <u>Theorem</u>. A bounded infinite set of real numbers has at least one limit point.

Proof. If $S \subset R$ and $|X| \leqq M$ for $X \in S$, then S is contained in the interval $I_0 = \{\lambda | \lambda \in R \text{ and } -M \leqq \lambda \leqq M\}$. Let $I_0' = \{\lambda | \lambda \in R \text{ and } -M \leqq \lambda \leqq 0\}$ and $I_0'' = \{\lambda | \lambda \in R \text{ and } 0 \leqq \lambda \leqq M\}$. Then $I_0 = I_0' \cup I_0''$ so at least one of these two sub-intervals must contain an infinite number of points of S, since S is assumed to be an infinite set. Let I_1 be either I_0' or I_0'', selected so that I_1 contains an infinite number of points of S. Next consider the two sub-intervals I_1' and I_1'' of I_1 determined by the midpoint of I_1; for example, if $I_1 = I_0''$, then $I_1' = \{\lambda | \lambda \in R \text{ and } 0 \leqq \lambda \leqq M/2\}$ and

$I_1'' = \{\lambda | \lambda \in R$ and $M/2 \leqq \lambda \leqq M\}$; then $I_1 = I_1' \cup I_1''$. Let I_2 be either I_1' or I_1'', selected so that I_2 contains an infinite number of points of S. This process may be continued indefinitely, and we obtain a sequence of intervals $I_0, I_1, I_2, \ldots,$ each containing an infinite number of points of S, with $I_m \subset I_k$ for $m > k$ and the length of I_k is $2M/2^k = M/2^{k-1}$. If $I_k = \{\lambda | \lambda \in R$ and $a_k \leqq \lambda \leqq b_k\}$, the a_k's form a bounded monotone increasing sequence of real numbers, which therefore has a limit, say a, as $k \longrightarrow \infty$, and $a_k \leqq a$ for all k. The b_k's form a bounded monotone decreasing sequence of real numbers, which therefore has a limit, say b, as $k \longrightarrow \infty$, and $b \leqq b_k$ for all k. Since $b_k - a_k = M/2^{k-1} \longrightarrow 0$, we must have $a = b$. Let A denote the point $a = b$. Then $A \in I_k$ for every k. Further, A is a limit point of S. In fact, any open D with $A \in D$ contains $B_\delta(A)$, for some $\delta > 0$, and therefore contains I_k for $M/2^{k-1} < \delta$, and therefore contains an infinite number of points of S.

2.20. <u>Theorem</u>. A bounded infinite set of a finite dimensional vector space has at least one limit point.

<u>Proof</u>. Let $A_1, \ldots, A_n$ be an orthonormal basis for the finite dimensional vector space V, and suppose $|X| \leqq M$ for $X \in S \subset V$, where S is an infinite subset of V. Every $X \in V$ is expressible in the form $X = \Sigma_{i=1}^n x^i(X) A_i$, where $x^i(X) = X \cdot A_i$. Then $|x^i(X)| \leqq M$ for $X \in S$, $i = 1, \ldots, n$. The set of real numbers $x^1(X)$, for example, for $X \in S$ forms a bounded set of real numbers, but not necessarily an infinite one if $n > 1$

(although the set $x^i(X)$ must be infinite for some choice of i, since S is assumed to be an infinite set). Consequently, theorem 2.19 cannot be applied directly. However, the method of proof of Theorem 2.19 can be adapted to the present case as follows. Let $I_o = \{X | X = \Sigma_{i=1}^n x^i(X)A_i$ and $-M \leq x^i(X) \leq M,\ i = 1,\ldots,n\}$. Then $S \subset I_o$. We consider the 2^n subsets of I_o obtained by choosing either $[-M, 0]$ or $[0, M]$ as the range of the coefficients $x^i(X)$, $i = 1,\ldots,n$, for points X in the subset. We then choose one of these subsets to be I_1, such that I_1 contains an infinite number of points of S. Then we consider the 2^n subsets of I_1 defined in the same way as above, etc. Thus we construct a sequence $I_o, I_1, I_2,\ldots$ of subsets of V, each containing an infinite number of points of S, with $I_m \subset I_k$ for $m > k$. If $I_k = \{X | X = \Sigma_{i=1}^n x^i(X)A_i$ and $a_k^i \leq x^i(X) \leq b_k^i$, $i = 1,\ldots,n\}$, then $b_k^i - a_k^i = M/2^{k-1}$. For each i, the monotonic sequences a_k^i and b_k^i determine a common limit a^i. Then the point $A = \Sigma_{i=1}^n a^i A_i \in I_k$ for each k. Further, A is a limit point of S. In fact, any open D with $A \in D$ contains $B_\delta(A)$ for some $\delta > 0$, and therefore contains I_k for $M/2^{k-1} < \delta/\sqrt{n}$, and therefore contains an infinite number of points of S.

2.21. <u>Construction</u>. Let S be any non-empty set in a finite dimensional vector space V, and let δ be any positive number. Let X_1 be any point of S; let X_2 be any point of S with $|X_2 - X_1| \geq \delta$; ... ; let X_k be a point of S with $|X_k - X_i| \geq \delta$, $i = 1,\ldots,k-1$; ... .

(i) If the construction is not possible after choosing

$X_1, \ldots, X_p$, then every point X of S satisfies $|X - X_i| < \delta$ for some choice of $i = 1, \ldots, p$, and the set S is bounded (cf. the proof of Proposition 2.17).

(ii) If the construction yields an infinite set $P = \{X_1, X_2, \ldots\}$, $X_i \in S$ (which is not possible if S is finite), then the set P can have no limit point. In fact, for any $A \in V$, the set $P \cap B_\delta(A)$ consists of a single point, or is empty.

Now Theorem 2.20 excludes the possibility (ii) if S is bounded, since the subset P of S would also be bounded. That is,

2.22. <u>Proposition</u>. The construction of 2.21 is finite (case (i)) if and only if the set S is bounded.

2.23. <u>Proposition</u>. Let S be a closed bounded set of a finite dimensional vector space V. If $S \subset \bigcup \mathfrak{D}$, where $\mathfrak{D}$ is a collection of open sets of V, then there exists a positive number δ (depending on $\mathfrak{D}$) such that, for each $X \in S$, $B_\delta(X) \subset D$ for some $D \in \mathfrak{D}$.

<u>Proof</u>. If no δ satisfies the conditions of the proposition, then for every δ there is an exception and, in particular, for each $\delta = 1, 1/2, \ldots, 1/m, \ldots$, there is a point X_m such that $B_{1/m}(X_m)$ is not contained in any $D \in \mathfrak{D}$. The points X_m need not be all distinct, but they form an infinite subset P of S. (Each $X \in S$ satisfies $X \in D$ for some $D \in \mathfrak{D}$; since D is open, there is some $B_\delta(X) \subset D$, and this implies $X \neq X_m$ for $1/m < \delta$). By Theorem 2.20, the set P must have a limit point, say A, since S is bounded. Then A is also a limit point of

S and $A \in S$ since S is closed. Now $A \in D$ for some $D \in \mathfrak{D}$, since $\mathfrak{D}$ gives a covering of S, and there is a $\delta > 0$ such that $B_\delta(A) \subset D$ since D is open. Take m so large that $1/m < \delta/2$. Since A is a limit point of P, the set $P \cap B_{1/m}(A)$ cannot be empty or even finite, so there is a $k > m$ such that $X_k \in B_{1/m}(A)$. Then for $Y \in B_{1/k}(X_k)$ we have $|Y - A| \leq |Y - X_k| + |X_k - A| \leq 1/k + 1/m < \delta$; that is, $B_{1/k}(X_k) \subset B_\delta(A) \subset D$, where $D \in \mathfrak{D}$, which is contrary to the construction of X_k.

2.24. <u>Theorem</u>. If S is a closed bounded set in a finite dimensional vector space V, then for any open covering of S there is a finite subcovering of S. That is, S is compact.

<u>Proof</u>. If $S = \emptyset$, the proposition is trivial, so we suppose $S \neq \emptyset$. Let $\mathfrak{D}$ be a collection of open sets of V such that $S \subset \cup \mathfrak{D}$. By Proposition 2.23, there is a positive δ such that $B_\delta(X) \subset D$ for each $X \in S$ for some $D \in \mathfrak{D}$. By Proposition 2.22, any construction of the type 2.21 for this δ yields a finite set $Q = \{X_1, \ldots, X_p\}$ such that each $X \in S$ satisfies $X \in B_\delta(X_i)$ for some choice of $i = 1, \ldots, p$; that is, $S \subset \cup_{i=1}^p B_\delta(X_i)$. Now each $B_\delta(X_i)$ is contained in some $D \in \mathfrak{D}$, so we may choose a finite subcollection $\mathfrak{D}'$ of $\mathfrak{D}$ such that $B_\delta(X_i)$ is contained in some $D \in \mathfrak{D}'$, $i = 1, \ldots, p$. Then $S \subset \cup_{i=1}^p B_\delta(X_i) \subset \cup \mathfrak{D}'$.

2.25. <u>Theorem</u>. Let V and W be Hausdorff spaces, and let $S \subset V$ be compact. If $F: S \longrightarrow W$ is continuous, then

$F(S) \subset W$ is compact.

Proof. Let $\tilde{\mathfrak{D}}$ be any open covering of $F(S)$. Since F is continuous, $D = F^{-1}(\tilde{D})$ is open in S for each $\tilde{D} \in \tilde{\mathfrak{D}}$. The sets D obtained in this way give an open covering $\mathfrak{D}$ of S. Since S is compact, a finite subcovering $\mathfrak{D}'$ of S can be selected from $\mathfrak{D}$. Let $\tilde{\mathfrak{D}}'$ be the finite subcollection of $\tilde{\mathfrak{D}}$ determined by the condition $\tilde{D} \in \tilde{\mathfrak{D}}'$ if $D = F^{-1}(\tilde{D}) \in \mathfrak{D}'$. Then $\tilde{\mathfrak{D}}'$ is a covering of $F(S)$.

§3. Some theorems in analysis

The set R of real numbers will always be assumed to have the standard topology. If $F: V \longrightarrow W$ is assumed to be continuous, it is assumed that $V \neq \emptyset$ and W are topological spaces (V may be a subspace of another topological space).

3.1. Proposition. Let $F: V \longrightarrow W$ be continuous, where W is a Hausdorff space, and let $B \in W$. Then (i) the set of points X of V for which $F(X) \neq B$ is open, and (ii) the set of points X of V for which $F(X) = B$ is closed.

Proof. Since W is a Hausdorff space, the set consisting of the point B alone is closed, and $B^{\#}$ is open. (i) Since F is continuous, $F^{-1}(B^{\#})$ is open. Statement (ii) then follows from $F^{-1}(B) = (F^{-1}(B^{\#}))^{\#}$.

If we take $W = R$ and $B = 0$, then

3.2. Corollary. Let $F: V \longrightarrow R$ be continuous, and let the set $S \subset V$ be determined by the equation $F(X) = 0$; that is, $X \in S$ if and only if $F(X) = 0$. Then S is closed.

3.3. Corollary. If $F: V \longrightarrow R$ is continuous and $F(A) \neq 0$, then there is an open set D in V with $A \in D$ such that $F(X) \neq 0$ for all $X \in D$. If $F(A) > 0$ and $0 < \varepsilon < F(A)$, then D can be chosen so that $F(X) > \varepsilon$ for all $X \in D$.

3.4. Definition. Let $F: V \longrightarrow W$ where W is a finite dimensional vector space, and let $S \subset V$ (the case $S = V$ not excluded). Then F is said to be bounded on S if the set $F(S) \subset W$ is bounded; that is, if for a given scalar product on W there exists a number μ such that $|F(X)| \leq \mu$ for all $X \in S$. For the case $W = R$, the function is said to be bounded above on S if the set $F(S) \subset R$ is bounded above, that is, if there is a number α such that $F(X) \leq \alpha$ for $X \in S$; F is said to be bounded below on S if the set $F(S) \subset R$ is bounded below, that is, if there is a number β such that $\beta \leq F(X)$ for $X \in S$. If F is bounded below on S, then the set $F(S) \subset R$ has a greatest lower bound β_0, which is called greatest lower bound of F on S (abbreviated g.l.b. or inf). It is easily checked that $\beta_0 \in \overline{F(S)}$. If $\beta_0 \in F(S)$, it is called the minimum (value) of F on S. If F is bounded above on S, then the least upper bound α_0 of F on S (abbreviated l.u.b. or sup) is defined and lies in $\overline{F(S)}$. If $\alpha_0 \in F(S)$, it is called the maximum (value) of F on S.

3.5. Theorem. A real-valued function which is continuous on a compact set is bounded and attains its maximum and minimum values.

Proof. If $F: S \longrightarrow R$ is continuous and S is compact,

then $F(S)$ is compact by Theorem 2.25, and therefore closed and bounded by Theorem 2.18. Since $\overline{F(S)} = F(S)$, we have $\alpha_0 \in F(S)$, $\beta_0 \in F(S)$; that is, there exists at least one point $X_1 \in S$ with $F(X_1) = \alpha_0$ and at least one point $X_2 \in S$ with $F(X_2) = \beta_0$.

3.6. <u>Theorem</u>. Let $F: S \longrightarrow W$ be continuous on $S \subset V$ where V and W are finite dimensional vector spaces. If S is compact, then F is <u>uniformly continuous</u> on S; that is, if continuity is computed by means of the ε, δ Definition (C) of §1, for any choice of scalar products on V and W, and $\varepsilon > 0$ is given, there exists a $\delta > 0$ such that $F(S \cap B_\delta(X)) \subset B_\varepsilon(F(X))$ for every $X \in S$.

<u>Proof</u>. Let $\varepsilon > 0$ be given. F is continuous, so for each $X \in S$ there is a $\delta_X > 0$ such that $F(S \cap B_{\delta_X}(X)) \subset B_{\varepsilon/2}(F(X))$. Since $S \subset \cup_{X \in S} B_{\delta_X/2}(X)$ and S is compact, there is a finite subset Q of S such that $S \subset \cup_{A \in Q} B_{\delta_A/2}(A)$. Let δ be chosen $\leq \delta_A/2$ for all $A \in Q$. For any point $X \in S$ there is some $A \in Q$ such that $X \in B_{\delta_A/2}(A)$. Then $B_\delta(X) \subset B_{\delta_A}(A)$ for this choice of A. In fact, if $Y \in B_\delta(X)$, then

$$|Y - A| \leq |Y - X| + |X - A| < \delta + \delta_A/2 \leq \delta_A \quad .$$

Further, $F(S \cap B_\delta(X)) \subset F(S \cap B_{\delta_A}(A)) \subset B_{\varepsilon/2}(F(A))$, so for $Y \in B_\delta(X)$, we have

$$|F(Y) - F(X)| \leq |F(Y) - F(A)| + |F(A) - F(X)| < \varepsilon/2 + \varepsilon/2 = \varepsilon \ ;$$

that is, $F(S \cap B_\delta(X)) \subset B_\varepsilon(F(X))$.

§4. The inverse and implicit function theorems

In the case of a linear transformation T from one vector space to another, we know (Proposition II, 7.4) that, if T is bijective, then T^{-1} is also linear. Moreover, if V and W are finite dimensional, with $\dim V = \dim W$, it is a relatively simple matter to determine, from an explicit expression of a given T in terms of bases for V and W, whether or not T is bijective and even the explicit expression of T^{-1}.

In §1, we noted that for the case of continuous functions F from one topological space to another it need not follow, from the fact that F is bijective, that F^{-1} is continuous. The situation is not improved by taking the topological spaces to be finite dimensional vector spaces, for which we have considerably more information; nor is it a simple matter to determine whether a particular function F, given explicitly in terms of bases, is injective or bijective, etc. Here we can consider a "better" class of functions, by supposing F to be continuously differentiable, and still not have a result about F^{-1} in general. However, in this case a simple (and computable) criterion can be given which will enable us to make a "local" statement about F^{-1}.

In Chapter VIII we considered a continuous map $F: D \longrightarrow W$, where $D \subset V$ is open and V, W are finite dimensional vector spaces. If F is continuously differentiable, that is, if $F'(X, Y)$ exists for each $X \in D$, $Y \in V$, and is continuous as a function of X for each fixed Y, we defined the derivative F' of F to be the function $F': D \longrightarrow L(V, W)$

where $F'(X)Y = F'(X, Y)$.

4.1. <u>Theorem</u>. Let $F: D \longrightarrow W$ be continuously differentiable, where $D \subset V$ is open, and V, W are vector spaces of dimension n. Let $A \in D$. If $F'(A): V \longrightarrow W$ is non-singular, then there is an open set $D_o \subset D$ with $A \in D_o$ such that $F(D_o) \subset W$ is open and $F: D_o \longrightarrow F(D_o)$ is a homeomorphism.

<u>Remarks</u>. The above result is an existence theorem: there exists an open set D_o for which the conclusion holds (an actual D_o can be exhibited for any particular choice of F and of A such that $F'(A) \neq \tilde{0}$ but this D_o may be unnecessarily small) and $F^{-1}: F(D_o) \longrightarrow D_o$ exists and is continuous (but there is no way of giving an explicit formula for F^{-1} in general). Actually, F^{-1} is continuously differentiable on $F(D_o)$ as well (see Theorem 4.5). Note that no hypothesis is needed that $F: D_o \longrightarrow F(D_o)$ is bijective. This is a conclusion following from the assumptions that F is continuously differentiable and that $F'(A)$ is non-singular.

Recall that for any continuously differentiable map $F: D \longrightarrow W$, $D \subset V$, and for a given choice of a basis $A_1, \ldots, A_k$ for V and a basis $B_1, \ldots, B_n$ for W, any $X \in D$ may be represented in the form $X = \Sigma_{j=1}^{k} x^j A_j$ and $F(X) \in W$ in the form $F(X) = \Sigma_{i=1}^{n} f^i(X) B_i$, where each component function $f^i(X) = f^i(x^1, \ldots, x^k)$ is continuously differentiable in the variables $x^1, \ldots, x^k$ for values which correspond to points of D. For $Y = \Sigma_{j=1}^{k} y^j A_j \in V$, the linear transformation $F'(X): V \longrightarrow W$ is given by

$$F'(X)Y = \Sigma_{i=1}^{n} \Sigma_{j=1}^{k} y^j \frac{\partial f^i}{\partial x^j} B_i$$

and the matrix $(\partial f^i/\partial x^j)$ of this representation is called the Jacobian matrix of F (Definition VIII, 1.8) relative to the given choices of bases for V and W.

If $k = n$, we can define a function $J: D \longrightarrow R$ associated with this representation by taking $J(X)$ to be the determinant of the Jacobian matrix $(\partial f^i/\partial x^j)$ at X. The function J is continuous in D, since J is a polynomial in the continuous functions $\partial f^i/\partial x^j$. The linear transformation $F'(X)$ at X is non-singular if and only if $J(X) \neq 0$ (Proposition IX, 9.4). (This property is independent of the particular representation considered, although the actual value of $J(X)$ depends on the representation.) Thus, given an explicit representation of a continuously differentiable F in terms of component functions $f^i(X)$, $X \in D$, it is sufficient to compute $J(A)$ in order to know if the hypothesis of Theorem 4.1 is satisfied.

The conclusion of the theorem can also be expressed as a result about component functions. Let $A = (x_o^1, \ldots, x_o^n)$, and $F(A) = (z_o^1, \ldots, z_o^n)$. If the functions $f^i(x^1, \ldots, x^n)$, $i = 1, \ldots, n$, are continuously differentiable in an open set containing $(x_o^1, \ldots, x_o^n)$, viz. D, and if $J(x_o^1, \ldots, x_o^n) \neq 0$, then there is an open set $F(D_o)$ containing $(z_o^1, \ldots, z_o^n)$ and there are n functions $g^j(z^1, \ldots, z^n)$, $j = 1, \ldots, n$, which are continuous (actually continuously differentiable) on $F(D_o)$ and which satisfy

$$f^i(g^1(z^1,\ldots,z^n),\ldots,g^n(z^1,\ldots,z^n)) = z^i \ , \qquad i = 1,\ldots,n \ ,$$

$$g^j(f^1(x^1,\ldots,x^n),\ldots,f^n(x^1,\ldots,x^n)) = x^j \ , \qquad j = 1,\ldots,n \ ,$$

for $(z^1,\ldots,z^n) \in F(D_o)$ or $(x^1,\ldots,x^n) \in D_o$.

<u>Proof</u>. The proof will be divided into three sections: (I) Selection of a suitable open set D_o; (II) Proof that $F: D_o \longrightarrow W$ is injective (therefore $F: D_o \longrightarrow F(D_o)$ is bijective); and (III) Proof that $F: D_o \longrightarrow W$ is open (then, in particular, $F(D_o)$ is open in W and $F: D_o \longrightarrow F(D_o)$ is a homeomorphism, by Proposition 1.21 and Theorem 1.12).

The entire proof will be carried out in terms of a fixed representation of F determined by a choice of a basis $A_1,\ldots,A_n$ for V and a basis $B_1,\ldots,B_n$ for W, with notations as above. We begin by giving three lemmas which will be used below.

4.2. <u>Lemma</u>. Let $A_1,\ldots,A_n$ be a basis for V. For $X \in V$, define

$$\|X\| = \max_j \ |x^j| \ , \qquad\qquad X = \Sigma_{j=1}^n \ x^j A_j \ .$$

Then (cf. Theorem III, 3.2 and Corollary III, 3.3)

(i) For each $X \in V$, $\|X\| \geqq 0$;

(ii) If $X \in V$, then $\|X\| = 0$ if and only if $X = \tilde{0}$;

(iii) For each $X \in V$ and $\lambda \in R$, $\|\lambda X\| = |\lambda| \ \|X\|$;

(iv) For each pair $X, Y \in V$, $\|X + Y\| \leqq \|X\| + \|Y\|$.

In particular, for $X = A + B$, $Y = -B$, we have $\|A + B\| \geqq \|A\| - \|B\|$. Further, we have

(v) $|X| \leqq \sqrt{n} \ \|X\|$, $\|X\| \leqq |X|$, where $|X|$ is determined from

the scalar product on V defined by taking $A_1,\ldots,A_n$ as an orthonormal basis on V.

The proof of this lemma is left as an exercise.

4.3. Lemma. The topology on V generated by $\emptyset$, V, and the sets

$$P_r(A) = \{X \mid X \in V \text{ and } \|X - A\| < r\}, \quad A \in V,\ r > 0,$$

is the same as the standard topology on V. In particular, if the analogous $\| \ \|$ on W is defined with respect to a basis $B_1,\ldots,B_n$ for W, then open sets, limits, and continuity can be computed according to Definitions (A), (B), and (C) of §1, with $\| \ \|$ in place of $| \ |$.

Proof. By (v) of Lemma 4.2, we have

$$B_r(A) \subset P_r(A), \qquad P_r(A) \subset B_{\sqrt{n}r}(A)$$

We then apply Exercise 5.1.

Remarks. Then we also have that the set $\{X \mid X \in V$ and $\|X - A\| < r\}$ is open for any choice of A and $r > 0$, that the set $\{X \mid X \in V$ and $\|X - A\| \leq r\}$ is compact, that $\|X\|$ is a continuous function of X, etc.

4.4. Lemma. Let S, T be elements of $L(V, W)$, $\dim V = \dim W = n$, with matrix representations (a^i_j) and (b^i_j) respectively, relative to a fixed choice of bases $A_1,\ldots,A_n$ for V and $B_1,\ldots,B_n$ for W. Let $|a^i_j - b^i_j| \leq M$, $i, j = 1,\ldots,n$. Suppose T is non-singular and let $U = T^{-1}$ be represented by (c^i_j) with $\max_{i,j} |c^i_j| = M'$ (then $M' > 0$ since $U \neq \tilde{0}$. Then

$$\|SY - TY\| \leq n^2 MM' \|TY\| , \qquad \text{for all } Y \in V .$$

<u>Proof</u>. If $Y = \Sigma_{j=1}^n y^j A_j \in V$, then

$$\begin{aligned} \|SY - TY\| &= \| \Sigma_{i=1}^n \Sigma_{j=1}^n y^j (a_j^i - b_j^i) B_i \| \\ &= \max_i |\Sigma_{j=1}^n y^j (a_j^i - b_j^i)| \\ &\leq nM \max_j |y^j| = nM \|Y\| . \end{aligned}$$

In particular,

$$\|Y\| = \|UTY\| = \|\tilde{O}TY - UTY\| \leq nM' \|TY\| .$$

Combining these two results, we obtain the required estimate.

(I) Now take $T = F'(A)$ which is non-singular by hypothesis (so, for our given choice of bases, $M' > 0$ is a given number), and $S = F'(X)$ for $X \in D$. Then $a_j^i = \partial f^i/\partial x^j|_X$ and $b_j^i = \partial f^i/\partial x^j|_A$. Since D is open, there is a δ' such that $B_{\delta'}(A) \subset D$; and since J is continuous and $J(A) \neq 0$, there is a δ'' such that $J(X) \neq 0$ for $X \in B_{\delta''}(A)$, by Corollary 3.3. Since $\partial f^i/\partial x^j$ is continuous in D, for any given $\varepsilon > 0$ we can choose $\delta_j^i > 0$ with $\delta_j^i \leq \min(\delta', \delta'')$ so that $\left|\partial f^i/\partial x^j|_X - \partial f^i/\partial x^j|_A\right| < \varepsilon$ for $|X - A| < \delta_j^i$ and therefore for $\|X - A\| < \delta_j^i/\sqrt{n}$. If we set $\delta = \min_{i,j}(\delta_j^i/\sqrt{n})$, then for $\|X - A\| < \delta$ we have $X \in D$, $J(X) \neq 0$, and

$$\|F'(X)Y - F'(A)Y\| \leq n^2 \varepsilon M' \|F'(A)Y\| , \qquad \text{for all } Y \in V .$$

Now let $\eta = 1 - n^2 \varepsilon M'$. If ε is large, η may be

negative, but for ε sufficiently small, $\eta > 0$. For $\varepsilon_0 = 1/n^2M' > 0$ we have $\eta = 0$, so we can certainly choose $\delta > \eta$ for $\varepsilon = \varepsilon_0$. As ε is chosen still smaller, η increases, and the choice of δ corresponding to ε may decrease, but for ε only slightly smaller than ε_0 it is still possible to choose $\delta \geqq \eta > 0$, so that we may then take $\delta = \eta$. Now set

$$D_0 = \{X \mid X \in V \text{ and } \|X - A\| < \eta\} \quad .$$

Then D_0 is open, $D_0 \subset D$, $A \in D_0$, and for all $X \in D_0$ we have $J(X) \neq 0$, that is, $F'(X)$ non-singular, and

$$(1_A) \qquad \|F'(X)Y - F'(A)Y\| \leqq (1 - \eta)\|F'(A)Y\| \ , \quad \text{for all } Y \in V \quad .$$

<u>Remarks</u>. It is clear that if a given value of η can be chosen, any η' with $0 < \eta' < \eta$ would do also, since $1 - \eta' > 1 - \eta$. Consequently, if $X_1 \in D_0$, we have $J(X_1) \neq 0$, and we can repeat the above discussion with A replaced by X_1, to show that we can choose η_1 sufficiently small that $D_1 = \{X \mid X \in V \text{ and } \|X - X_1\| < \eta_1\} \subset D_0$ with $X_1 \in D_1$, while for $X \in D_1$, we have $J(X) \neq 0$ and

$$(1_{X_1}) \qquad \|F'(X)Y - F'(X_1)Y\| \leqq (1 - \eta_1)\|F'(X_1)Y\| \ , \quad \text{for all } Y \in V \quad .$$

This result will be used in (III).

To show that a choice of η as described above is possible, we may argue as follows. If we suppose that the choice of η is not possible, then for every $\eta > 0$ and corresponding $\varepsilon = (1 - \eta)/n^2M'$, there is no choice of $\delta \geqq \eta$ so that

$\left| \partial f^i/\partial x^j|_X - \partial f^i/\partial x^j|_A \right| < \varepsilon$ for all X satisfying $\|X - A\| < \delta$, for all $i, j = 1,\ldots,n$. In particular, for $\eta = 1/m \longrightarrow 0$, $m = 1, 2,\ldots,$ with $\varepsilon_m = (1 - 1/m)/n^2M' \longrightarrow \varepsilon_o$, the choice $\delta = 1/m$ will not do, so there is a point X_m with $\|X_m - A\| < 1/m$ such that we have $\left|\partial f^i/\partial x^j|_{X_m} - \partial f^i/\partial x^j|_A\right| \geq \varepsilon_m$ for some choice of i, j. Since there are only a finite number of functions $\partial f^i/\partial x^j$, some function must provide the exception for an infinite number of values of m. Now for m sufficiently large, $\varepsilon_m \geq \varepsilon_o/2 > 0$, so that this particular $\partial f^i/\partial x^j$ cannot satisfy $\lim_A \partial f^i/\partial x^j = \partial f^i/\partial x^j|_A$. This is a contradiction, since $\partial f^i/\partial x^j$ is continuous, by hypothesis.

(II) Now let $X, Z \in D_o$ and set $Y = Z - X$. Then $L: [0, 1] \longrightarrow V$, defined by $L(t) = X + tY$, $0 \leq t \leq 1$, is a curve in D_o (why?) and $G = FL: [0, 1] \longrightarrow W$ is a curve in W given by $G(t) = F(X + tY)$. Then $G'(t) = F'(X + tY)Y$, by Exercise VIII, 3.6, or by computing explicitly from $G(t) = F(X + tY) = \Sigma_{i=1}^n f^i(X + tY)B_i$:

$$G'(t) = \Sigma_{i=1}^n \frac{d}{dt} f^i(X + tY)B_i$$

$$= \Sigma_{i=1}^n \Sigma_{j=1}^n \frac{\partial f^i}{\partial x^j}(X + tY)y^jB_i = F'(X + tY)Y \quad .$$

Now $\int_o^1 \frac{d}{dt} f^i(X + tY)dt = f^i(X + Y) - f^i(X)$, $i = 1,\ldots,n$, so if we integrate the component functions of the vector equation $G'(t) = F'(X + tY)Y$ with respect to t between $t = 0$ and $t = 1$, we obtain the vector equation

$$(2_A)\qquad \begin{aligned} F(X+Y) - F(X) &= \int_0^1 F'(X+tY)Y\,dt \\ &= \int_0^1 F'(A)Y\,dt + \int_0^1 [F'(X+tY) - F'(A)]Y\,dt\ . \end{aligned}$$

Then, for X and $X + Y$ in D_0, we have

$$(3_A)\qquad \begin{aligned} \|F(X+Y) - F(X)\| &\geqq \|\textstyle\int_0^1 F'(A)Y\,dt\| - \|\int_0^1 [F'(X+tY) - F'(A)]Y\,dt\| \\ &\geqq \|F'(A)Y\| - (1-\eta)\|F'(A)Y\| \\ &= \eta\|F'(A)Y\|\ . \end{aligned}$$

(To get the second estimate, we use the fact that if $U(t) = \Sigma_{i=1}^n u^i(t)dt$, and if $\|U(t)\| \leqq M$ for $0 \leqq t \leqq 1$, then $\|\int_0^1 U(t)dt\| \leqq M$, with equality if U does not depend on t, so $M = |u^i|$ for some i. These statements follow from the estimate

$$|\textstyle\int_0^1 u^i(t)dt| \leqq \int_0^1 |u^i(t)|dt \leqq \int_0^1 M\,dt = M\ ,$$

which holds for each $i = 1,\ldots,n$.)

If we suppose that $F(X) = F(Z) = F(X+Y)$, then (3_A) implies $\|F'(A)Y\| = 0$ since $\eta > 0$, and therefore $F'(A)Y = \tilde{0}$. But then $Y = \tilde{0}$ since $F'(A)$ is non-singular. That is, $X = Z$. Thus F is injective on D_0.

(III) To show that $F: D_0 \longrightarrow W$ is an open map, we must show that for any open $\tilde{D} \subset D_0$, the set $F(\tilde{D})$ is open. To do this, it is sufficient, by Lemma 4.3, to show that for any $Z_1 \in F(\tilde{D})$ there is an $\varepsilon > 0$ such that the open set

$$P_\varepsilon(Z_1) = \{Z \mid Z \in W \text{ and } \|Z - Z_1\| < \varepsilon\} \subset F(\tilde{D})\ .$$

Let X_1 be the point in $\tilde{D} \subset D_0$ such that $F(X_1) = Z_1$. By the remark following (I), there is an $\eta_1 > 0$ such that

$$D_1 = \{X \mid X \in V \text{ and } \|X - X_1\| < \eta_1\} \subset D_0$$

and such that for $X \in D$, $Y \in V$, we have (1_{X_1}). Then by (2_{X_1}), we have, for $X_1 + Y \in D_1$,

$$(3_{X_1}) \qquad \|F(X_1 + Y) - F(X_1)\| \geqq \eta_1 \|F'(X_1)Y\| \quad .$$

Let β denote the minimum value of $\|F'(X_1)Y\|$ as Y varies on the compact set $\|Y\| = \eta_1/2$. Then $\beta > 0$, since $F'(X_1)$ is non-singular. We choose $\varepsilon = \eta_1\beta/2$. For any Z with $\|Z - Z_1\| < \varepsilon$, we shall show that there is a point $X \in \tilde{D}$ with $F(X) = Z$; that is, $Z \in F(\tilde{D})$.

For fixed Z, the function $\|F(X) - Z\|$ is continuous on the compact set $\|X - X_1\| \leqq \eta_1/2$ and attains its minimum value for some X_2 with $\|X_2 - X_1\| \leqq \eta_1/2$, by Theorem 3.5. Then

$$\|F(X) - Z\| \geqq \|F(X_2) - Z\| \ , \qquad \text{for } \|X - X_1\| \leqq \eta_1/2 \quad .$$

We shall show that $F(X_2) = Z$. We note that the minimum value of $\|F(X) - Z\|$ for $\|X - X_1\| \leqq \eta_1/2$ must be smaller than ε since

$$\|F(X_1) - Z\| = \|Z_1 - Z\| < \varepsilon \quad .$$

Consequently, X_2 must satisfy

$$\|X_2 - X_1\| < \eta_1/2 \quad .$$

In fact, for $\|Y\| = \eta_1/2$, we have

$$\|F(X_1 + Y) - Z\| = \|F(X_1 + Y) - F(X_1) + Z_1 - Z\|$$

$$\geqq \|F(X_1 + Y) - F(X)_1)\| - \|Z_1 - Z\|$$

$$\geqq \eta_1\beta - \varepsilon = \varepsilon ,$$

by (3_{X_1}) and the choice of β and ε.

If $F(X_2) - Z = B \neq \tilde{0}$, we can reach a contradiction as follows. Since $F'(X_1)$ is non-singular, there is a Y_o such that $F'(X_1)Y_o = B$. Then for $\mu > 0$ sufficiently small, the point $X = X_2 - \mu Y_o$ satisfies $\|X - X_1\| \leqq \eta_1/2$ and

$$\|F(X) - Z\| < \|F(X_2) - Z\| ,$$

contrary to the choice of X_2.

In fact, from (2_{X_1}), since

$$\int_o^1 F'(X_1)(-\mu Y_o)dt = -\mu F'(X_1)Y_o ,$$

and from (1_{X_1}), we have

$$\|F(X_2 - \mu Y_o) - F(X_2) + \mu F'(X_1)Y_o\|$$

$$= \|\int_o^1 [F'(X_2 - t\mu Y_o) - F'(X_1)](-\mu Y_o)dt\|$$

$$\leqq (1 - \eta_1)\|F'(X_1)(-\mu Y_o)\| = (1 - \eta_1)\mu\|B\|$$

On the other hand,

$$\|F(X_2) - Z + F'(X_1)(-\mu Y_o)\| = \|B - \mu B\| = (1 - \mu)\|B\|$$

Then

$$\|F(X_2 - \mu Y_o) - Z\|$$

$$= \|F(X_2 - \mu Y_o) - F(X_2) + \mu F'(X_1)Y_o + F(X_2) - Z - \mu F'(X_1)Y_o\|$$

$$\leqq (1 - \eta_1)\mu\|B\| + (1 - \mu)\|B\| = (1 - \eta_1\mu)\|B\|$$

$$< \|B\| = \|F(X_2) - Z\| \quad .$$

4.5. Theorem. The inverse function in Theorem 4.1 is continuously differentiable.

Proof. In Theorem 4.1, starting from a continuously differentiable map $F: D \longrightarrow W$, where $D \subset V$ is open, with derivative $F': D \longrightarrow L(V, W)$, we have shown that, if $F'(A)$ is non-singular, then there is an open set $D_o \subset D$ with $A \in D_o$ such that $F: D_o \longrightarrow F(D_o) = \tilde{D}_o$ is a homeomorphism, with $F'(X)$ non-singular for $X \in D_o$. In particular, $G = F^{-1}: \tilde{D}_o \longrightarrow D_o$ is defined and continuous, and satisfies $G(\tilde{X}) = X \in D_o$ if and only if $F(X) = \tilde{X} \in D_o$. We must show that $G': \tilde{D}_o \longrightarrow L(W, V)$ exists and is continuous in $\tilde{X} \in \tilde{D}_o$ for each fixed $\tilde{Y} \in W$.

Now we already have a continuous function $H: \tilde{D}_o \longrightarrow L(W, V)$ defined as follows: $H(\tilde{X})\tilde{Y} = Y \in V$ if and only if $F'(X)Y = \tilde{Y}$ for $X = G(\tilde{X})$. To see that this function is defined, note that $\tilde{X}$ determines X uniquely and that $F'(X)$ is non-singular. To see that $H(\tilde{X})$ is continuous, we return to the particular representation of F used in the proof of Theorem 4.1. If we write

$$H(\tilde{X})\tilde{Y} = \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} \tilde{y}^j h_j^i(\tilde{X}) A_i ,$$

it is clear that the matrix $(h_j^i(\tilde{X}))$ is the inverse of the matrix $(\partial f^i/\partial x^j(X))$. That is, $h_j^i(\tilde{X})$ is a polynomial in the continuous functions $\partial f^i/\partial x^j(X)$, divided by the polynomial $J(X) \neq 0$, and is obviously continuous as a function of $X = G(\tilde{X})$ and therefore as a function of $\tilde{X}$ (Exercise 5.3).

To show that $G' = H$, we must show that $\lim_{h \to 0} \frac{1}{h}[G(\tilde{X} + h\tilde{Y}) - G(\tilde{X})]$ exists and is $H(\tilde{X})\tilde{Y}$ for each $\tilde{X} \in \tilde{D}_0$, $\tilde{Y} \in W$, or

$$\lim_{h \to 0} \frac{1}{h}[G(\tilde{X} + h\tilde{Y}) - G(\tilde{X}) - hH(\tilde{X})\tilde{Y}] = 0 \quad .$$

Let $X = G(\tilde{X})$, $Y = H(\tilde{X})\tilde{Y}$; then $F(X) = \tilde{X}$, $F'(X)hY = h\tilde{Y}$, but $F(X + hY) \neq \tilde{X} + h\tilde{Y}$ in general. For h sufficiently small, $\tilde{X} + h\tilde{Y} \in \tilde{D}_0$, and we write $G(\tilde{X} + h\tilde{Y})$ in the form $X + hY_h \in D_0$. Then we must show that, for any $\varepsilon > 0$,

$$\|X + hY_h - X - hY\| = |h| \, \|Y_h - Y\| \leqq |h| \varepsilon$$

for sufficiently small values of $|h|$.

Since $F'(X)$ exists, for any $\mu > 0$ we have

$$\|F(X + hY) - F(X + hY_h)\| = \|F(X + hY) - \tilde{X} - h\tilde{Y}\|$$

$$= \|F(X + hY) - F(X) - hF'(X)Y\| \leqq |h|\mu$$

for sufficiently small $|h|$. We shall take $\mu = \varepsilon\eta/nM'$ where M' is the maximum value of $|h_j^i(A)|$ in the matrix representation

of $H(A) = (F'(A))^{-1}$, and $\eta > 0$ is the value which was used in determining D_o. By (3_A) we have

$$\|F(X + hY_h) - F(X + hY)\| = \|F(X + hY + h(Y_h - Y)) - F(X + hY)\|$$

$$\geqq \eta\|F'(A)h(Y_h - Y)\| \geqq \frac{\eta|h|}{nM'} \|Y_h - Y\| ,$$

the last equality following as in the proof of Lemma 4.4. Combining these two estimates, we have

$$|h| \ \|Y_h - Y\| \leqq |h|\mu nM'/\eta = |h| \ \varepsilon$$

for $|h|$ sufficiently small, as was to be shown.

4.6. Definition. Let $F: D \longrightarrow W$ be continuously differentiable, where $D \subset V$ is open. Let $k = \dim V$ and $n = \dim W$. For each $X \in D$, define

$$r(X) = k - \dim (\ker F'(X)) \leqq k \quad .$$

By Theorem II, 2.5, we also have

$$\dim (\operatorname{im} F'(X)) = r(X) \leqq n \quad .$$

The integer $r(X)$ is called the rank of F at X.

Remarks. For Theorems 4.1 and 4.5, we considered the case $k = n$. This is the only case for which it is possible to have $F'(X)$ non-singular, that is, $\dim (\ker F'(X)) = 0$ and $\dim (\operatorname{im} F'(X)) = n$.

If $\operatorname{im} F'(A)$ contains r, but not more than r, linearly independent vectors, then the Jacobian matrix of F at

A, corresponding to any choice of bases in V and W, must contain at least one submatrix of order r whose determinant does not vanish, and no submatrix of order $r + 1$ can have a non-vanishing determinant. (That is, $r(A)$ coincides with the ordinary rank of the Jacobian matrix of F at A.)

If we select a submatrix of order r with non-vanishing determinant at A, then this same submatrix has non-vanishing determinant on an open set D_o with $A \in D_o$, since the entries in the submatrix are continuous functions. This implies $r(X) \geq r(A)$ for $X \in D_o$. In general, however, we cannot exclude the possibility $r(X) > r(A)$ for some $X \in D_o$ for every open D_o containing A; there may exist a submatrix of order $r(A) + 1$ whose determinant vanishes at A but does not vanish identically on any open set D_o containing A.

4.7. <u>Definition</u>. A continuously differentiable map $F: D \longrightarrow W$ is said to be <u>regular</u> at $A \in D$ if there is an open set $D_o \subset D$ with $A \in D_o$, such that $r(X) = r(A)$ for $X \in D_o$.

<u>Remark</u>. The map F is surely regular if $r(X) = k$ or $r(X) = n$. Thus, regularity is an additional condition only in the case that $r(X) < k$ <u>and</u> $r(X) < n$.

4.8. <u>Theorem</u>. Let $F: D \longrightarrow W$ be continuously differentiable, where $D \subset V$ is open, and V and W are vector spaces of dimension k and n respectively. Let A be a fixed point of D and let the rank of F at A be r. If $r < k$ and $r < n$, assume further that F is regular at A. Then there is an open set $D_o \subset D$ with $A \in D_o$ such that $F: D_o \longrightarrow W$

can be given in the form $F = EP_1\hat{F}$, where (i) $\hat{F}$ is a homeomorphism of D_o with a product $D_1 \times D_2$, with D_1 an open set in a vector space of dimension r, and D_2 an open set in a vector space of dimension $k - r$; (ii) $P_1: D_1 \times D_2 \longrightarrow D_1$ is the usual projection; and (iii) $E: D_1 \longrightarrow W$ is injective; and (iv) $\hat{F}$, P_1, and E are continuously differentiable. That is, $F: D_o \longrightarrow W$ is given by the composition

$$D_o \xrightarrow{\hat{F}} D_1 \times D_2 \xrightarrow{P_1} D_1 \xrightarrow{E} W \quad .$$

If $r = k$, then $D_2 = \tilde{0}$ and P_1 is a homeomorphism. If $r = n$, then E is a homeomorphism of D_1 with $F(D_o) \subset W$.

<u>Proof</u>. We choose a fixed basis $A_1,\ldots,A_k$ for V and a fixed basis $B_1,\ldots,B_n$ for W and suppose that the submatrix of the Jacobian matrix of F which corresponds to the indices $i, j = 1,\ldots,r$ has non-vanishing determinant at A (this is always possible, for example, by changing the order in which the basis elements are listed). Let V_1 denote the vector space spanned by $A_1,\ldots,A_r$, and V_2 the vector space spanned by $A_{r+1},\ldots,A_k$ (where $V_2 = \tilde{0}$ if $r = k$). Then $V = V_1 \times V_2$, and any point of V may be considered as a pair (U, X) where

$$U = \Sigma_{j=1}^{r} u^j A_j, \quad X = \Sigma_{m=1}^{k-r} x^m A_{r+m} \quad .$$

Similarly, $W = W_1 \times W_2$ consists of pairs (Z, T), with

$$Z = \Sigma_{i=1}^{r} z^i B_i, \; T = \Sigma_{\ell=1}^{n-r} t^\ell B_{r+\ell}$$

(where $W_2 = \tilde{0}$ if $r = n$). If $p_1: W \longrightarrow W_1$ and $p_2: W \longrightarrow W_2$

are the usual projections, then $F_1 = p_1F: D \longrightarrow W_1$ and $F_2 = p_2F: D \longrightarrow W_2$ are continuously differentiable (Exercise 5.4). In terms of the component functions of F, we have

$$(F_1) \qquad z^i = f^i(U, X) , \qquad i = 1,\ldots,r ,$$

$$(F_2) \qquad t^\ell = f^{r+\ell}(U, X) , \qquad \ell = 1,\ldots,n - r .$$

We define Z_o, T_o by $F(A) = (F_1(A), F_2(A)) = (Z_o, T_o)$, $A = (U_o, X_o)$. Since the matrix $(\partial z^i/\partial u^j)$ has non-vanishing determinant at A (and therefore also on an open set containing A), the rank of F_1 at A is r and A is a regular point for F_1 in any case.

Let $\hat{V}$ be the vector space $W_1 \times V_2$ of dimension k, and define $\hat{F}: D \longrightarrow \hat{V}$ by $\hat{F}(U, X) = (F_1(U, X), P_2(U, X))$, where $P_2: V_1 \times V_2 \longrightarrow V_2$ is the usual projection. For example, $\hat{F}(A) = (F_1(U_o, X_o), P_2(X_o)) = (Z_o, X_o)$. Now the Jacobian matrix of $\hat{F}$, which is given by

$$\begin{pmatrix} \partial z^i/\partial u^j & \partial z^i/\partial x^m \\ 0 & \begin{matrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{matrix} \end{pmatrix},$$

has non-vanishing determinant at A, so by Theorems 4.1, 4.5, there is an open set $\hat{D}_o \subset D$ with $A \in \hat{D}_o$ such that $\hat{F}: \hat{D}_o \longrightarrow \hat{F}(\hat{D}_o) = D'$ is a homeomorphism, where $D' \subset W_1 \times V_2$ is open. Moreover, we suppose that $\hat{D}_o$ is chosen sufficiently small that $r(X) = r$ for $X \in \hat{D}_o$. Then $\hat{F}$ has a continuously

differentiable inverse $G: D' \longrightarrow \hat{D}_o$ with $G(Z_o, X_o) = (U_o, X_o) = A$. The relation $\hat{F}G$ = identity map on D' is equivalent to

$$(4) \qquad F_1G(Z, X) \equiv Z, \quad P_2G(Z, X) \equiv X , \qquad (Z, X) \in D' \quad .$$

The second equation of (4) implies that the expression of G in terms of the basis has the form

$$(5) \qquad G(Z, X) = \Sigma_{j=1}^{r} g^j(Z, X)A_j + \Sigma_{m=1}^{k-r} x^m A_{r+m}$$

Next we choose an open set $D'' \subset D'$, with $(Z_o, X_o) \in D''$, which is the intersection of the two open sets of $\hat{V}$ determined by a connected open set $D_2 \subset V_2$ with $X_o \in D_2$ and an open set $D_1 \subset W_1$ with $Z_o \in D_1$ (cf. Definition 1.8). Then we take

$$D_o = G(D'') = G(D_1 \times D_2) \quad .$$

The projection $P_1: D_1 \times D_2 \longrightarrow D_1$, that is, $P_1(Z, X) = Z$, is obviously well-defined and continuously differentiable, by Exercise 5.4, and is surjective, as is the composition $P_1\hat{F}: D_o \longrightarrow D_1$.

To complete the proof, we must show the existence of a map $E: D_1 \longrightarrow W$ such that

$$EP_1\hat{F}(U, X) = F(U, X) = (F_1(U, X), F_2(U, X)) ,$$

for all $(U, X) \in D_o$, that is

$$(6) \qquad E(F_1(U, X)) = (F_1(U, X), F_2(U, X)) \quad .$$

It is sufficient to show that, if $Z = F_1(U, X) = F_1(\tilde{U}, \tilde{X})$, then

$F_2(U, X) = F_2(\tilde{U}, \tilde{X})$. If $r = n$, this is trivial, since $F_2 = \tilde{0}$. If $r = k$, then there is no X, and F_1 coincides with $\hat{F}$, which is bijective, so $U = \tilde{U}$ and therefore $F_2(U) = F_2(\tilde{U})$. For the case $r < k$ and $r < n$, the proof depends on the fact that F is assumed to be regular on D_0.

For fixed $Z \in D_1$, the function $H_1(X) = F_1G(Z, X)$ is constant: $H_1(X) = Z$, by (4), so $H_1' = \tilde{0}$. The equations stating that the entries in the Jacobian matrix of H_1 vanish for $X \in D_2$ are:

$$(7) \quad \Sigma_{j=1}^{r} \frac{\partial f^i}{\partial u^j}(U, X) \frac{\partial g^j}{\partial x^m}(X) + \frac{\partial f^i}{\partial x^m}(U, X) = 0,$$

$$i = 1,\ldots,r;\ m = 1,\ldots,k - r\ ,$$

where $(U, X) = G(Z, X)$. Let $H_2(X) = F_2G(Z, X)$. Then for $Y = \Sigma_{m=1}^{k-r} y^m A_{r+m}$, we have

$$H_2'(X)Y = \Sigma_{\ell=1}^{n-r}[\Sigma_{m=1}^{k-r}(\Sigma_{j=1}^{r} \frac{\partial f^{r+\ell}}{\partial u^j}(U, X) \frac{\partial g^j}{\partial x^m}(X) + \frac{\partial f^{r+\ell}}{\partial x^m}(U, X))y^m]B_{r+\ell} \quad .$$

In order that $H_2'(X) \neq \tilde{0}$ at $X \in D_2$, there must be some Y such that at least one coefficient of the $B_{r+\ell}$'s does not vanish, and for this ℓ, at least one choice of m for which

$$(8) \qquad \Sigma_{j=1}^{r} \frac{\partial f^{r+\ell}}{\partial u^j}(U, X) \frac{\partial g^j}{\partial x^m}(X) + \frac{\partial f^{r+\ell}}{\partial x^m}(U, X) = \alpha_m^\ell \neq 0 \quad .$$

However, since $r(U, X) = r$ there can be no equation of type (8) with $\alpha_m^\ell \neq 0$, and this implies $H_2' = \tilde{0}$, or $H_2 = \text{constant}$ (since D_2 is connected). In fact, the system of equations (7) for the above choice of m, for $i = 1,\ldots,r$, together with (8), would

give a system of $r + 1$ linear equations satisfied by the non-trivial set $(\partial g^1/\partial x^m, \ldots, \partial g^r/\partial x^m, 1)$. For $\alpha_m^\ell \neq 0$ this is not possible unless the determinant of coefficients of this system does not vanish; that is, unless there is a submatrix of order $r + 1$ of the Jacobian matrix of F at (U, X) with non-vanishing determinant, or $r(U, X) > r$.

For each $Z \in D_1$, the set $F_1^{-1}(Z)$ consists of the points $(U, X) = G(Z, X)$, $X \in D_2$, since $H_1(X) = F_1G(Z, X) = Z$. Since $H_2(X) = F_2G(Z, X)$ is constant, it follows that F_2 maps all these points into a common value in W_2. Thus a map E satisfying (6) exists, and can be given explicitly by

$$E(Z) = (Z, F_2G(Z, X))$$

from which it also follows that E is continuously differentiable. It is obvious that $E: D_1 \longrightarrow W_1 \times W_2 = W$ is injective (although the map $D_1 \longrightarrow W_2$ defined by $Z \longrightarrow F_2G(Z, X)$ need not be injective).

4.9. <u>Theorem</u> (Implicit function theorem). Given a set of real-valued functions $f^i(u^1, \ldots, u^r; x^1, \ldots, x^m) = f^i(U, X)$, $i = 1, \ldots, r$, which are assumed to be continuously differentiable for (U, X) in an open set $D \subset R^{r+m}$, such that the determinant of the matrix $(\partial f^i/\partial u^j)$ is different from zero at the point $(U_o, X_o) = (u_o^1, \ldots, u_o^r; x_o^1, \ldots, x_o^m) \in D$, then there is an open set $D_2 \subset R^m$ with $X_o = (x_o^1, \ldots, x_o^m) \in D_2$ and a unique set of continuously differentiable functions $\varphi^j(x^1, \ldots, x^m)$, $i = 1, \ldots, r$, defined on D_2 and satisfying

$$(9) \qquad f^i(\varphi^1(x^1,\ldots,x^m),\ldots,\varphi^r(x^1,\ldots,x^m);\ x^1,\ldots,x^m) \equiv f^i(u_o^1,\ldots,u_o^r;\ x_o^1,\ldots,x_o^m)$$

and

$$(10) \qquad \varphi^i(x_o^1,\ldots,x_o^m) = u_o^i$$

for $i = 1,\ldots,r$.

Proof. The given functions can be considered as the component functions of a continuously differentiable map $F\colon D \longrightarrow R^r$. We then apply Theorem 4.8 for the case $n = r$, $F_2 = \tilde{\tilde{0}}$, $F = F_1$. For $Z_o = F(U_o, X_o)$, we take $\varphi^j(X) = g^j(Z_o, X)$, $j = 1,\ldots,r$, where the $g^j(Z, X)$ are component functions of $G = \hat{F}^{-1}$, as in (5). Then (9) is obtained from the identity (4): $FG(Z_o, X) = Z_o = F(U_o, X_o)$ and (10) is obtained from $G(Z_o, X_o) = (U_o, X_o)$. The uniqueness follows from the fact that $\hat{F}$ is injective.

Note that this method of proof shows also that the solutions $\varphi^j(X)$ of the problem vary continuously and differentiably with respect to the parameters U_o, X_o.

§5. Exercises

1. Show that a subset S of a topological space V is open if and only if for each $X \in S$ there is an open set D in V such that $X \in D$ and $D \subset S$.

2. Let V be a vector space whose topology is defined by means of a scalar product, and let $S \subset V$. Show that a subset

$D \subset S$ is open in S if and only if for each $X \in D$ there exists a positive number δ such that $S \cap B_\delta(X) \subset D$.

3. Let V, W, Z be topological spaces, and let $F: V \longrightarrow W$ and $G: W \longrightarrow Z$ be continuous. Show that $GF: V \longrightarrow Z$ is continuous.

4. Let $F: V \longrightarrow U \times W$, where V, U, W are topological spaces. Show that F is continuous if and only if the functions $P_U F: V \longrightarrow U$ and $P_W F: V \longrightarrow W$ are continuous. If V, U, W are finite dimensional vector spaces, show that F is continuously differentiable if and only if $P_U F$ and $P_W F$ are continuously differentiable.

5. A subset C of a vector space V is called <u>convex</u> if, whenever $A, B \in C$, then $X \in C$ for any X on the straight line segment joining A and B. Show that the intersection of two convex sets is convex. If V is finite dimensional, show that any ball $B_r(X)$, for any choice of r, X and scalar product, is a convex set, as is $\overline{B_r(X)}$.

6. The <u>convex hull</u> of a subset S of a vector space V is defined to be $\cap\, \mathcal{C}$, where $\mathcal{C}$ is the collection of all closed convex sets C satisfying $S \subset C$. Then the convex hull of S is a closed set. Let S consist of k points $A_1, \ldots, A_k \in V$ (not necessarily independent). Show that $L(A_1, \ldots, A_k)$ (Definition I, 7.3) is a closed convex set containing S, and determine the convex hull of S as a particular subset of $L(A_1, \ldots, A_k)$.

7. Let V be a vector space of dimension 2. Choose a basis A_1, A_2 for V and write $X = xA_1 + yA_2$ for $X \in V$.

Let $\hat{V} = V \cap \tilde{0}^{\#}$ and let $F: \hat{V} \longrightarrow R$ be defined by $F(X) = xy/(x^2 + y^2)$. Let m be any real number and let S be the (closed) subset of V defined by the equation $G(X) = 0$ where $G(X) = y - mx$. Show that F is continuous in $S \cap \tilde{0}^{\#}$, that $\lim_{\tilde{0}} F$ exists, say M, as $X \in S \cap \tilde{0}^{\#}$ tends to $\tilde{0}$. Then the function $\bar{F}: V \longrightarrow R$ defined by $\bar{F}(X) = F(X)$ for $X \in V$, $F(\tilde{0}) = M$ is continuous in S (why?). Show that $\bar{F}$ is not continuous in V.

8. If $F: V \longrightarrow R$ is continuous, show that the set $D = \{X \mid X \in V \text{ and } F(X) > 0\}$ is open and that the set $E = \{X \mid X \in V \text{ and } F(X) \geq 0\}$ is closed. Then $\bar{D} \subset E$ (why?). Consider the example that V is a vector space of dimension 2, with notation as in Exercise 7, with $F(X) = x^2(x - 1) - y^2$, and show that $\bar{D} \neq E$. What property of the set $O = \{X \mid X \in V \text{ and } F(X) = 0\}$ makes this possible?

9. For any function $F: S \longrightarrow W$, where S is a subset of a topological space V and W is a finite dimensional vector space, the <u>support</u> of F is defined to be the closure of the set $\{X \mid X \in S \text{ and } F(X) \neq \tilde{0}\}$. Give an alternative definition of the support of F as the intersection of a certain collection of closed sets, and prove that the two definitions are equivalent.

10. Let V be a topological space and let $F: D \longrightarrow R$ be continuous where D is an open set in V. Suppose that the support of F is compact and contained in D, and define $\hat{F}: V \longrightarrow R$ by $\hat{F}(X) = F(X)$ for $X \in D$ and $\hat{F}(X) = 0$ for $X \in D^{\#}$. Show that $\hat{F}$ is continuous and bounded on V.

11. Let a topology $\mathfrak{J}$ be defined on the set R of real numbers by taking $\mathfrak{J}_0$ to include $\emptyset$, R, and the sets $D_r = \{\lambda \mid \lambda \in R \text{ and } \lambda < r\}$, for all $r \in R$. Show (i) that $\bar{\mathfrak{J}}_0 = \mathfrak{J}$ (what properties of the real number system have you used to obtain this result?), (ii) that the set R with the topology $\mathfrak{J}$ is not a Hausdorff space, and (iii) that $\mathfrak{J} \subset \mathfrak{J}^*$, where $\mathfrak{J}^*$ denotes the usual topology for R.

12. A real-valued function which is continuous relative to the topology $\mathfrak{J}$ of Exercise 11 is called <u>upper semi-continuous</u>. Show that a real-valued continuous function is upper semi-continuous. Take V to be a vector space of dimension 1, and construct some examples of real-valued functions which are continuous on $V \cap \mathfrak{O}^{\#}$ and upper semi-continuous, but not continuous, on V.

13. Let N be the set consisting of the non-negative integers and let $\bar{N} = N \cup \infty$, where ∞ denotes a set consisting of a single element (denoted by the same symbol). (i) Show that $\bar{N}$ can be assigned a topology by the definition: a subset of $\bar{N}$ is closed if it is a finite subset of N or if it is all of $\bar{N}$. (ii) Show that $\bar{N}$, with this topology, is not a Hausdorff space, but that the properties (a) and (b) of Proposition 2.4 hold for any $A \in \bar{N}$, and that property (c) holds only for $A = \infty$. (Consequently, ∞ can be considered as a limit point of N).

14. A function $F: N \longrightarrow W$, where W is a topological space, is called a <u>sequence</u>. Suppose that W is a Hausdorff space.

In analogy with Definition 2.7, state the condition that $B \in W$ be the limit of F as $n \in N$ tends to ∞, and show that this definition agrees with the usual one for the case $W = R$.

Remark. The functions $x^j(X)$ may be considered as the component functions of the linear transformation T (Corollary II, 7.6) identifying V with euclidean space R^n where $T(A_1) = (1, 0,\ldots,0)$, $T(A_2) = (0, 1,\ldots,0),\ldots,T(A_n) = (0, 0,\ldots,1)$, and $T(X) = (x^1(X), x^2(X),\ldots,x^n(X))$. By Proposition 1.3, T and $S = T^{-1}$ are homeomorphisms of class C^∞. In particular, the real-valued functions $X \longrightarrow x^j(X)$ are of class C^∞, by Proposition 1.5. The values of the functions $x^j(X)$, for any $X \in D$, are the euclidean coordinates of $T(X)$ and are then called the coordinates of X. Further, if D is an open set in V, then $\hat{D} = T(D)$ is an open set in R^n. Any function $f: D \longrightarrow R$ determines a function $fS: \hat{D} \longrightarrow R$. It is usual to omit explicit mention of the identification S and to write

$$f(X) = f(x^1,\ldots,x^n)$$

where the x^j are considered as coordinates on D rather than as coordinates on $\hat{D}$. [For a further discussion of coordinate systems, see §3.]

1.7. Proposition. A function $f: D \longrightarrow R$, $D \subset V$, is of class C^k, $k \geq 1$, if and only if it has continuous partial derivatives of all orders $\leq k$ when expressed in terms of linear coordinates on V.

Proof. For $k = 1$, this is essentially Proposition VII, 1.8. The proof of the "only if" statement can be modified as follows. For $Y = \Sigma_{j=1}^n y^j A_j$, where the y^j are constants, and a function $f: D \longrightarrow R$ of class C^1, the formula (8) of Chapter VII,

§1 becomes

$$(2) \qquad \mathfrak{D}_Y f = \Sigma_{j=1}^n y^j \mathfrak{D}_{A_j} f = \Sigma_{j=1}^n y^j \frac{\partial f}{\partial x^j} \quad .$$

Since the y^j can be chosen arbitrarily, the continuity of $\mathfrak{D}_Y f$ for all $Y \in V$ implies the continuity of each partial derivative $\partial f/\partial x^j$, $j = 1,\ldots,n$. For $k > 1$, the same argument may be used inductively. For example, if f is of class C^2, then the above formula for the function $\mathfrak{D}_{Y_2} f$ of class C^1, gives

$$(3) \qquad \mathfrak{D}_{Y_1}(\mathfrak{D}_{Y_2} f) = \Sigma_{\ell,j=1}^n y_1^\ell y_2^j \frac{\partial^2 f}{\partial x^\ell \partial x^j} \; ;$$

again the y_1^ℓ, y_2^j can be chosen arbitrarily, etc.

Combining Propositions 1.5 and 1.7 we obtain

1.8. <u>Proposition</u>. A function $F: D \longrightarrow W$, $D \subset W$, is of class C^k if and only if the component functions, relative to any choice of base for W, have continuous partial derivatives of all orders $\leq k$ when expressed in terms of any (linear) coordinate system on V.

In computing $F'(X, Y)$, the values of X and of Y are kept fixed; consequently, we may equally well consider the case that Y depends on X; that is, Y is a vector field $Y: D \longrightarrow V$ (Definition VII, 6.1). Then (2) becomes

$$(4) \qquad \begin{aligned} (\mathfrak{D}_Y f)(X) &= \Sigma_{j=1}^n y^j(X)(\mathfrak{D}_{A_j} f)(X) \\ &= \Sigma_{j=1}^n y^j(x^1,\ldots,x^n) \frac{\partial f}{\partial x^j}(x^1,\ldots,x^n) \quad . \end{aligned}$$

Using (4), (1) and Propositions 1.2 and 1.4, we obtain

1.9. Proposition. If $F: D \longrightarrow W$ is a function of class C^k, $k \geq 1$, and if Y is a vector field of class C^{k-1}, then $\mathfrak{D}_Y F$ is a function of class C^{k-1}.

1.10. Remarks. A constant vector $Y \in V$ as in Definition 1.1, etc., may also be considered as the vector field defined by the constant map $D \longrightarrow Y$ and, by Proposition 1.3, is a vector field of class C^∞. All the earlier propositions remain valid if Y denotes a vector field of class C^∞, not necessarily constant. However, if Y_2 is a vector field, (3) is to be replaced by

$$\mathfrak{D}_{Y_1}(\mathfrak{D}_{Y_2} f) = \Sigma_{\ell, j=1}^{n} y_1^\ell y_2^j \frac{\partial^2 f}{\partial x^\ell \partial x^j} + \mathfrak{D}_Z f$$

where $Z = \mathfrak{D}_{Y_1} Y_2 = \Sigma_{\ell, j=1}^{n} y_1^\ell \frac{\partial y_2^j}{\partial x^\ell} A_j$. In particular, $\mathfrak{D}_{Y_1} \mathfrak{D}_{Y_2} \neq \mathfrak{D}_{Y_2} \mathfrak{D}_{Y_1}$ in general. This non-commutativity may arise in dealing with systems of linear differential equations with non-constant coefficients, and is "measured" by the operator $\mathfrak{D}_Z = \mathfrak{D}_{Y_1} \mathfrak{D}_{Y_2} - \mathfrak{D}_{Y_2} \mathfrak{D}_{Y_1}$ where $Z = \mathfrak{D}_{Y_1} Y_2 - \mathfrak{D}_{Y_2} Y_1 = [Y_1, Y_2]$ is a vector field called the "bracket product" of the vector fields Y_1 and Y_2. It is left as an exercise to verify that in addition to the obvious identities

$$[Y, Y] = \tilde{0}, \quad [Y_1, Y_2] = -[Y_2, Y_1],$$

the bracket product satisfies the "Jacobi identity"

$$[[Y_1, Y_2], Y_3] + [[Y_2, Y_3], Y_1] + [[Y_3, Y_1], Y_2] = \tilde{0} .$$

(Note that $[Y_1, Y_2]$ is of class C^{k-1} if Y_1, Y_2 are of class C^k, and cannot be defined if the vector fields are merely continuous.)

1.11. Proposition. Let $F: D \longrightarrow W$, $D \subset V$, and $G: \tilde{D} \longrightarrow \tilde{W}$, $F(D) \subset \tilde{D} \subset \tilde{W}$, be of class C^k. Then $GF: D \longrightarrow \tilde{W}$ is of class C^k.

Proof. For $k = 0$, this is Exercise X, 5.3. For $k \geq 1$, this result follows from Proposition 1.8 and the usual rule for differentiation of composite functions in euclidean space. For example, let $(x^1,\ldots,x^n)$ be linear coordinates corresponding to a choice of base for V; let $(\tilde{x}^1,\ldots,\tilde{x}^m)$ be linear coordinates in W corresponding to a choice of base in W, and let f^i be the component functions of F corresponding to the same base for W; finally, let g^ℓ be the component functions of G relative to a choice of base $\tilde{B}_1,\ldots,\tilde{B}_p$ for $\tilde{W}$. Then

$$GF(X) = \Sigma_{\ell=1}^{p} g^\ell(f^1(x^1,\ldots,x^n),\ldots,f^m(x^1,\ldots,x^n))\tilde{B}_\ell$$

and (2) gives

$$(5) \qquad (\mathcal{D}_Y(GF))(X) = \Sigma_{\ell=1}^{p} \Sigma_{j=1}^{n} \Sigma_{i=1}^{m} y^j \frac{\partial g^\ell}{\partial \tilde{x}^i}(F(X)) \frac{\partial f^i}{\partial x^j}(X)\tilde{B}_\ell$$

with continuous coefficients; for example, $\partial g^\ell/\partial \tilde{x}^j$ is a continuous function in the coordinates $(\tilde{x}^1,\ldots,\tilde{x}^m)$, since G is of class C^1, and the $\tilde{x}^i$ are continuous functions in the $(x^1,\ldots,x^n)$ since F is of class C^1 and therefore also of class C^0.

For later use we note that (5) is the same as

(6) $$(\mathcal{D}_Y(GF))(X) = (\mathcal{D}_{F'(X,\ Y)}G)(F(X)) \ , \quad X \in D \ , \ Y \in V \ .$$

1.12. Corollary to Theorems X, 4.1, 4.5. If the function $F: D \longrightarrow W$ in Theorem X, 4.1 is of class C^k, $k > 1$, rather than of class C^1, then the inverse function is also of class C^k.

Proof. The proof of Theorem X, 4.5 consisted in showing that the derivative G' of the inverse function G coincided with a certain continuous function, there denoted by H. Thus it is sufficient to show that H is of class C^{k-1} under the assumption that F is of class C^k and the inductive assumption that G is of class C^{k-1}; this follows by the same argument, replacing "continuous function" by "function of class C^{k-1}", using Proposition 1.4, and using Proposition 1.11 at the last step.

§2. Associated structures

Any function $F: D \longrightarrow W$ can be considered as a section of the product space $D \times W$ over D, that is, as a function $\hat{F}: D \longrightarrow D \times W$ satisfying $\pi\hat{F}$ = identity map on D, where $\pi: D \times W \longrightarrow D$ is the projection as in Definition X, 1.8. In fact, the points of $D \times W$ are pairs (X, Z), $X \in D$ and $Z \in W$, and $\hat{F}: D \longrightarrow D \times W$ is given by $X \longrightarrow (X, F(X))$.

The notion of considering a function $F: D \longrightarrow W$ as a section of $D \times W$ over D is the exact analogue of representing a function $y = f(x)$ by its graph in the (x, y)-plane. As in the elementary case, both F and $\hat{F}$ will usually be denoted by the

same symbol F. The difference between the two approaches may readily be seen by considering the two interpretations of $F: D \longrightarrow W$ if F is a constant map.

The term "section" is borrowed from the theory of "fibre bundles" (we have here a very special case: a product bundle). Borrowing from the same source, we shall call the set of points $\pi^{-1}(X)$, for any $X \in D$, the <u>fibre</u> over X. Then a section is a function which assigns to each $X \in D$ a point of the fibre over X. As we consider different choices of W we shall usually use the same symbol π to denote the projection onto D.

Let V, W be finite dimensional vector spaces, and D an open subset of V. By generalizing Exercise X, 5.4 ($k = 0, 1$) to arbitrary integers k, we have that $\hat{F}$ is of class C^k if and only if $\sigma\hat{F} = F: D \longrightarrow W$ is of class C^k, where σ denotes the projection $D \times W \longrightarrow W$. (The other condition, that $\pi\hat{F}$ be of class C^k, is trivial since the identity map is of class C^∞ by Proposition 1.3.)

2.1. <u>Functions</u>. For $W = R$, the sections are the functions on D. [The word "function" will henceforth denote a real-valued function; for arbitrary choices of W, we shall use the term "W-valued function" when the interpretation as section of $D \times W$ is intended, and "map" otherwise. Particular names will be introduced below in the case of certain particular choices of W.] If D is an open subset of a finite dimensional vector space, then the notion of functions of class C^k, $k = 0, 1, 2, \ldots,$ on D has

been defined (Definition 1.1, etc.), and we denote by $\mathfrak{F}^k = \mathfrak{F}^k(D)$ the set of functions of class C^k on D. (Wherever possible, the superscript k and the reference to D will be omitted.) By Proposition 1.2, $\mathfrak{F}^k \subset \mathfrak{F}^\ell$ if $k > \ell$. Combining Propositions 1.4 and 1.3, we see that each set $\mathfrak{F}$ may be described as a commutative ring with unit element (Chapter IX, 2.14) which contains a subring (the set of constant functions) isomorphic to R. With the interpretation of functions as sections of $D \times R$, one now visualizes the pointwise addition or multiplication of two functions $f, g \in \mathfrak{F}$ as taking place in the fibre over X, for each $X \in D$.

2.2. <u>Vector-valued functions</u>. Let W be an arbitrary vector space of dimension m, and let $\mathfrak{W}^k = \mathfrak{W}^k(D; W)$ denote the set of W-valued functions on D of class C^k. Then $\mathfrak{W}^k$ may be described as a free $\mathfrak{F}^k$-module of dimension m (see Chapter IX, 2.14), that is, a set which satisfies the axioms of Definition I, 1.1, with V replaced by $\mathfrak{W}^k$ and the field of real numbers by the ring $\mathfrak{F}^k$ and, further, there is a set of m elements of $\mathfrak{W}^k$ which forms a base for $\mathfrak{W}^k$ using coefficients from $\mathfrak{F}^k$. In fact, if $B_1, \ldots, B_m$ is a base for W then, for any $F \in \mathfrak{W}^k$, we have the representation

$$F(X) = \Sigma_{i=1}^m f^i(X)B_i$$

in which the functions f^i are in $\mathfrak{F}^k$, by Proposition 1.5. All that is necessary is to define $B_i \in \mathfrak{W}^k$ to be the (constant) section $D \longrightarrow B_i$; then we have

$$F = \Sigma_{i=1}^{m} f^i B_i ,$$

for some choice of the $f^i \in \mathfrak{F}^k$, for each $F \in \mathfrak{W}^k$.

Note that addition and scalar multiplication in $\mathfrak{W}^k$ are defined by carrying out these operations fibrewise, that is, in the fibre over X, for each $X \in D$. That the result of the computation gives an element of $\mathfrak{W}^k$ follows from Propositions 1.4 and 1.5.

Since scalar multiplication is defined for elements of the subring $R \subset \mathfrak{F}^k$, the set $\mathfrak{W}^k$ may also be considered as a vector space (of infinite dimension). In the case of a "linear transformation" on $\mathfrak{W}^k$, we shall include the phrase "over R" to indicate transformations which are not linear with respect to the scalars $\mathfrak{F}^k$. Any transformation which is linear over $\mathfrak{F}^k$ is linear over R as well.

2.3. <u>Tangent vectors and vector fields</u>. Here $W = V$ if $D \subset V$. We write $T = T(D) = D \times V$; the space T is called the <u>tangent space of D</u>, and $T_X = \pi^{-1}(X)$, $\pi: T \longrightarrow D$, is called the space of <u>tangent vectors to D at X</u> (or the tangent space to D at X) and is a copy of V. The sections of T of class C^0 have been called <u>vector fields</u> and we denote by $\tau^k = \tau^k(D)$ the set of vector fields of class C^k on D, and by v an element of τ^k. Each element v of τ^k assigns to each $X \in D$ a tangent vector $u = v(X) \in T_X$.

By 2.2, the set τ^k is a free $\mathfrak{F}^k$-module of dimension n; that is, the functions act on the vector fields by scalar

multiplication, etc. On the other hand, the vector fields and tangent vectors act on functions by differentiation. Each $v \in \tau^k$ determines a correspondence $\mathfrak{D}_v\colon \mathfrak{F}^{k+1} \longrightarrow \mathfrak{F}^k$, $k > 0$, by $f \longrightarrow \mathfrak{D}_v f$ (Proposition 1.9); each $u \in T_X$ determines a correspondence $\mathfrak{D}_u\colon \mathfrak{F}^k \longrightarrow R$, $k \geqq 1$, by $f \longrightarrow \mathfrak{D}_u f = \mathfrak{D}_u f(X)$, and

(1) $\mathfrak{D}_u(rf + sg) = r\mathfrak{D}_u f + s\mathfrak{D}_u g$, $f, g \in \mathfrak{F}^k$, $k \geqq 1$, $r, s \in R$,

(2) $\mathfrak{D}_u(fg) = (\mathfrak{D}_u f)g(X) + f(X)(\mathfrak{D}_u g)$, $f, g \in \mathfrak{F}^k$, $k \geqq 1$.

<u>Remarks</u>. It should be noted that, although each $u \in T_X$ is considered as a vector at X, a function f must be given for all points near X as well, in order to compute all $\mathfrak{D}_u f$. We have assumed that f is defined on all of D; actually to compute $\mathfrak{D}_u f$ for all $u \in T_X$ it is sufficient if f is given on some open set containing X. Also, it should be noted in what sense u is a tangent vector <u>at X</u>. In computing $\mathfrak{D}_u f$ we have used the value $Y \in V$ obtained from u by the projection $\sigma\colon T \longrightarrow V$, and the values of f at the points $X + tY$ for small values of t. Now Theorem VII, 2.1 shows that the derivative of a function of class C^1 at X along any smooth curve C through X will have the same value provided the curve C is tangent to the curve $X + tY$ at X, in the sense that both curves have the same instantaneous velocity vector at X; that is, each $u \in T_X$ represents an equivalence class in the set of all curves through X, two curves through X lying in the same

equivalence class if they are "tangent" to each other at X. Thus, for functions f of class C^k, $k \geq 1$, a differential operator $\mathfrak{D}_u$ corresponds to an equivalence class; in computing $\mathfrak{D}_u f$ we have merely used the simplest curve of the equivalence class, viz. a straight line. [But "straight" is a very special property of a curve; cf. §3.]

2.4. Associated base. Any choice of base $A_1, \ldots, A_n$ for V serves to introduce (linear) coordinates $(x^1, \ldots, x^n)$ on V, as described in Definition 1.6. The same choice also leads to a base for τ^k (scalars $\mathfrak{F}^k$) as described in 2.2 and for T_X for each $X \in D$ (scalars R). By Proposition VII, 1.8 we have

$$(\mathfrak{D}_{A_j} f)(X) = \frac{\partial f}{\partial x^j}(x^1, \ldots, x^n)$$

when f is expressed in terms of the corresponding coordinate system. Consequently, instead of using the notation A_j, the basis elements are usually denoted symbolically by $\frac{\partial}{\partial x^j}$, $j = 1, \ldots, n$, (and the same symbol is used for the corresponding differential operator); that is, $\frac{\partial}{\partial x^j}$ denotes the vector field $D \longrightarrow A_j$ and also the tangent vector at X which corresponds to A_j. For any $v \in \tau^k$, we have

$$v = \Sigma_{j=1}^n v^j \frac{\partial}{\partial x^j}$$

where $v^j \in \mathfrak{F}^k$. It is easy to evaluate the functions v^j corresponding to a given v. For any $f \in \mathfrak{F}^1$ we have

$$\mathfrak{D}_v f = \Sigma_{j=1}^n v^j \frac{\partial f}{\partial x^j} ;$$

in particular, for the function $f^\ell : X \longrightarrow x^\ell(X)$, we have $\frac{\partial f^\ell}{\partial x^j} = \delta^\ell_j$ for all $X \in D$. Then

$$\mathfrak{D}_v f^\ell = \Sigma_{j=1}^n v^j \frac{\partial f^\ell}{\partial x^j} = v^\ell \ ;$$

that is, $v^j = \mathfrak{D}_v f^j = \mathfrak{D}_v x^j$. Similarly, if $u \in T_X$, we have $u = \Sigma_{j=1}^n u^j \frac{\partial}{\partial x^j}$, where $u^j \in R$ is given by $(\mathfrak{D}_u x^j)(X)$. The u^j, together with x^j for $X = \pi(u)$, give a system of (linear) coordinates on T.

<u>Remarks</u>. (i) An alternative notation for $\mathfrak{D}_u f$, in the case of <u>functions</u> only, is $u \cdot f$. [There should be no confusion with the scalar product (defined only for pairs of the same kind) whenever it is clear that the first symbol denotes a vector and the second a function; $v \cdot f$ must also be distinguished from fv if v denotes a vector field.] Then the above formulas are written as

$$v = \Sigma_{j=1}^n v^j \frac{\partial}{\partial x^j} \ , \qquad\qquad v^j = v \cdot x^j \ ,$$

and

$$v \cdot f = \Sigma_{j=1}^n v^j \frac{\partial f}{\partial x^j} \ ,$$

with similar formulas for the case of a tangent vector u ~~at X~~.

(ii) If $X \in D$, we can construct the set T_X axiomatically, starting from the set $\mathfrak{F} = \mathfrak{F}^\infty(D)$, as follows: T_X is the set of all maps L of $\mathfrak{F}$ into R having the properties

(3) $\qquad L(rf + sg) = rL(f) + sL(g), \qquad f, g \in \mathfrak{F}, \ r, s \in R \ ,$

(4) $\qquad L(fg) = L(f)g(X) + f(X)L(g), \qquad f, g \in \mathfrak{F} \ .$

Note first that these axioms imply

(5) $$L(f) = 0 \quad \text{if} \quad f \quad \text{is a constant function.}$$

In fact, if $f(X) \equiv 1$, and g is arbitrary, then (4) gives

$$L(g) = L(1)g(X) + 1(X)L(g) = L(1)g(X) + L(g) \quad .$$

If we take g such that $g(X) \neq 0$, this implies $L(1) = 0$. An arbitrary constant function may be written in the form rf, $r \in R$, $f(X) \equiv 1$, and (3) then gives $L(rf) = 0$.

Let $(x^1,\ldots,x^n)$ be a linear coordinate system on D. Then $f = f(x^1,\ldots,x^n)$ for any $f \in \mathfrak{F}$. For each set $(u^1,\ldots,u^n)$ of real numbers, we have a corresponding L by taking $L(f) = \Sigma_{j=1}^n u^j \frac{\partial f}{\partial x^j}(x^1(X),\ldots,x^n(X))$. It remains only to prove that all maps $L: \mathfrak{F} \longrightarrow R$ satisfying (3) and (4) are of this type.

Let $D_o \subset D$ be a <u>star-shaped</u> open set about X, that is, an open set such that, if $Z \in D_o$, then $X + t(Z - X) \in D_o$ for $0 \leq t \leq 1$. For $Z \in D_o$ and any $f \in \mathfrak{F}$ we have

$$\begin{aligned} f(Z) &= f(X) + \int_0^1 \frac{d}{dt} f(X + t(Z - X))dt \\ &= f(X) + \Sigma_{j=1}^n (x^j(Z) - x^j(X)) \int_0^1 \frac{\partial f}{\partial x^j}(X + t(Z - X))dt \\ &= f(X) + \Sigma_{j=1}^n (x^j(Z) - x^j(X))g_j(Z) \end{aligned}$$

where $g_j \in \mathfrak{F}(D_o)$ and $g_j(X) = (\frac{\partial f}{\partial x^j})(X)$. For any L satisfying (3), (4), and therefore (5), we must have (X fixed, Z variable)

$$\begin{aligned} L(f) &= 0 + \Sigma_{j=1}^n \{L(x^j(Z) - x^j(X))\cdot g_j(X) + (x^j(X) - x^j(X))\cdot L(g_j)\} \\ &= \Sigma_{j=1}^n u^j \frac{\partial f}{\partial x^j}(X), \qquad\qquad u^j = L(x^j) \ ; \end{aligned}$$

that is, L is of the type already found.

Since this construction also assigns coordinates $(u^1,\ldots,u^n)$ to each $L \in T_X$, the constructed "tangent vectors" can be put together to form a space T which is equivalent to $D \times R^n$.

The above proof fails if we start from $\mathfrak{F} = \mathfrak{F}^k(D)$ for $k < \infty$, since the functions g_j are not in the set $\mathfrak{F}$ for which the axioms (3) and (4) hold. However, the same conclusion is obtained if we start from a single stronger axiom to the effect that each L satisfies the identities corresponding to the usual "function of a function rule" for arbitrary composite functions.

2.5. <u>Tensor fields, q-vectors, and differential forms</u>. Given an open $D \subset V$, we now have: (1) for each non-negative integer k the set $\mathfrak{F}^k$ of functions on D of class C^k; (2) the tangent space T of D whose fibres T_X are isomorphic copies of the fixed vector space V; and (3) for each non-negative integer k the set τ^k of vector fields on D of class C^k, all regarded as part of the "built-in" structure associated with D. We next enlarge this structure by means of the constructions studied in Chapter IX.

Starting from V, we can construct the vector spaces $\overset{s}{\otimes} V$, $s = 0, 1, \ldots,$ and $\Lambda^q V$, $q = 0, 1, \ldots$. We have $\overset{0}{\otimes} V = \Lambda^0 V = R$, $\overset{1}{\otimes} V = \Lambda^1 V = V$ and $\Lambda^q V = \tilde{0}$ for $q > n$; also, $\overset{s}{\otimes} V$ has dimension n^s (Proposition IX, 2.18) and $\Lambda^q V$ has dimension $\binom{n}{q}$ (Theorem IX, 7.11). Then we consider the product spaces $\overset{s}{\otimes} T = D \times \overset{s}{\otimes} V$ and $\Lambda^q T = D \times \Lambda^q V$ over D. [The

notation is symbolic: T is not a vector space, so we cannot form $T \otimes T$, etc.] Note that $(\overset{s}{\otimes} T)_X$ is isomorphic to $\overset{s}{\otimes}(T_X)$, and $(\Lambda^q T)_X$ to $\Lambda^q(T_X)$ (why?). An element of $(\overset{s}{\otimes} T)_X$ is called a <u>contravariant</u> (see §3) <u>tensor at X</u> of order s, and an element of $(\Lambda^q T)_X$ is called a <u>q-vector at X</u>. The elements of the set $\overset{s}{\otimes}\tau^k$ of sections (of class C^k) of $\overset{s}{\otimes} T$ are the <u>contravariant tensor fields</u> of order s on D and the elements of the set $\Lambda^q\tau^k$ of sections (of class C^k) of $\Lambda^q T$ are the <u>q-vector fields</u> on D. [Here the notation is not necessarily symbolic. Since $\tau = \tau^k$ is a free $\mathfrak{F}^k$-module of dimension n, it is quite possible to carry out the construction of $\tau \otimes \tau$, etc., as in Chapter IX, with V replaced by τ and the set of real numbers by $\mathfrak{F} = \mathfrak{F}^k$ (so the tensor product is taken with respect to $\mathfrak{F}$ rather than R); but it requires a proof to show that the result is equivalent to that obtained by constructing $\overset{s}{\otimes} T$ first, and then taking sections.]

Let $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ denote the associated base for τ, as in 2.4, corresponding to a particular choice of (linear) coordinates on D; this gives a base, denoted by the same symbols, for T_X for each $X \in D$. By Proposition IX, 2.18, a basis for $(\overset{s}{\otimes} T)_X$ is given by the collection of elements of the form

$$\frac{\partial}{\partial x^{j_1}} \otimes \frac{\partial}{\partial x^{j_2}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_s}},$$ for all choices of $j_1, j_2, \ldots, j_s$ between 1 and n; any tensor $t \in \overset{s}{\otimes}(T_X)$ can be expressed in the form

$$t = \Sigma^n_{j_1, j_2, \ldots, j_s = 1} \; t^{j_1 j_2 \cdots j_s} \frac{\partial}{\partial x^{j_1}} \otimes \frac{\partial}{\partial x^{j_2}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_s}} \;, \qquad t^{j_1 j_2 \cdots j_s} \in R \;. \tag{6}$$

A section t of $\overset{s}{\otimes} T$ over D, that is, a tensor field, can be expressed in the same way except that the coefficients $t^{j_1 j_2 \cdots j_s}$ vary with X and thus determine functions, which are of class C^k if and only if $t \in \overset{s}{\otimes} \tau^k$. Each symbol $\frac{\partial}{\partial x^{j_1}} \otimes \frac{\partial}{\partial x^{j_2}} \otimes \cdots \otimes \frac{\partial}{\partial x^{j_s}}$ represents a section of $\overset{s}{\otimes} T$ of class C^∞ (why?).

Similarly, by Theorem IX, 7.11, the collection of all q-vectors of the form $\frac{\partial}{\partial x^{j_1}} \wedge \frac{\partial}{\partial x^{j_2}} \wedge \cdots \wedge \frac{\partial}{\partial x^{j_q}}$, with $j_1 < j_2 < \cdots < j_q$ between 1 and n, forms a basis for the q-vectors $(\wedge^q T)_X$ at X, and any q-vector $u \in (\wedge^q T)_X$ can be expressed as

$$u = \Sigma_{j_1 < j_2 < \cdots < j_q \leq n} \; u^{j_1 j_2 \cdots j_q} \frac{\partial}{\partial x^{j_1}} \wedge \frac{\partial}{\partial x^{j_2}} \wedge \cdots \wedge \frac{\partial}{\partial x^{j_q}} \;, \qquad u^{j_1 j_2 \cdots j_q} \in R \;. \tag{7}$$

Any q-vector field $v \in \wedge^q \tau^k$ can be expressed in the same way with coefficients varying with X and determining functions of class C^k. The basis elements $\frac{\partial}{\partial x^{j_1}} \wedge \frac{\partial}{\partial x^{j_2}} \wedge \cdots \wedge \frac{\partial}{\partial x^{j_q}}$ are sections of $\wedge^q T$ of class C^∞.

Next we form the dual space V^* of V, and the product space $T^* = D \times V^*$ over D [notation T^* symbolic]. The elements

of $(T^*)_X = (T_X)^*$ are called <u>1-forms at X</u>. The elements of the set $(\tau^k)^*$ of sections (of class C^k) of T^* are the 1-forms on D, or <u>differential forms</u> of degree 1 (of class C^k). Each $\omega \in (T^*)_X$ defines a linear transformation of T_X into the real numbers R; each $\omega \in (\tau^k)^*$ determines a linear (with respect to $\mathfrak{F} = \mathfrak{F}^k$ as scalars) transformation of $\tau = \tau^k$ into $\mathfrak{F}$, which is defined pointwise, that is, in the fibres over X for $X \in D$.

If $\frac{\partial}{\partial x^1},\ldots,\frac{\partial}{\partial x^n}$ is the base for τ or for T_X associated to a particular choice of (linear) coordinates on D, we denote the dual basis (Definition IX, 1.3) for $(T^*)_X$ by $dx^1,\ldots,dx^n$ (the reason for the choice of notation will appear in §4). That is, dx^i is the 1-form at X for which the corresponding linear transformation of T_X into R has the values

$$(8) \qquad < \frac{\partial}{\partial x^j}, dx^i > = \delta^i_j , \qquad j = 1,\ldots,n ,$$

for the base $\frac{\partial}{\partial x^1},\ldots,\frac{\partial}{\partial x^n}$ of T_X. If we allow X to vary, we also denote by dx^i, $i = 1,\ldots,n$, the differential form of degree 1 given by $X \longrightarrow dx^i$. Then (8) remains true when the symbols denote the vector field and the differential form respectively. Any 1-form ω can be expressed as

$$\omega = \Sigma_{i=1}^n \omega_i dx^i$$

where $\omega_i = < \frac{\partial}{\partial x^i}, \omega >$. Here $\omega_i \in R$ if $\omega \in (T^*)_X$ and $\omega_i \in \mathfrak{F}^k$ if $\omega \in (\tau^k)^*$. More generally, if $u = \Sigma_{j=1}^n u^j \frac{\partial}{\partial x^j}$, then

$$(9)\qquad \begin{aligned} < u, \omega > &= \Sigma_{i,j=1}^{n} < u^j \frac{\partial}{\partial x^j}, \omega_i dx^i > \\ &= \Sigma_{i,j=1}^{n} u^j \omega_i \delta_j^i = \Sigma_{i=1}^{n} u^i \omega_i \ . \end{aligned}$$

The symbol $< u, \omega >$ denotes a real number if ω and u are given at a point $X \in D$ (the <u>same</u> point for both), or a function in $\mathfrak{F}^k$ if $\omega \in (\tau^k)^*$ and $u \in \tau^k$.

Finally, we may form the vector spaces $\overset{r}{\otimes} V^*$, $\overset{r,s}{\otimes} V = V^* \otimes \ldots \otimes V^* \otimes V \otimes \ldots \otimes V$ (with V^* appearing r times and V appearing s times) and $\Lambda^p V^*$, $p = 0, 1, \ldots, n$, and then the corresponding product spaces $\overset{r}{\otimes} T^* = D \times \overset{r}{\otimes} V^*$, $\overset{r,s}{\otimes} T = D \times \overset{r,s}{\otimes} V$, and $\Lambda^p T^* = D \times \Lambda^p V^*$. An element in a fibre over X is called, respectively, <u>a covariant tensor at X</u> of order r, a <u>mixed tensor at X, covariant of order r and contravariant of order s</u>, and a <u>p-form at X</u>. The corresponding sets of sections of class C^k, denoted by $\overset{r}{\otimes} (\tau^k)^*$, $\overset{r,s}{\otimes} \tau^k$, and $\Lambda^p (\tau^k)^*$, are <u>covariant tensor fields</u> on D of order r, <u>mixed tensor fields</u> on D, and <u>differential forms of degree p</u>.

Using the results of Chapter IX we have the following expressions in terms of associated basis elements, for $t \in (\overset{r}{\otimes} T^*)_X = (\overset{r}{\otimes} (T_X))^*$, $t \in (\overset{r,s}{\otimes} T)_X$, and $\varphi \in (\Lambda^p T^*)_X = (\Lambda^p (T_X))^*$, respectively,

$$(10)\qquad t = \Sigma_{i_1, i_2, \ldots, i_r = 1}^{n} t_{i_1 i_2 \ldots i_r} \, dx^{i_1} \otimes dx^{i_2} \otimes \ldots \otimes dx^{i_r} \ ,$$

where

$$t_{i_1 i_2 \ldots i_r} = < \frac{\partial}{\partial x^{i_1}} \otimes \frac{\partial}{\partial x^{i_2}} \otimes \ldots \otimes \frac{\partial}{\partial x^{i_r}}, t > \in R$$

when t is considered as an element of $(\overset{r}{\otimes}(T_X))^*$;

$$t = \Sigma^n_{i_1,i_2,\ldots,i_r=1} \; \Sigma^n_{j_1,j_2,\ldots,j_s=1} \; t_{i_1 i_2 \ldots i_r}{}^{j_1 j_2 \ldots j_s} \cdot$$

(11)

$$dx^{i_1} \otimes dx^{i_2} \otimes \ldots \otimes dx^{i_r} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \frac{\partial}{\partial x^{j_2}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_s}} ,$$

where

$$t_{i_1 i_2 \ldots i_r}{}^{j_1 j_2 \ldots j_s}$$

$$= < \frac{\partial}{\partial x^{i_1}} \otimes \frac{\partial}{\partial x^{i_2}} \otimes \ldots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes dx^{j_2} \otimes \ldots \otimes dx^{j_s}, \; t > \in R \; ;$$

and

(12) $$\varphi = \Sigma_{i_1<i_2<\ldots<i_p\leq n} \; \varphi_{i_1 i_2 \ldots i_p} \; dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_p}$$

where

(13) $$\varphi_{i_1 i_2 \ldots i_p} = < \frac{\partial}{\partial x^{i_1}} \wedge \frac{\partial}{\partial x^{i_2}} \wedge \ldots \wedge \frac{\partial}{\partial x^{i_p}}, \; \varphi > \in R \quad .$$

Note that (13) defines a number $\varphi_{i_1 i_2 \ldots i_p}$ even if we do not have $i_1 < i_2 < \ldots < i_p$ and that these numbers are skew symmetric in their indices. The ordering of the indices is merely an arbitrary device to select a basis for $(\wedge^p T^*)_X$, in particular a linearly independent set of elements $(p \geqq 2)$, so as to have uniquely determined coefficients. For example, if φ is a 2-form at X, we may write

$$\varphi = \varphi_{12}\, dx^1 \wedge dx^2$$

$$= \tfrac{1}{2}\varphi_{12}\, dx^1 \wedge dx^2 + \tfrac{1}{2}\varphi_{21}\, dx^2 \wedge dx^1$$

$$= \tfrac{1}{3}\varphi_{12}\, dx^1 \wedge dx^2 + \tfrac{2}{3}\varphi_{21}\, dx^2 \wedge dx^1 \ ,$$

etc. The later formulas are still equal to φ, but do not represent φ in terms of a basis.

In all the above formulas we may allow X to vary. Then the coefficients become functions, which are of class C^k if the tensor field or differential form is of class C^k, and any tensor field or differential form can be represented in this way. The resulting sections give free $\mathfrak{F}^k$-modules of suitable dimension, which will be denoted by $\overset{r}{\otimes}(\tau^k)^* = (\overset{r}{\otimes}\tau^k)^*$, $\overset{r,s}{\otimes}\tau^k$, $\wedge^p(\tau^k)^* = (\wedge^p\tau^k)^*$. The symbolic identifications implied above represent the identifications which can be made at each X. For example, a differential form of degree p can be considered as a section of class C^k of $D \times \wedge^p V^*$ or as a linear (over $\mathfrak{F}^k$) form on $\wedge^p\tau^k$.

Any two forms φ, ψ at X, or any two differential forms, can be combined by the operation of exterior multiplication to give a form $\varphi \wedge \psi$ which is of degree $p + q$ if φ is of degree p and ψ of degree q (and therefore is $\tilde{\tilde{0}}$ if $p + q > n$). It is left as an exercise to show that the exterior product of two differential forms of class C^k, computed fibrewise, is a differential form of class C^k. Thus we can consider

the "exterior algebra" of differential forms.

§3. Maps; change of coordinates

In Chapter X, we considered sets D, $\tilde{D}$, etc., to which topological structures had been assigned, and a map $F: D \longrightarrow \tilde{D}$ was called continuous (that is, of class C^0) if it induced an inverse correspondence from the open sets of $\tilde{D}$ to the open sets of D (Definition X, 1.22). Further, in the remark following Theorem X, 1.12, we noted that a homeomorphism (a bijective map of class C^0 whose inverse is of class C^0) gives an isomorphism between the topological structures on D and $\tilde{D}$.

We now consider the analogous questions in the case that D and $\tilde{D}$ are open subsets of finite dimensional vector spaces V and $\tilde{V}$ (and thus have the additional structure described in §2), and the map F is of class C^k. This question has not only theoretical, but also practical importance. For example, if we make a change of variables in working out a problem, we want to know whether a solution of the problem in the new variables actually leads to a solution of the original problem.

3.1. Assumptions and notation (through 3.10). Let D be a fixed open subset of a given vector space V of dimension n, let $\tilde{D}$ be an open subset of another vector space $\tilde{V}$ of dimension m, and let

$$F: D \longrightarrow \tilde{D} ,$$

$X \in D \longrightarrow \tilde{X} = F(X) \in \tilde{D}$, be given, where F is of class C^k; the

integer k is fixed but arbitrary, with $k > 0$ (except for 3.2).

We shall give with each construction its expression in terms of coordinates and associated bases, where $(x^1,\ldots,x^n)$ denotes a (linear) coordinate system on V, and $(\tilde{x}^1,\ldots,\tilde{x}^m)$ a (linear) coordinate system on W.

In terms of coordinates, F is expressed by a system of functions

$$(1) \qquad \tilde{x}^i = f^i(X) = f^i(x^1,\ldots,x^n) , \qquad i = 1, 2,\ldots,m ,$$

where $f^i \in \mathfrak{F}^k(D)$ by Proposition 1.5 and has continuous partial derivatives, with respect to the x^j, of all orders $\leqq k$ for values of $(x^1,\ldots,x^n)$ corresponding to points of D (Proposition 1.7).

3.2. <u>Proposition</u>. The map F of 3.1 induces a correspondence

$$F^*: \mathfrak{F}^\ell(\tilde{D}) \longrightarrow \mathfrak{F}^\ell(D) , \qquad \ell \leqq k ,$$

defined for any function $\tilde{f}$ on $\tilde{D}$, by

$$\tilde{f} \longrightarrow f = \tilde{f}F .$$

If F is not of class C^{k+1}, then there exists a function $\tilde{f} \in \mathfrak{F}^{k+1}(\tilde{D})$ such that $F^*\tilde{f} \notin \mathfrak{F}^{k+1}(D)$.

<u>Proof</u>. It is clear that $\tilde{f}F: D \longrightarrow R$ (that is, $F^*\tilde{f}$ is a function on D, for any function $\tilde{f}$ on $\tilde{D}$) and is of class C^ℓ if f is of class C^ℓ by Proposition 1.11, $\ell \leqq k$, since F is of class C^ℓ (Proposition 1.2). If F is not of class C^{k+1},

then at least one function f^i in (1) is not of class C^{k+1} (Proposition 1.5). Since $f^i = F^*\tilde{x}^i$ and the function $\tilde{x}^i$ is of class C^∞ (therefore of class C^{k+1}) on $\tilde{D}$, the function $f = \tilde{x}^i$ provides the required example.

Remark. In terms of coordinates, $f = F^*\tilde{f}$ is expressed by $f(x^1,\ldots,x^n) = \tilde{f}(f^1(x^1,\ldots,x^n),\ldots,f^m(x^1,\ldots,x^n))$. Thus it is clear that F^* sends a constant function on $\tilde{D}$ into a constant function on D, and that $F^*(\tilde{f} + \tilde{g}) = F^*\tilde{f} + F^*\tilde{g}$, $F^*(\tilde{f}\tilde{g}) = (F^*\tilde{f})(F^*\tilde{g})$; that is, F^* may be described as a ring homomorphism which preserves the subring corresponding to the real numbers. If F is not surjective, the values $\tilde{f}(\tilde{X})$ for $\tilde{X}$ not in $F(D)$ are not used in defining $F^*\tilde{f}$.

3.3. Proposition. For each $X \in D$, the map F of 3.1 induces a linear transformation (writing $\tilde{T} = T(\tilde{D})$)

$$F_*: T_X \longrightarrow \tilde{T}_{\tilde{X}} , \qquad \tilde{X} = F(X) ,$$

of the tangent vectors at X into the tangent vectors at $\tilde{X} = F(X)$ where, for $u \in T_X$, the tangent vector $\tilde{u} = F_*u \in \tilde{T}_{F(X)}$ is determined by the condition

$$F_*u\cdot\tilde{f} = u\cdot F^*\tilde{f} , \tag{2}$$

the right-hand member being computed at $X \in D$.

Proof. To see that (2) determines $\tilde{u} = F_*u$ uniquely, recall (2.4, Remarks) that a tangent vector at $\tilde{X} = F(X)$ is uniquely determined by its action on the functions $\tilde{f} \in \mathfrak{F}(\tilde{D})$, and note that this action, as prescribed by (2), has the required

properties (since u is a tangent vector and F^* is a ring homomorphism). Similarly, the definition (2) implies that $F_*(ru + sv) = rF_*u + sF_*v$, $u, v \in T_X$, $r, s \in R$, so F_* is a linear transformation from the vector space T_X into the vector space $\tilde{T}_{\tilde{X}}$.

Remarks. (i) If (2) is compared with (6) of §1, we see that for any $X \in D$, $F_* \in L(T_X, \tilde{T}_{F(X)})$ is essentially the same as $F'(X) \in L(V, \tilde{V})$ of Definition VIII, 1.7. The only difference is that now we consider the transformation as going from T_X, an isomorphic copy of V thought of as standing over the point X, into $\tilde{T}_{F(X)}$ which is a copy of $\tilde{V}$ over $\tilde{X} = F(X)$.

(ii) Equation (2) may be used directly to express F_* in terms of coordinates. The coefficients in the expression $\tilde{u} = \Sigma_{i=1}^m \tilde{u}^i \frac{\partial}{\partial \tilde{x}^i} \in \tilde{T}_{\tilde{X}}$ have the values $\tilde{u}^i = \tilde{u}\cdot\tilde{x}^i$ (see 2.4), so for $\tilde{u} = F_*u$, where $u = \Sigma_{j=1}^n u^j \frac{\partial}{\partial x^j}$, we have

$$(3) \quad \tilde{u}^i = \tilde{u}\cdot\tilde{x}^i = F_*u\cdot\tilde{x}^i = u\cdot F^*\tilde{x}^i = u\cdot f^i = \Sigma_{j=1}^n u^j \frac{\partial f^i}{\partial x^j} = \Sigma_{j=1}^n u^j \frac{\partial \tilde{x}^i}{\partial x^j}$$

(where we have used the standard notation $\frac{\partial \tilde{x}^i}{\partial x^j}$ in place of $\frac{\partial f^i}{\partial x^j}$, since it is clear that $\tilde{x}^i$ is being considered as a function of $(x^1, \ldots, x^n)$, that is, as $F^*\tilde{x}^i$) and

$$(4) \qquad F_*u = \Sigma_{i=1}^m \Sigma_{j=1}^n u^j \frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial}{\partial \tilde{x}^i}$$

Equation (4) also leads to the apparent rule $\frac{\partial}{\partial x^j} = \Sigma_{i=1}^m \frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial}{\partial \tilde{x}^i}$ (correct version: $F_* \frac{\partial}{\partial x^j} = \Sigma_{i=1}^m \frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial}{\partial \tilde{x}^i}$) and to the historical choice of the adjective contravariant for the

case of tangent vectors and those tensors (see below) for which the induced transformation is of the type F_*, that is, from D to $\tilde{D}$, with <u>covariant</u> corresponding to F^*.

(iii) In general, if u is a vector field on D, the values of F_*u, as X varies, need not determine a vector field on $\tilde{D}$, that is, a section of $\tilde{T}$ over $\tilde{D}$. If F is not injective, we may have more than one point X mapping into the same point $\tilde{X}$ and, in general, more than one image point F_*u in the fibre $\tilde{T}_{\tilde{X}}$. If F is not surjective, then for $\tilde{X}$ not in $F(D)$ there is no image point F_*u in $\tilde{T}_{\tilde{X}}$.

3.4. <u>Proposition</u>. For each $X \in D$, the map F of 3.1 induces linear transformations

$$\overset{s}{\otimes} F_*: (\overset{s}{\otimes} T)_X \longrightarrow (\overset{s}{\otimes} \tilde{T})_{\tilde{X}} ,$$

$$\wedge^q F_*: (\wedge^q T)_X \longrightarrow (\wedge^q \tilde{T})_{\tilde{X}} , \qquad \tilde{X} = F(X) ,$$

of the contravariant tensors at X of order s into the contravariant tensors at $\tilde{X} = F(X)$ of order s, $s = 1, 2, \ldots,$ and of the q-vectors at X into the q-vectors at $\tilde{X} = F(X)$, $q \leq n$.

<u>Proof</u>. The linear transformation $\wedge^q F_*$ is obtained from F_* of 3.3 by Theorem IX, 7.5 (and will be the zero transformation if $q > m$, $q \leq n$, which can occur only if $m < n$). The linear transformation $\overset{s}{\otimes} F_*$ is obtained in the same way from Theorem IX, 6.9.

<u>Remarks</u>. Both $\overset{s}{\otimes} F_*$ and $\wedge^q F_*$ are usually denoted simply by F_*. A linear transformation is determined by its values

for basis elements. For $F_* = \overset{s}{\otimes} F_*$ we have

$$F_*\left(\frac{\partial}{\partial x^{j_1}} \otimes \frac{\partial}{\partial x^{j_2}} \otimes \dots \otimes \frac{\partial}{\partial x^{j_s}}\right) = \left(F_*\frac{\partial}{\partial x^{j_1}}\right) \otimes \left(F_*\frac{\partial}{\partial x^{j_2}}\right) \otimes \dots \otimes \left(F_*\frac{\partial}{\partial x^{j_s}}\right) .$$

To obtain an expression in terms of coordinates, we compute the right-hand member according to the computing rules for tensor products. If $t \in (\overset{s}{\otimes} T)_X$ is given by (6) of §2, then the coefficients of $\tilde{t} = F_* t$ in the analogous expression at $\tilde{X}$ are given by

$$(5)\quad \tilde{t}^{i_1 i_2 \dots i_s} = \Sigma_{j_1=1}^{n} \Sigma_{j_2=1}^{n} \dots \Sigma_{j_s=1}^{n} t^{j_1 j_2 \dots j_s} \frac{\partial \tilde{x}^{i_1}}{\partial x^{j_1}} \frac{\partial \tilde{x}^{i_2}}{\partial x^{j_2}} \dots \frac{\partial \tilde{x}^{i_s}}{\partial x^{j_s}} .$$

In the case of q-vectors $(q \leq n,\ q \leq m)$ we have

$$F_*\left(\frac{\partial}{\partial x^{j_1}} \wedge \frac{\partial}{\partial x^{j_2}} \wedge \dots \wedge \frac{\partial}{\partial x^{j_q}}\right) = \left(F_*\frac{\partial}{\partial x^{j_1}}\right) \wedge \left(F_*\frac{\partial}{\partial x^{j_2}}\right) \wedge \dots \wedge \left(F_*\frac{\partial}{\partial x^{j_q}}\right)$$

$$= \Sigma_{i_1<i_2<\dots<i_q} \frac{\partial(\tilde{x}^{i_1}, \tilde{x}^{i_2}, \dots, \tilde{x}^{i_q})}{\partial(x^{j_1}, x^{j_2}, \dots, x^{j_q})} \frac{\partial}{\partial \tilde{x}^{i_1}} \wedge \frac{\partial}{\partial \tilde{x}^{i_2}} \wedge \dots \wedge \frac{\partial}{\partial \tilde{x}^{i_q}} ,$$

where

$$(6)\quad \frac{\partial(\tilde{x}^{i_1}, \tilde{x}^{i_2}, \dots, \tilde{x}^{i_q})}{\partial(x^{j_1}, x^{j_2}, \dots, x^{j_q})} = \det \begin{pmatrix} \frac{\partial \tilde{x}^{i_1}}{\partial x^{j_1}} & \frac{\partial \tilde{x}^{i_1}}{\partial x^{j_2}} & \dots & \frac{\partial \tilde{x}^{i_1}}{\partial x^{j_q}} \\ \frac{\partial \tilde{x}^{i_2}}{\partial x^{j_1}} & \frac{\partial \tilde{x}^{i_2}}{\partial x^{j_2}} & \dots & \frac{\partial \tilde{x}^{i_2}}{\partial x^{j_q}} \\ \cdot & \cdot & \dots & \cdot \\ \frac{\partial \tilde{x}^{i_q}}{\partial x^{j_1}} & \frac{\partial \tilde{x}^{i_q}}{\partial x^{j_2}} & \dots & \frac{\partial \tilde{x}^{i_q}}{\partial x^{j_q}} \end{pmatrix}$$

For the particular case $q = m = n$, this is the Jacobian determinant of the coordinate representation of F and the notation is borrowed from the standard notation for Jacobian determinants. If $u \in (\Lambda^q T)_X$ is given by (7) of §2, then in the analogous expression for $\tilde{u} = F_* u$, the coefficients are

$$(7) \quad \tilde{u}^{i_1 i_2 \cdots i_q} = \Sigma_{j_1 < j_2 < \cdots < j_q} \frac{\partial(\tilde{x}^{i_1}, \tilde{x}^{i_2}, \ldots, \tilde{x}^{i_q})}{\partial(x^{j_1}, x^{j_2}, \ldots, x^{j_q})} u^{j_1 j_2 \cdots j_q} .$$

3.5. <u>Proposition</u>. Let $\tilde{\tau} = \tau(\tilde{D})$. Then the map F of 3.1 induces a linear transformation (over R)

$$F^*: (\tilde{\tau}^\ell)^* \longrightarrow (\tau^\ell)^* , \qquad \ell \leq k - 1 ,$$

of the differential forms of degree 1 on $\tilde{D}$ into the differential forms of degree 1 on D where, for $\tilde{\omega} \in \tilde{\tau}^*$, the differential form $F^*\tilde{\omega}$ is determined at each $X \in D$ by the condition

$$(8) \qquad < u, F^*\tilde{\omega} > = < F_* u, \tilde{\omega} > \qquad \text{for all } u \in T_X ,$$

the right-hand member being evaluated at $\tilde{X} = F(X)$.

<u>Proof</u>. Obviously, (8) is the condition that

$$(9) \qquad F^*: (\tilde{T}^*)_{F(X)} \longrightarrow (T^*)_X$$

be the transpose of F_* of 3.3, so F^* sends a 1-form at $F(X)$ into a 1-form at X. It is also clear that the image points $F^*\tilde{\omega}$ of a section $\tilde{\omega}$ of $\tilde{T}^*$ over $\tilde{D}$ determine a section of T^* over D, since a unique value is determined in each fibre $(T^*)_X$. It remains only to show that $\omega = F^*\tilde{\omega}$ is of class C^ℓ if $\tilde{\omega}$ is of

class C^{ℓ}, $\ell \leqq k - 1$; this will follow from the expression of ω in terms of coordinates on V. If $\tilde{\omega} = \Sigma_{i=1}^{m} \tilde{\omega}_i d\tilde{x}^i$, and $\omega = F^*\tilde{\omega} = \Sigma_{j=1}^{n} \omega_j dx^j$, then

$$\begin{aligned} \omega_j &= < \frac{\partial}{\partial x^j}, \omega > = < \frac{\partial}{\partial x^j}, F^*\tilde{\omega} > = < F_*\frac{\partial}{\partial x^j}, \tilde{\omega} > \\ &= < \Sigma_{i=1}^{m} \frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial}{\partial \tilde{x}^i}, \tilde{\omega} > = \Sigma_{i=1}^{m} \frac{\partial \tilde{x}^i}{\partial x^j} \tilde{\omega}_i \end{aligned} \tag{10}$$

and

$$F^*\tilde{\omega} = \Sigma_{j=1}^{n} \Sigma_{i=1}^{m} \tilde{\omega}_i \frac{\partial \tilde{x}^i}{\partial x^j} dx^j .$$

In (10), as X varies, we have $\omega_j(X) = \Sigma_{i=1}^{m} \tilde{\omega}_i(F(X)) \frac{\partial \tilde{x}^i}{\partial x^j}(X)$. The factors $\tilde{\omega}_i(F(X))$ are of class C^{ℓ} for $\ell \leqq k$ if $\tilde{\omega}_i$ is of class C^{ℓ}, by Proposition 1.11, but the factors $\frac{\partial \tilde{x}^i}{\partial x^j}$ are only class C^{k-1} in general. If F is not of class C^{k+1}, then $\frac{\partial \tilde{x}^i}{\partial x^j}$ is not of class C^k for some choice of i and j, and it is clear that, for $\tilde{\omega} = d\tilde{x}^i$ for this i, we have $(F^*\tilde{\omega})_j$ not of class C^k.

3.6. <u>Proposition</u>. The map F of 3.1 induces linear transformations (over R)

$$\otimes^r F^* : \otimes^r (\tilde{\tau}^{\ell})^* \longrightarrow \otimes^r (\tau^{\ell})^* ,$$

$$\wedge^p F^* : \wedge^p (\tilde{\tau}^{\ell})^* \longrightarrow \wedge^p (\tau^{\ell})^* , \qquad \ell \leqq k - 1 ,$$

of the covariant tensor fields of order r on $\tilde{D}$ into the covariant tensor fields of order r on D, and of the differential forms of degree p on $\tilde{D}$ into the differential forms of degree p on D.

<u>Proof</u>. The linear transformations F^* may be defined

either by extending (9) to the tensor and exterior products formed from $(\tilde{T}^*)_{F(X)}$ and $(T^*)_X$ or, in analogy with (8), by taking the duals of $\overset{r}{\otimes} F_*$ and $\wedge^p F_*$; the resulting linear transformations are the same (by Exercise IX, 10.2 and Theorem IX, 9.9) since the vector spaces in question are finite dimensional. For example, if $\tilde{\varphi} \in \wedge^p(\tilde{\tau}^\ell)^*$, the p-form $F^*\tilde{\varphi}$ is determined at each $X \in D$ by the condition

$$(11) \qquad \langle u, F^*\tilde{\varphi} \rangle = \langle F_* u, \tilde{\varphi} \rangle \quad \text{for all } u \in (\wedge^p T)_X ,$$

the right-hand member being evaluated at $\tilde{X} = F(X)$; $F^*\tilde{\varphi}$ is zero if $p > n$.

The expression of F^* in terms of coordinates is then obtained by duality from the expressions for F_*. Just as (10) is dual to (3), we obtain from (7)

$$(12) \quad \varphi_{j_1 j_2 \ldots j_p} = \Sigma_{i_1 < i_2 < \ldots < i_p} \frac{\partial(\tilde{x}^{i_1}, \tilde{x}^{i_2}, \ldots, \tilde{x}^{i_p})}{\partial(x^{j_1}, x^{j_2}, \ldots, x^{j_p})} \tilde{\varphi}_{i_1 i_2 \ldots i_p}$$

if

$$\tilde{\varphi} = \Sigma_{i_1 < i_2 < \ldots < i_p} \tilde{\varphi}_{i_1 i_2 \ldots i_p} \, d\tilde{x}^{i_1} \wedge d\tilde{x}^{i_2} \wedge \ldots \wedge d\tilde{x}^{i_p} ,$$

and

$$\varphi = F^*\tilde{\varphi} = \Sigma_{j_1 < j_2 < \ldots < j_p} \varphi_{j_1 j_2 \ldots j_p} \, dx^{j_1} \wedge dx^{j_2} \wedge \ldots \wedge dx^{j_p} .$$

Similarly, for $t = F^*\tilde{t}$, where $\tilde{t} \in \overset{r}{\otimes}(\tilde{\tau}^\ell)^*$, the dual of (5) gives

$$(13) \quad t_{j_1 j_2 \ldots j_r} = \Sigma_{i_1=1}^m \Sigma_{i_2=1}^m \ldots \Sigma_{i_r=1}^m \frac{\partial \tilde{x}^{i_1}}{\partial x^{j_1}} \frac{\partial \tilde{x}^{i_2}}{\partial x^{j_2}} \ldots \frac{\partial \tilde{x}^{i_r}}{\partial x^{j_r}} \tilde{t}_{i_1 i_2 \ldots i_r}$$

for the coefficients of t in terms of those for $\tilde{t}$ in the expressions for t, $\tilde{t}$ corresponding to (10) of §2.

Clearly, the coefficients of $\varphi = F^*\tilde{\varphi}$ and of $t = F^*\tilde{t}$, as X varies, cannot be better than class C^{k-1}, if F is not of class C^{k+1}, and are of class C^ℓ if $\tilde{\varphi}$ or $\tilde{t}$, respectively, are of class C^ℓ, $\ell \leqq k - 1$.

3.7. Proposition. The linear transformation F^* induced by F of 3.1 is a homomorphism of the exterior algebras of differential forms:

$$F^*(\tilde{\varphi} + \tilde{\psi}) = F^*\tilde{\varphi} + F^*\tilde{\psi}\ , \tag{14}$$

$$F^*(\tilde{\varphi} \wedge \tilde{\psi}) = (F^*\tilde{\varphi}) \wedge (F^*\tilde{\psi})\ . \tag{15}$$

Remarks. The formulas (14) and (15) follow from the properties of F^* for each $X \in D$. To interpret F^* as a homomorphism, we must specify what is meant by the "exterior algebra of differential forms". (i) For fixed class C^ℓ, $\ell \leqq k - 1$, we have a commutative graded algebra $Z = Z(D)$, scalars R, with

$$(Z)_0 = \mathfrak{F}^\ell\ ,$$

$$(Z)_p = \Lambda^p(\tau^\ell)^*\ , \qquad 1 \leqq p \leqq n\ ,$$

$$(Z)_p = \tilde{0}\ , \qquad p > n\ ,$$

and an analogous $\tilde{Z}$ corresponding to $\tilde{D}$. Then F^* is a homomorphism, since F^* sends $R\subset(\tilde{Z})_0$ into $R\subset(Z)_0$ and $F^*(r\tilde{\varphi}) = rF^*\tilde{\varphi}$, by (15). In this case, addition in (14) is defined only for differential forms of the same degree. (ii) As in

Chapter IX, we may construct the direct sum of the (infinite dimensional) vector spaces $(Z)_p$, $p = 0, 1, \ldots, n$. We then have addition of differential forms of arbitrary degree, and an algebra (scalars R) using the exterior product of differential forms as multiplication. Then F^* is a homomorphism. (iii) We may also construct the direct sum of the $\mathfrak{F}^\ell$-modules $\mathfrak{F}^\ell$, $\Lambda^p(\tau^\ell)^*$, $p = 1, \ldots, n$, and use the exterior product of differential forms to make the resulting $\mathfrak{F}^\ell$-module into an "algebra" with scalars $\mathfrak{F}^\ell$. Then F^* is not a homomorphism because the resulting algebras on D and $\tilde{D}$ do not have the same ring of scalars. In fact, F^* is not even linear for the same reason.

<u>Summary of 3.2 - 3.7</u>. If $F: D \longrightarrow \tilde{D}$ is a map of class C^k, $k > 1$, then F induces linear transformations F_* of contravariant objects at $X \in D$ into the contravariant objects of the same type at $\tilde{X} = F(X)$, and linear transformations (over R) F^* of covariant sections over $\tilde{D}$ into the covariant sections over D of the same type, which preserve structure of class C^ℓ, $\ell \leq k - 1$.

<u>Remark</u>. Note that the "differentiable structure" is used only in defining F^* on functions and F_* on tangent vectors. The remaining correspondences are then determined algebraically, by the constructions of Chapter IX, at each point.

3.8. <u>Theorem</u>. Let $m = n$ in 3.1; then $F: D \longrightarrow \tilde{D}$ of class C^k, $k > 1$, is a homeomorphism with inverse $G: \tilde{D} \longrightarrow D$ of class C^k if and only if (a) F is bijective and (b) F_* is non-singular for each $X \in D$.

Proof. If F satisfies (a), there is a unique map $G: \tilde{D} \longrightarrow D$ satisfying GF = identity on D. By (b) and Theorems X, 4.1, 4.5 and Corollary 1.12, for each $X \in D$ there is an open set D_o containing X such that G is of class C^k on the open set $F(D_o)$ containing $\tilde{X} = F(X)$. Then G is of class C^k on $\tilde{D}$ since every $\tilde{X} \in \tilde{D}$ is contained in an open set of the type $F(D_o)$.

Conversely, if F has an inverse G, then F is bijective. If F and G are of class C^k, $k > 1$, then

$$G_*F_*: T_X \longrightarrow T_X$$

is defined and is the identity transformation on T_X, so F_* is non-singular. In fact, for $u \in T_X$, G_*F_*u is determined by the condition

$$G_*F_*u \cdot f = u \cdot f \qquad \text{for all } f \in \mathfrak{F}^k(D) ,$$

which implies that $G_*F_*u = u$, since a tangent vector is uniquely determined by its action on functions (see 2.4).

Remarks. If F is expressed in terms of coordinates on D and $\tilde{D}$, then (b) becomes: (b') the Jacobian determinant $\dfrac{\partial(\tilde{x}^1, \ldots, \tilde{x}^n)}{\partial(x^1, \ldots, x^n)}$ is different from zero at each $X \in D$. Similarly, G_*F_* = identity is expressed by

$$\Sigma_{i=1}^n \frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial x^\ell}{\partial \tilde{x}^i} = \delta_j^\ell \tag{15}$$

and F_*G_* = identity by

$$\Sigma_{j=1}^n \frac{\partial x^j}{\partial \tilde{x}^i} \frac{\partial \tilde{x}^\ell}{\partial x^j} = \delta_i^\ell \quad . \tag{16}$$

3.9. Corollary. If $D \subset V$ is given, where D is open and V is of dimension n, and if $f^1, \ldots, f^n \in \mathcal{F}^k(D)$, $k \geq 1$, satisfy (a') for each pair $X_1, X_2 \in D$, we have $f^i(X_1) \neq f^i(X_2)$ for at least one value of $i = 1, 2, \ldots, n$; (b') for each $X \in D$, the Jacobian determinant $\dfrac{\partial(f^1, f^2, \ldots, f^n)}{\partial(x^1, x^2, \ldots, x^n)} \neq 0$, where $(x^1, \ldots, x^n)$ are coordinates on D, then the map $F: D \longrightarrow R^n$ determined by the component functions f^i is of class C^k and gives a homeomorphism of D with an open set $F(D) \subset R^n$; moreover, $G = F^{-1}$ exists and is of class C^k.

3.10. Theorem. If $m = n$ in 3.1 and if $F: D \longrightarrow \tilde{D}$ of class C^k, $k \geq 1$, has an inverse $F^{-1}: \tilde{D} \longrightarrow D$ of class C^k, then the associated structures on D and $\tilde{D}$ are equivalent up to class C^{k-1}.

Proof. The word "equivalent" means that there is a bijective correspondence between structures on D and structures on $\tilde{D}$, for each type of structure, and that the notion of class C^ℓ, $\ell \leq k - 1$, is preserved

For example, in addition to

$$F^*: (\tilde{\tau}^\ell)^* \longrightarrow (\tau^\ell)^*, \qquad \ell \leq k - 1,$$

of Proposition 3.5, we also have

$$F^{-1*}: (\tau^\ell)^* \longrightarrow (\tilde{\tau}^\ell)^*, \qquad \ell \leq k - 1,$$

and it is easily verified that $(F^{-1})^* F^*$ is the identity correspondence, so F^* is bijective.

Further, there is no longer any difficulty about allowing X to vary as we consider F_* (see Remark (iii) after 3.3). Since F is bijective, two different fibres T_{X_1} and T_{X_2} cannot map into the same fibre $\tilde{T}_{\tilde{X}}$; since F_* is non-singular, the map $T_X \longrightarrow \tilde{T}_{\tilde{X}}$, $\tilde{X} = F(X)$, is bijective; since F is surjective, every $\tilde{T}_{\tilde{X}}$ is the bijective image of some T_X. Thus

$$F_*: T \longrightarrow \tilde{T}$$

is bijective, and fibre-preserving (definition ?) and of class C^{k-1} with inverse of class C^{k-1}. To verify this last statement, note that a (linear) coordinate system $(x^1,\ldots,x^n,u^1,\ldots,u^n)$ can be introduced on T and a (linear) coordinate system $(\tilde{x}^1,\ldots,\tilde{x}^n, \tilde{u}^1,\ldots,\tilde{u}^n)$ on $\tilde{T}$, in terms of which F_* has the component functions

$$\tilde{x}^i = f^i(x^1,\ldots,x^n) ,$$

$$\tilde{u}^i = \Sigma_{j=1}^n u^j \frac{\partial f^i}{\partial x^j} (x^1,\ldots,x^n) , \qquad i = 1,\ldots,n .$$

The f^i are the component functions of the given map F and are of class C^k (therefore also of class C^{k-1}) with respect to the coordinates x^i, and also with respect to the u^j which do not appear in the expression for f^i. The remaining component functions $\Sigma_{j=1}^n u^j \frac{\partial f^i}{\partial x^j}$ are of class C^{k-1} with respect to the coordinates x^i and linear in the coordinates u^j, so are of class C^{k-1} with respect to $(x^1,\ldots,x^n, u^1,\ldots,u^n)$.

Similarly, it can be verified that

$$F_*: \tau^\ell \longrightarrow \tilde{\tau}^\ell , \qquad \ell \leqq k - 1,$$
$$F^*: \tilde{T}^* \longrightarrow T^*$$

etc., where all maps are bijective.

Finally, because of the existence of F^{-1}, a transformation of mixed tensors is defined. For example,

$$dx^i \otimes \frac{\partial}{\partial x^{j_1}} \otimes \frac{\partial}{\partial x^{j_2}} \longrightarrow ((F^{-1})^* dx^i) \otimes (F_* \frac{\partial}{\partial x^{j_1}}) \otimes (F_* \frac{\partial}{\partial x^{j_2}}) ,$$

and it can be verified that these maps are bijective and take tensor fields of class C^ℓ into fields of the same type, of class C^ℓ, $\ell \leqq k - 1$.

3.11. Remarks. In theoretical work on this subject it is frequently found convenient to consider only class C^∞. It is clear that this is the only class which is preserved under differentiation and mappings. Further, the assumption of class C^∞ avoids cluttering up statements of theorems with precise statements of differentiability hypotheses; in many cases, simpler proofs or axioms can be used for this class (see, for example, 2.4). Thus the overall picture can be presented more clearly. Finally, no finite value of k selected in advance can be sufficient for a complete theory (for example, partial differential equations of high order would be excluded). However, a finite value of k is often sufficient for any particular theorem or problem. [A different type of theoretical question is that of determining the minimum differentiability conditions

under which a given theorem can be proved; in this type of work further subdivisions of the classes C^k are introduced. Also it should be mentioned that an assumption stronger than class C^∞ can be made: that of real analyticity, usually denoted by C^ω. Analyticity is needed to use power series methods.]

3.12. Definition. The adjective differentiable will be used in the sense of "differentiable of class C^∞". A map F will be called bidifferentiable if it is differentiable and if F^{-1} exists and is differentiable. If V is a vector space of dimension n and $D \subset V$ is open, the set of differential forms on D of degree p of class C^∞ will be denoted by $A^p = A^p(D)$, where $A^0 = \mathcal{F}^\infty(D)$; then $A = A(D)$ will denote the exterior algebra

$$A = A^0 \oplus A^1 \oplus \ldots \oplus A^n$$

where n is the dimension of V, $D \subset V$.

3.13. Change of coordinates. A linear coordinate system on $D \subset V$ is introduced (Definition 1.6) by means of a linear transformation $T: V \longrightarrow R^n$, that is, a bidifferentiable map with inverse $S = T^{-1}$, which identifies D with the open set $T(D) \subset R^n$; the coordinates assigned to a point $X \in D$ are those of the point $T(X) \in R^n$. The expression of a function f on D in terms of these coordinates is the same as the expression of the function $fS = S^*f$ on $T(D)$ in terms of the euclidean coordinates (which give a linear coordinate system on $T(D)$), but the explicit mention of S is suppressed. The expression of a tangent vector u at $X \in D$ in terms of a base associated

with these coordinates is the same as the expression for T_*u at $T(X)$, but T_* is not mentioned explicitly.

An arbitrary bidifferentiable map $F: D \longrightarrow \tilde{D} \subset R^n$ may be used in exactly the same way to introduce a coordinate system on D, starting from the euclidean coordinates $(\tilde{x}^1,\ldots,\tilde{x}^n)$ on $\tilde{D} = F(D)$. Each point $X \in D$ is assigned the coordinates of the point $F(X) \in \tilde{D}$; corresponding to these coordinates, each tangent vector u at $X \in D$ is assigned the expression for F_*u at $F(X) \in \tilde{D}$, etc. Any computation, such as adding two tangent vectors u, v at $X \in D$, may equally well be carried out in terms of the new coordinate representation, since $F_*u + F_*v = F_*(u + v)$, etc., and all correspondences are bijective.

If we also have a linear coordinate system $(x^1,\ldots,x^n)$ on D, based on a linear transformation $T: D \longrightarrow T(D)$, the relation between the coefficients in the two expressions for $u \in T_X$, say, is given by (3); but since the expression for F_*u is now the expression for u in terms of the new coordinates, we write (instead of (4)),

$$u = \Sigma_{j=1}^n u^j \frac{\partial}{\partial x^j} \left(= \Sigma_{i,j=1}^n u^j \frac{\partial \tilde{x}^i}{\partial x^j} \frac{\partial}{\partial \tilde{x}^i}\right) = \Sigma_{i=1}^n \tilde{u}^i \frac{\partial}{\partial \tilde{x}^i} .$$

The component functions $\tilde{x}^i = f^i(x^1,\ldots,x^n)$ expressing F in terms of linear coordinates are also the component functions of the map

$$FT^{-1}: T(D) \longrightarrow F(D) .$$

Similarly, if $\hat{F}: D \longrightarrow \hat{F}(D) \subset R^n$ is another map used to introduce

still another system of coordinates on D, no new formulas are needed to express the change from one system of coordinates to another. For example, if u is a tangent vector at $X \in D$, the expressions for u in terms of the two coordinate systems are the same as those for F_*u and $\hat{F}_*u$, respectively, and the relation between them is computed from the expressions for the map $\hat{F}F^{-1}: F(D) \longrightarrow \hat{F}(D)$ in terms of the linear coordinates on R^n. This interpretation also shows that any relation previously stated only for linear coordinate systems is also valid for arbitrary coordinate systems.

The many formulas expressing the transformation of coordinates, etc., need not be memorized, but can be reconstructed as needed, once the following rules are noted.

(i) The indices of the coordinates of a point X are written in superscript position, as in x^i; then the indices of the associated contravariant basis elements appear in subscript position, as in $\frac{\partial}{\partial x^i}$, and those of covariant basis elements in superscript position, as in dx^i.

(ii) The indices attached to the coefficients of a contravariant object are written as superscripts (and those of a covariant object as subscripts) so that summation is always over a pair of indices with one above one below, and <u>both referring to the same coordinate system</u>. [When the "summation convention" is used, the symbol Σ is omitted when the range of the summation is clear, and it is

understood that the expression is to be summed over any such pair of indices.]

(iii) Any index which is not summed must appear on both sides of an equation, and as a superscript on both sides, or as a subscript on both sides.

(iv) The same general rules apply in the case of the formulas expressing the correspondences induced by a differentiable map F in terms of coordinates except that, if F is given by (1), for example, any formula implying that x^j can be expressed as a function of the $\tilde{x}^i$'s must be discarded. This is obvious in the case of factors such as $\frac{\partial x^j}{\partial \tilde{x}^i}$, but is also the reason why X cannot be allowed to vary in a formula such as (4) which is valid at $\tilde{X} = F(X)$ but involves $u^j = u^j(X) = u^j(x^1,\ldots,x^n)$ if u is a vector field.

3.14. <u>Differentiable manifolds</u>. We have considered the case that D is an open subset of a finite dimensional vector space V. More generally, if D is a topological space, a homeomorphism F of D with an open subset of R^n, say, will serve to introduce the same associated structures for D: a function $f: D \longrightarrow R$ is called differentiable on D if and only if the function $fF^{-1}: F(D) \longrightarrow R$ is differentiable. The tangent vectors at $X \in D$ are defined by their action on differentiable functions (cf. Remark (ii), 2.4), etc. Then the map F is bidifferentiable by construction. Any bidifferentiable change of coordinates on $F(D)$ gives a change of coordinates on D.

More generally, the structure of a "differentiable manifold of dimension n" is defined on a topological space D (usually assumed to be a Hausdorff space, etc.) by giving a collection of homeomorphisms of open sets of D with open subsets of R^n in such a way that (i) each point $X \in D$ is contained in some open $D_\alpha \subset D$ for which a homeomorphism F_α of D_α with an open subset of R^n is given, and (ii) different homeomorphisms agree in determining whether or not a given function f is differentiable in a neighborhood of X, that is, if $X \in D_\alpha \cap D_\beta$, then $F_\alpha F_\beta^{-1}$ is bidifferentiable on the open set $F_\beta^{-1}(D_\alpha \cap D_\beta) \subset R^n$. The coordinates of $F_\alpha(D_\alpha) \subset R^n$ serve to introduce coordinates on D_α which are called <u>local coordinates</u> since the coordinate system is not given for the whole of D. [Note: for an arbitrary topological space D, it need not be possible to construct a suitable collection of local homeomorphisms with open sets in R^n, for any choice of n.]

In the case $D \subset V$, a local coordinate system can be introduced to express a computation, etc. which depends only on the structure in a neighborhood of a fixed point $X \in D$. We shall require only that the local homeomorphism used shall be bidifferentiable, so as to give an exact correspondence of all structures involved.

3.15. <u>Remarks</u>. So far we have emphasized the properties which are invariant under bidifferentiable maps or coordinate changes. It is clear that a differential form or vector field may have constant coefficients relative to a given coordinate system, but will not have the same property relative to another coordinate

system if general changes of coordinates are allowed. [The use of change of variables in solving problems is a sort of converse: look for a coordinate system in which the given elements of a problem have as simple an expression as possible.] Similarly, a straight line segment in D need not correspond to a straight line segment in R^n for a general system of coordinates, and vice versa. [For this reason, coordinate systems not obtained by means of a linear transformation are sometimes called curvilinear coordinate systems.] Finally, computations involving the operator $\mathfrak{D}_u$, except on functions, do not preserve their form. For example, $F_*(\mathfrak{D}_u v) \neq \mathfrak{D}_{F_*u} F_* v$ in general, where v is a vector field (unless F is obtained from a linear transformation $V \longrightarrow \tilde{V}$). Again, if $F: D \longrightarrow \tilde{D}$ is an arbitrary map (not necessarily a homeomorphism) and if $\tilde{\varphi}$ is a differential form of degree $p > 0$ on $\tilde{D}$ then, in general, $\mathfrak{D}_u(F^*\tilde{\varphi}) \neq F^*(\mathfrak{D}_{F_*u}\tilde{\varphi})$. [This does not contradict (6) of §1, since $F^*\tilde{\varphi}$ is not given by $\tilde{\varphi}F$, for $p > 0$.]

One solution is to define another differentiation operator for vector <u>fields</u> only, the so-called "Lie derivative", which satisfies $F_*(\mathfrak{L}_u v) = \mathfrak{L}_{F_*u} F_* v$, etc., and which satisfies $\mathfrak{L}_u = \mathfrak{D}_u$ when u is a constant vector field on D, but involves a more complicated procedure in the case of non-constant vector fields, such as F_*u under a non-linear bidifferentiable map F.

Another solution is to restrict our attention to those combinations of differential operators which do preserve their form under mappings. Such operators, which have an invariant

meaning, usually have a close relation to "good" problems, which are themselves invariant. One such combination is the bracket product (§1) of two vector fields, and it is left as an exercise to verify that $F_*[u, v] = [F_*u, F_*v]$. Another such is the exterior derivative, operating on differential forms, which will be studied in §4.

§4. The exterior derivative

Let

$$A = A^0 \oplus A^1 \oplus \ldots \oplus A^n$$

be the exterior algebra of differential forms associated with an open set $D \subset V$, where $A^p = A^p(D)$ is the set of differential forms on D of degree p, of class C^∞, with $A^0 = \mathcal{F}^\infty(D)$, and $A^p = \tilde{0}$ for $p > n = \dim V$. The algebra A is a finite dimensional A^0-module (with the additional operation of exterior multiplication) if the elements of the ring A^0 are taken as scalars, or an infinite dimensional vector space over the real numbers, which correspond to the constant functions. The form which is the zero form in each fibre will be denoted simply by 0.

4.1. Definition. The exterior derivative is the operator

$$d: A \longrightarrow A$$

defined by the following axioms:

(1) $\quad d(\varphi + \psi) = d\varphi + d\psi \,, \qquad \varphi, \psi \in A \,,$

(2) $\quad d(\varphi \wedge \psi) = d\varphi \wedge \psi + (-1)^p \varphi \wedge d\psi \,, \qquad \varphi \in A^p, \psi \in A \,,$

(3) for $f \in A^0$, df is the differential form of degree 1 determined at each $X \in D$ by

$$< u, df > = u \cdot f \qquad \text{for all } u \in T_X ,$$

(4) $$d^2 f = d(df) = 0 , \qquad f \in A^0 .$$

<u>Proof</u>. We must show that an operator d satisfying the above axioms exist and is uniquely determined by them.

For a function $f \in A^0$, the condition (3) clearly defines a 1-form at each $X \in D$ and therefore determines a section df of T^* over D; but it must be shown that this section is differentiable, that is, $df \in A^1 \subset A$. Let $(x^1,\ldots,x^n)$ be a system of coordinates on D. For any 1-form ω, we have $\omega = \Sigma_{i=1}^n \omega_i \, dx^i$ with $\omega_i = < \frac{\partial}{\partial x^i}, \omega >$. In particular,

(5) $$df = \Sigma_{i=1}^n < \frac{\partial}{\partial x^i}, df > dx^i = \Sigma_{i=1}^n \frac{\partial f}{\partial x^i} dx^i ,$$

which is differentiable since the coefficients are differentiable. Note that, for $f = x^i$, df coincides with the basis element dx^i, which explains the choice of notation for the associated basis elements.

By (4), $d^2 x^i = 0$, $i = 1,\ldots,n$; then by (2),

(6) $$d(dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_p}) = 0$$

for the basis elements of A^p, $p > 0$. If $\varphi \in A^p$, we have

(7) $$\varphi = \Sigma_{i_1<i_2<\ldots<i_p} \varphi_{i_1 i_2 \ldots i_p} dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_p}$$

where $\varphi_{i_1 i_2 \ldots i_p} = < \frac{\partial}{\partial x^{i_1}} \wedge \frac{\partial}{\partial x^{i_2}} \wedge \ldots \wedge \frac{\partial}{\partial x^{i_p}}, \varphi > \in A^0$. Then by (1), (2), and (6),

$$(8) \qquad d\varphi = \Sigma_{i_1<i_2<\ldots<i_p} d\varphi_{i_1 i_2 \ldots i_p} \wedge dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_p}$$

with $d\varphi_{i_1 i_2 \ldots i_p} \in A^1$. In particular, $d\varphi \in A^{p+1}$, that is,

$$(9) \qquad \begin{aligned} &d: A^p \longrightarrow A^{p+1}, \qquad p = 0, 1, \ldots, n-1, \\ &d: A^n \longrightarrow \tilde{0}. \end{aligned}$$

An arbitrary $\varphi \in A$ is a sum of forms of degree p, $p = 0, 1, \ldots, n$, and $d\varphi$ can be computed by (1).

Further, (4) and (8) show that

$$(10) \qquad d^2\varphi = 0, \qquad \varphi \in A.$$

The above considerations show that an operator d exists, since we can compute $d\varphi$ for any $\varphi \in A$, and it is clear that each step of the computation is uniquely determined by the axioms, which are themselves stated independently of any coordinate system.

<u>Remarks</u>. If A is considered as a graded vector space, then D is a linear transformation of degree 1 (using Proposition 4.2). Because of (2), the operator d does not give a homomorphism; an operator having the properties (1) and (2) is called an <u>anti-derivation</u>.

For given $\varphi \in A^p$, as in (7), formula (8) does not give an expression for $d\varphi$ in terms of a basis for A^{p+1}. If we write

$$d\varphi = \Sigma_{j_1<j_2<\cdots<j_{p+1}} (d\varphi)_{j_1 j_2 \cdots j_{p+1}} \, dx^{j_1} \wedge dx^{j_2} \wedge \cdots \wedge dx^{j_{p+1}} ,$$

then the coefficient $(d\varphi)_{j_1 j_2 \cdots j_{p+1}} \in A^0$ is given by

$$(d\varphi)_{j_1 j_2 \cdots j_{p+1}} = \left\langle \frac{\partial}{\partial x^{j_1}} \wedge \frac{\partial}{\partial x^{j_2}} \wedge \cdots \wedge \frac{\partial}{\partial x^{j_{p+1}}}, d\varphi \right\rangle$$

(11)

$$= \Sigma_{\ell=1}^{p+1} (-1)^{\ell-1} \frac{\partial \varphi_{j_1 \cdots j_{\ell-1} j_{\ell+1} \cdots j_{p+1}}}{\partial x^{j_\ell}} ,$$

which is obtained from (8) by computing $d\varphi_{i_1 i_2 \cdots i_p}$ as in (5) and rearranging terms. For example, for $p = 1$, if $\omega = \Sigma_{i=1}^n \omega_i dx^i$, we compute

$$\begin{aligned}
d\omega &= \Sigma_{i=1}^n d\omega_i \wedge dx^i \\
&= \Sigma_{i=1}^n \Sigma_{k=1}^n \frac{\partial \omega_i}{\partial x^k} dx^k \wedge dx^i \\
&= \Sigma_{i=1}^n \left(\Sigma_{k=1}^{i-1} \frac{\partial \omega_i}{\partial x^k} dx^k \wedge dx^i + \Sigma_{k=i+1}^n \frac{\partial \omega_i}{\partial x^k} dx^k \wedge dx^i \right) \\
&= \Sigma_{k=1}^n \Sigma_{i=1}^{k-1} \frac{\partial \omega_k}{\partial x^i} dx^i \wedge dx^k + \Sigma_{i=1}^n \Sigma_{k=i+1}^n \frac{\partial \omega_i}{\partial x^k} dx^k \wedge dx^i \\
&= \Sigma_{i<k} \left(\frac{\partial \omega_k}{\partial x^i} - \frac{\partial \omega_i}{\partial x^k} \right) dx^i \wedge dx^k .
\end{aligned}$$

At the next to the last step we have used the "trick" of relabelling indices which are summed out anyway. A judicious choice here makes the final collection of terms obvious.

A general formula for $\langle u_1 \wedge \cdots \wedge u_{p+1}, d\varphi \rangle$, expressed in terms of invariant operations and for arbitrary vector fields u_ℓ, will not be given here; see Exercise 6.5.

It is left as an exercise to verify that, for $\varphi \in A^p$, the form $d\varphi \in A^{p+1}$ coincides with the linear transformation on the $(p+1)$-vectors defined by

$$(12)\quad \begin{aligned} &< u_1 \wedge \ldots \wedge u_{p+1}, d\varphi > \\ &= \Sigma_{\ell=1}^{p+1} (-1)^{\ell-1} < u_1 \wedge \ldots \wedge u_{\ell-1} \wedge u_{\ell+1} \wedge \ldots \wedge u_{p+1}, \mathfrak{D}_{u_\ell} \varphi > \end{aligned}$$

for tangent vectors or vector fields u_ℓ, $\ell = 1, \ldots, p+1$. Formula (11) coincides with (12) in the case $u_\ell = \frac{\partial}{\partial x^{j_\ell}}$ if the coordinate system is linear (this being the only case in which we know how to compute $\mathfrak{D}_{u_\ell} \varphi$ in terms of coordinates). [Hints: Note first that the right-hand member of (12) is linear and skew-symmetric in the u_ℓ's and therefore determines a differential form "$d\varphi$" of degree $p+1$. It is then sufficient to verify that, if we take (12) as the definition of d for $\varphi \in A^p$, $p = 0, 1, \ldots, n$, the operator so defined satisfies the axioms of Definition 4.1. For (3), $p = 0$, this is trivial; for (4), compute $< u, \mathfrak{D}_v \omega >$ for any form ω of degree 1 and use this result to compute $< u \wedge v, d(df) >$. For (1) and (2), it is sufficient to verify that $\mathfrak{D}_u(\varphi + \psi) = \mathfrak{D}_u \varphi + \mathfrak{D}_u \psi$, and that $\mathfrak{D}_u(\varphi \wedge \psi) = \mathfrak{D}_u \varphi \wedge \psi + \varphi \wedge \mathfrak{D}_u \psi$.] Thus the operator d can be expressed in terms of the operators $\mathfrak{D}_u$.

4.2. <u>Proposition</u>. For $f \in A^0$, we have $df = 0$ if and only if f is constant on each connected component of D. If A is considered as a vector space over the real numbers, then d is a linear transformation $A^p \longrightarrow A^{p+1}$ and an endomorphism of A.

Proof. If f is constant, it is clear that $df = 0$. Conversely, if $df = 0$, we conclude from (5) that all partial derivatives of f vanish, since the dx^i form a basis for the 1-forms. However, unless D is connected, we cannot conclude that f is constant — if D consists of disjoint pieces, f can have a different constant value on each piece and still be differentiable on D. The linearity of d over the real numbers follows from (1) and from (2) with $\varphi \in A^0$ taken to be a constant function.

4.3. Theorem. If $F: D \longrightarrow \tilde{D}$ is differentiable, then

$$F^*d = dF^* , \tag{13}$$

that is,

$$F^*(d\tilde{\varphi}) = d(F^*\tilde{\varphi})$$

for any differential form $\tilde{\varphi}$ on $\tilde{D}$.

Remark. Formula (13) might have been written as

$$F^*\tilde{d} = dF^* , \tag{13'}$$

where $\tilde{d}$ denotes the exterior derivative on $\tilde{D}$; but the fact that (13') is true means that this distinction is not really necessary, so (13) is the customary form.

Proof. For $p = 0$, arbitrary $X \in D$, $u \in T_X$, $\tilde{f} \in \tilde{A}^0 = A^0(\tilde{D})$, we have

$$< u, F^*d\tilde{f} > = < F_*u, d\tilde{f} > = F_*u\cdot\tilde{f} = u\cdot F^*\tilde{f} = < u, dF^*\tilde{f} > \tag{14}$$

using, in turn, Proposition 3.5, axiom (3) on $\tilde{D}$, Proposition 3.3, and axiom (3) on D. Since (14) holds for each $u \in T_X$, we have $F^*d\tilde{f} = dF^*\tilde{f}$ at $X \in D$ for each $X \in D$, or

$$(15) \qquad F^*d\tilde{f} = dF^*\tilde{f}\ , \qquad \tilde{f} \in \tilde{A}^0\ ,$$

from which also follows $dF^*d\tilde{f} = 0$ (why?).

If we introduce a coordinate system $(\tilde{x}^1,\ldots,\tilde{x}^m)$ on $\tilde{D}$, and express $\tilde{\varphi} \in \tilde{A}^p$ in the form (7), then by Proposition 3.7 and (15) we have

$$\begin{aligned} dF^*\tilde{\varphi} &= dF^*(\Sigma_{i_1<\ldots<i_p}\ \tilde{\varphi}_{i_1\ldots i_p}\ d\tilde{x}^{i_1} \wedge\ldots\wedge d\tilde{x}^{i_p}) \\ &= d(\Sigma_{i_1<\ldots<i_p}\ (F^*\tilde{\varphi}_{i_1\ldots i_p})(F^*d\tilde{x}^{i_1}) \wedge\ldots\wedge (F^*d\tilde{x}^{i_p})) \\ &= \Sigma_{i_1<\ldots<i_p}\ (F^*d\tilde{\varphi}_{i_1\ldots i_p}) \wedge (F^*d\tilde{x}^{i_1}) \wedge\ldots\wedge (F^*d\tilde{x}^{i_p}) \\ &= F^*(\Sigma_{i_1<\ldots<i_p}\ d\tilde{\varphi}_{i_1\ldots i_p} \wedge d\tilde{x}^{i_1} \wedge\ldots\wedge d\tilde{x}^{i_p}) \\ &= F^*d\tilde{\varphi}\ . \end{aligned}$$

4.4. <u>Definition</u>. A differential form φ is called <u>closed</u> if $\varphi \in \ker d$, that is, if $d\varphi = 0$; φ is called <u>exact</u> if $\varphi \in \operatorname{im} d$, that is, if there exists a differential form η such that $\varphi = d\eta$. The set consisting of the closed differential forms of degree p is a vector space (why?) which will be denoted by $Z^p = Z^p(D)$; the vector space of the exact differential forms of degree p will be denoted by $B^p = B^p(D)$.

<u>Remarks</u>. By (9), an exact form must have positive

degree, that is, must have zero component in A^0 (in particular, $B^0 = \tilde{0}$), and $Z^n = A^n$. From (10), we have

4.5. <u>Proposition</u>. An exact form is closed.

Therefore we have $B^p \subset Z^p$. By Proposition 4.2, Z^0 consists of the functions which are constant on each connected component of D; in particular, if D is connected, $Z^0 = R$; if D consists of two components then $Z^0 = R \oplus R$, etc.

Definition 4.4 can be summarized in the statement that the following are exact sequences:

$$\tilde{0} \longrightarrow Z^p \xrightarrow{\iota} A^p \xrightarrow{d} B^{p+1} \longrightarrow \tilde{0}\,, \qquad p = 0, 1, \ldots, n-1\,,$$

$$\tilde{0} \longrightarrow Z^n \longrightarrow A^n \longrightarrow \tilde{0}\,.$$

The fact that $d^2 = 0$ gives

$$\tilde{0} \longrightarrow B^p \xrightarrow{\iota} Z^p\,, \qquad p = 0, 1, \ldots, n\,,$$

which can be completed to

$$\tilde{0} \longrightarrow B^p \xrightarrow{\iota} Z^p \longrightarrow H^p \longrightarrow \tilde{0}\,, \qquad p = 0, 1, \ldots, n\,,$$

by setting $H^p = H^p(D) = Z^p/B^p$. If it is true that every closed form (of positive degree) is exact, then $B^p = Z^p$ and $H^p = \tilde{0}$ $(p > 0)$ by Proposition II, 11.2. (The same proposition also gives $H^0 = Z^0$ in general, since $B^0 = \tilde{0}$.) However, the truth or falsity of this converse to Proposition 4.5 depends on the choice of D, and the dimension of the vector space $H^p(D)$ represents a measure of the failure of this converse for forms

of degree p, $p > 0$ (see Chapter XII, §6). The remainder of this section will be devoted to showing that the converse does hold for a certain class of domains D.

4.6. Definition (cf. §2.4). An open set $D \subset V$ is star-shaped relative to a point $X_o \in D$ if it has the property that the point $X_o + t(X - X_o) \in D$ for $0 \leqq t \leqq 1$ whenever $X \in D$.

4.7. Definition. An open set $D \subset V$ is called star-like if it is star-shaped relative to some point of D or if it is equivalent (by way of a bidifferentiable homeomorphism to a star-shaped domain in R^n $(n = \dim V)$.

Remarks. In the second case, there is an admissible coordinate system $(x^1,\ldots,x^n)$ on D in terms of which D has the following property: there is a point $X_o \in D$ with coordinates $(x_o^1,\ldots,x_o^n)$, say, such that the point X_t with coordinates $x_o^i + t(x^i - x_o^i)$, $i = 1,\ldots,n$, $0 \leqq t \leqq 1$, is in D whenever the point X with coordinates $(x^1,\ldots,x^n)$ is in D. In general, $X_t \neq X_o + t(X - X_o)$, and the path X_t joining X_o and X is not straight. Such a coordinate system will be called star-like.

Examples. An open convex set (Exercise X, 5.5) is star-shaped relative to any one of its points. Given an arbitrary open $D \subset V$ and $X_o \in D$, there exists an open set $D_o \subset D$ with $X_o \in D_o$ which is star-like; in fact, take any coordinate system $(x^1,\ldots,x^n)$ on D, with $X_o = (x_o^1,\ldots,x_o^n)$, and take

$$D_o = \{X \mid X = (x^1,\ldots,x^n) \in D \text{ and } \Sigma_{i=1}^n (x^i - x_o^i)^2 < \varepsilon\}$$

for a suitable choice of $\varepsilon > 0$.

4.8. Theorem. If $D \subset V$ is star-like, then every closed form on D of positive degree is exact.

The proof of this theorem will be given in 4.13 below.

4.9. Definition. If u is a differentiable q-vector field on D, that is, if $u \in \Lambda^q_\tau = \Lambda^q_\tau(D)$ where $\tau = \tau^\infty$, then

$$i(u): A^p \longrightarrow A^{p-q}, \qquad p = q,\ldots,n,$$

is defined as follows: for $\varphi \in A^p$, the form $i(u)\varphi \in A^{p-q}$ is determined by

$$< v, i(u)\varphi > = < u \wedge v, \varphi > \qquad \text{for } v \in \Lambda^{p-q}_\tau,$$

if $p > q$, with $i(u)\varphi = < u, \varphi > \in A^0$ if $p = q$ and $i(u)\varphi = 0$ if $p < q$. The differential form $i(u)\varphi$ is called the contraction of φ with u.

Remarks. A contraction operator can also be defined at a point $X \in D$ for q-vectors and p-forms at X, or for tensor fields contracted against tensor fields, etc.

It is left as an exercise to show that

$$(16) \quad i(u)(\varphi + \psi) = i(u)\varphi + i(u)\psi, \qquad u \in \Lambda^q_\tau,\ \varphi,\ \psi \in A,$$

$$(17) \quad i(u)(\varphi \wedge \psi) = i(u)\varphi \wedge \psi + (-1)^p \varphi \wedge i(u)\psi, \qquad u \in \tau,\ \varphi \in A^p,\ \psi \in A;$$

thus, for $u \in \tau$, the operator $i(u)$ is an anti-derivation of degree -1. In terms of coordinates, if $u = \Sigma_{j=1}^n u^j \frac{\partial}{\partial x^j}$ and φ is given by (7), then

in R^n by a fixed vector in a bidifferentiable map.] Then $x^i = t\tilde{x}^i$.

Finally, we define a linear transformation

$$k: A^p \longrightarrow A^{p-1}, \qquad p = 1,\ldots,n,$$

as follows: $k\varphi \in A^{p-1}$ is the differential form determined by

$$(24) \quad \begin{aligned} &< u_1 \wedge \ldots \wedge u_{p-1}, k\varphi > \\ &= P_* \int_0^1 < P^* u_1 \wedge \ldots \wedge P^* u_{p-1}, i(\frac{\partial}{\partial t})(F^*\varphi) > dt, \qquad u_\ell \in \tau, \end{aligned}$$

if $p > 1$, and

$$(24') \qquad k\varphi = P_* \int_0^1 i(\frac{\partial}{\partial t})(F^*\varphi)dt$$

if $p = 1$.

<u>Remarks</u>. Formula (24) does indeed define a differential form $k\varphi$ of degree $p - 1$ for each differential form φ of degree $p > 0$ on D. First, $F^*\varphi$ is a differential form of degree p on $\tilde{D}$; then $i(\frac{\partial}{\partial t})(F^*\varphi)$ is a differential form of degree $p - 1$ with the property that its expression in terms of coordinates does not involve dt (why?), although its coefficients still depend on t. After integrating with respect to t, however, we have values which do not depend on t, and therefore give functions, by way of P_*, on D. Note also that the formula is skew-symmetric in the u_ℓ's and linear with respect to scalars which are functions of $(x^1,\ldots,x^n)$, that is, $< fu_1 \wedge \ldots \wedge u_{p-1}, k\varphi > = f < u_1 \wedge \ldots \wedge u_{p-1}, k\varphi >$, etc.

(18) $i(u)\varphi = \Sigma_{i_1<\ldots<i_{p-1}} \Sigma_{j=1}^{n} u^j \varphi_{ji_1\ldots i_{p-1}} dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}$.

[Here, $\Sigma_{j=1}^{n}$ may equally well be written $\Sigma_{j\neq i_s}$, since the omitted terms vanish anyway (why?).] Note also that

$$< u \wedge v, i(u)\varphi > = 0 , \qquad u \in \tau, v \in \Lambda^{p-2}\tau, \varphi \in A^p .$$

4.10. <u>Definition</u>. For $u \in \tau$, the operator

$$\mathcal{L}_u : A^p \longrightarrow A^p , \qquad p = 0, 1, \ldots, n ,$$

is defined by

(19) $$\mathcal{L}_u = i(u)d + di(u) ,$$

that is,

(19') $$\mathcal{L}_u\varphi = i(u)d\varphi + d(i(u)\varphi) , \qquad \varphi \in A .$$

<u>Remarks</u>. The notation $\mathcal{L}_u$ is used for this operator because it happens to coincide with the Lie derivative, with respect to the vector field u, in the case of differential forms. By (1) and (16) we have

$$\mathcal{L}_u(\varphi + \psi) = \mathcal{L}_u\varphi + \mathcal{L}_u\psi ;$$

by (2) and (17) we have

$$\mathcal{L}_u(\varphi \wedge \psi) = \mathcal{L}_u\varphi \wedge \psi + \varphi \wedge \mathcal{L}_u\psi ;$$

and by (10), we have

$$d\,\mathcal{L}_u = \mathcal{L}_u d .$$

In terms of coordinates, if φ is given by (7) and F by (23), then

$$F^*\varphi = \Sigma_{i_1<\ldots<i_p} \varphi_{i_1\ldots i_p}(t\tilde{x}^1,\ldots,t\tilde{x}^n).$$
$$(td\tilde{x}^{i_1} + \tilde{x}^{i_1}dt)\wedge\ldots\wedge(td\tilde{x}^{i_p} + \tilde{x}^{i_p}dt)\ ,$$

and

$$i(\frac{\partial}{\partial t})(F^*\varphi)$$
$$= \Sigma_{i_1<\ldots<i_{p-1}}\Sigma_{j=1}^n t^{p-1}\tilde{x}^j\varphi_{ji_1\ldots i_{p-1}}(t\tilde{x}^1,\ldots,t\tilde{x}^n)d\tilde{x}^{i_1}\wedge\ldots\wedge d\tilde{x}^{i_{p-1}}$$

If

$$k\varphi = \Sigma_{i_1<\ldots<i_{p-1}}(k\varphi)_{i_1\ldots i_{p-1}}dx^{i_1}\wedge\ldots\wedge dx^{i_{p-1}}$$

then

$$(k\varphi)_{i_1\ldots i_{p-1}} = < \frac{\partial}{\partial x^{i_1}}\wedge\ldots\wedge\frac{\partial}{\partial x^{i_{p-1}}},\ k\varphi >$$

$$= P_*\int_0^1 < \frac{\partial}{\partial\tilde{x}^{i_1}}\wedge\ldots\wedge\frac{\partial}{\partial\tilde{x}^{i_{p-1}}},\ i(\frac{\partial}{\partial t})(F^*\varphi) > dt$$

(25)

$$= P_*\int_0^1 \Sigma_{j=1}^n t^{p-1}\tilde{x}^j\varphi_{ji_1\ldots i_{p-1}}(t\tilde{x}^1,\ldots,t\tilde{x}^n)dt$$

$$= \int_0^1 \Sigma_{j=1}^n t^{p-1}x^j\varphi_{ji_1\ldots i_{p-1}}(tx^1,\ldots,tx^n)dt$$

If $p = 1$, then $k\varphi \in A^0$ is given by

(25') $$k\varphi = \int_0^1 \Sigma_{j=1}^n x^j\varphi_j(tx^1,\ldots,tx^n)dt$$

In all cases, the coefficients in the expression for $k\varphi$ are functions which vanish at X_0 which corresponds to $x^j = 0$, $j = 1,\ldots,n$.

The integrals on the right in (24) are actually improper integrals, since the integrand is not defined for $t = 0$ and $t = 1$, but it is clear from (25), for example, that the integrals exist. To avoid improper integrals, we should have to take I to be the closed interval $0 \leq t \leq 1$, and modify all definitions, etc., to cover the fact that $D \times I$ is not an open subset of $V \oplus R$.

4.12. <u>Proposition</u>. The operator k defined by (24) satisfies

$$kd\varphi + dk\varphi = \varphi\,, \qquad \varphi \in A^p,\ p > 0\,, \tag{26}$$

$$kdf = f - f(X_0)\,, \qquad f \in A^0\,. \tag{27}$$

<u>Proof</u>. The identity (26) is equivalent to

$$(26')\ \langle u_1 \wedge \ldots \wedge u_p,\ kd\varphi \rangle + \langle u_1 \wedge \ldots \wedge u_p,\ dk\varphi \rangle = \langle u_1 \wedge \ldots \wedge u_p,\ \varphi \rangle$$

for arbitrary choices of $u_\ell \in \tau$, $\ell = 1,\ldots,p$, and it is sufficient to verify this identity on a basis for $\Lambda^p\tau$, i.e. for $u_\ell = \dfrac{\partial}{\partial x^{i_\ell}}$. We use the identity (19')

$$i(\tilde{u})d\tilde{\psi} + d(i(\tilde{u})\tilde{\psi}) = \mathcal{L}_{\tilde{u}}\tilde{\psi}$$

on $\tilde{D}$, with $\tilde{\psi} = F^*\varphi$ and $\tilde{u} = \dfrac{\partial}{\partial t}$. Then

$$P_*\int_0^1 < P^* \frac{\partial}{\partial x^{i_1}} \wedge \ldots \wedge P^* \frac{\partial}{\partial x^{i_p}}, \; i(\frac{\partial}{\partial t})(dF^*\varphi) > dt$$

$$(28) \qquad + P_* \int_0^1 < P^* \frac{\partial}{\partial x^{i_1}} \wedge \ldots \wedge P^* \frac{\partial}{\partial x^{i_p}}, \; d(i(\frac{\partial}{\partial t})(F^*\varphi)) > dt$$

$$= P_* \int_0^1 < P^* \frac{\partial}{\partial x^{i_1}} \wedge \ldots \wedge P^* \frac{\partial}{\partial x^{i_p}}, \; \mathcal{L}_{\frac{\partial}{\partial t}} F^*\varphi > dt$$

For the first term, we use Theorem 4.3 to give $dF^*\varphi = F^*d\varphi$, and then (24) to identify this term as $< \frac{\partial}{\partial x^{i_1}} \wedge \ldots \wedge \frac{\partial}{\partial x^{i_p}}, \; kd\varphi >$.

For the second, we argue from (11) that, since the contraction is with vectors $P^* \frac{\partial}{\partial x^{i_\ell}}$ "lifted up" from D, the function

$$< P^* \frac{\partial}{\partial x^{i_1}} \wedge \ldots \wedge P^* \frac{\partial}{\partial x^{i_p}}, \; d(i(\frac{\partial}{\partial t})(F^*\varphi)) >$$

involves no differentiation with respect to t, but only differentiations with respect to the $\tilde{x}^k$, which commute with the integration; then we use (21) to identity the second term with $< \frac{\partial}{\partial x^{i_1}} \wedge \ldots \wedge \frac{\partial}{\partial x^{i_p}}, \; dk\varphi >$.

Finally, we use (20) — on $\tilde{D}$ — to evaluate the right-hand side of (28). Note that $\frac{\partial}{\partial t}$ has constant coefficients in terms of the coordinates on $\tilde{D}$. For the case that F is given by (23), we obtain

$$P_*\int_0^1 \frac{\partial}{\partial t}(t^p\varphi_{i_1\ldots i_p}(tx^1,\ldots,tx^n))dt = t^p\varphi_{i_1\ldots i_p}(tx^1,\ldots,tx^n)\Big|_0^1$$

$$= \varphi_{i_1\ldots i_p}(x^1,\ldots,x^n) = \left\langle \frac{\partial}{\partial x^{i_1}} \wedge\ldots\wedge \frac{\partial}{\partial x^{i_p}}, \varphi \right\rangle .$$

In the case of a general F, not given by (23), the same conclusion is found, using the fact F is assumed to satisfy (22').

To prove (27), we use (24'), Proposition 4.3, (19'), and the fact that $i(\frac{\partial}{\partial t})(F^*f) = 0$ to obtain

$$kdf = P_*\int_0^1 i(\frac{\partial}{\partial t})(dF^*\varphi)dt = P_*\int_0^1(\mathcal{L}_{\frac{\partial}{\partial t}}F^*f)dt$$

$$= P_*\int_0^1(\frac{\partial}{\partial t}F^*f)dt = f(x^1,\ldots,x^n) - f(0,\ldots,0) .$$

<u>Remark</u>. Many different operators k can be constructed, which satisfy Proposition 4.12, corresponding to different choices of the differentiable map $F: D \times I \longrightarrow D$. For the map (23), corresponding to a given star-like coordinate system on D, the operator k can be <u>defined</u> by the equations (25) and the properties of Proposition 4.12 verified by direct computation.

4.13. <u>Proof of Theorem 4.8</u>. If $D \subset V$ is star-like, then there is an operator k on D satisfying Proposition 4.12. If φ satisfies $d\varphi = 0$ and is of positive degree, then (26) becomes $dk\varphi = \varphi$, so for $\eta = k\varphi$ we have $\varphi = d\eta$.

<u>Remarks</u>. For $\hat{\eta} = \eta + d\psi$, we also have $\varphi = d\hat{\eta}$; that is, the equation $d\eta = \varphi$, for given φ, does not determine η uniquely. The above construction gives a solution of the following system of differential equations:

$$(29) \qquad \Sigma_{\ell=1}^{p} (-1)^{\ell-1} \frac{\partial \eta_{j_1 \cdots j_{\ell-1} j_{\ell+1} \cdots j_p}}{\partial x^{j_\ell}} = \varphi_{j_1 \cdots j_p},$$

in the unknowns $\eta_{i_1 \cdots i_{p-1}}$ for given functions $\varphi_{j_1 \cdots j_p}$. The condition $d\varphi = 0$ is equivalent to the integrability conditions for the system (29).

4.14. Proposition (Poincaré Lemma). If $D \subset V$ is given, together with $X_o \in D$, then there is an open set D_o, containing X_o and contained in D, such that, for any closed differential form φ of positive degree on D, there is a differential form η, defined on D_o, satisfying $d\eta = \varphi$ in D_o.

§5. Riemannian metric

If V is a finite dimensional vector space with a scalar product (Definition III, 1.1), then for any open $D \subset V$ the scalar product on V serves not only to determine arc length for curves in D (Chapter VI, §3), but also to define a scalar product in each fibre T_X of the tangent space of D, since these fibres are copies of V. Further, as was seen in Chapter IX, the scalar product in V induces a scalar product in V^* and in all the vector spaces constructed as tensor or exterior products of V or V^*, and consequently in each fibre which is part of a structure associated with D.

A scalar product may be considered as an element of $L(V, V; R)$ or as an element of $L(V, V^*)$ or as an element of $L(V \otimes V, R) = (V \otimes V)^*$. Thus a scalar product on V defines a covariant tensor γ of order 2, at each $X \in D$, where

(1) $$< u \otimes v, \gamma > = u \cdot v , \qquad u, v \in T_X .$$

The properties S1, S4, and S5 of scalar product give

(2) $$< u \otimes v, \gamma > = < v \otimes u, \gamma > , \qquad u, v \in T_X ,$$

(3) $$< u \otimes u, \gamma > \geqq 0 , \qquad u \in T_X ,$$

with

(3') $$< u \otimes u, \gamma > = 0 \text{ if and only if } u = \tilde{0} .$$

These additional properties are described by saying that γ is symmetric and positive definite.

If $A_1, \ldots, A_n$ is an orthonormal basis for V, the corresponding linear transformation $V \longrightarrow R^n$ which introduces linear coordinates (Definition 1.6) on V (and therefore on D) is an isometry (Definition III, 6.1), and the linear coordinates are euclidean coordinates in the strict sense of the term. Further, the identification of T_X, $X \in D$, with the set of tangent vectors to R^n, at the point corresponding to X, is also an isometry. For such a coordinate system, the associated contravariant basis elements form an orthonormal basis in each fibre T_X. Thus, if these coordinates are denoted by $(\tilde{x}^1, \ldots, \tilde{x}^n)$, the corresponding expression of γ is

(4) $$\gamma = \Sigma_{k=1}^{n} d\tilde{x}^k \otimes d\tilde{x}^k ,$$

since $\frac{\partial}{\partial \tilde{x}^k} \cdot \frac{\partial}{\partial \tilde{x}^\ell} = \delta_{k\ell}$, that is, $< \frac{\partial}{\partial \tilde{x}^k} \otimes \frac{\partial}{\partial \tilde{x}^\ell}, \gamma > = \delta_{k\ell}$. (In particular, as X varies, it is clear that γ is differentiable).

However, if some other coordinate system $(x^1,\ldots,x^n)$ is used, the basis elements $\frac{\partial}{\partial x^i}$ do not, in general, give an orthonormal basis in each fibre T_X, and

$$\gamma = \Sigma_{i,j=1}^{n} g_{ij}\, dx^i \otimes dx^j \tag{5}$$

where

$$g_{ij} = < \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}, \gamma > = \frac{\partial}{\partial x^i} \cdot \frac{\partial}{\partial x^j} . \tag{6}$$

5.1. Definition. A riemannian metric on D is defined by a differentiable covariant tensor γ of order 2 which is symmetric and positive definite, that is, which has the properties (2), (3), and (3'). γ is then called the fundamental tensor of the riemannian space D, and determines a scalar product in T_X, for each $X \in D$, by taking (1) as the definition of the scalar product in T_X.

5.2. Definition. Let D, $\tilde{D}$ be riemannian spaces with fundamental tensors γ, $\tilde{\gamma}$ respectively. A bidifferentiable map $F: D \longrightarrow \tilde{D}$ is called an isometry if

$$\gamma = F^*\tilde{\gamma} .$$

It is left as an exercise to verify that this condition is equivalent to the statement that $F_*: T_X \longrightarrow \tilde{T}_{F(X)}$ is an isometry for each $X \in D$.

5.3. Definition. A riemannian metric γ on D is called flat (or euclidean) if (D, γ) is isometric to an open subset of R^n.

5.4. Proposition. A riemannian metric γ on D is

flat if and only if there exists an admissible coordinate system $(\tilde{x}^1,\ldots,\tilde{x}^n)$ on D in terms of which γ is given by (4).

Remarks. In general, a riemannian metric is not flat. This causes no difficulty in computation, provided it is remembered that an associated basis need not be an orthonormal basis. On the other hand, one must avoid those aspects of one's geometric intuition which are appropriate only to euclidean or flat spaces. For example, if $F: [0, 1] \longrightarrow D$ is a differentiable curve in D, then the length of F from 0 to t is a function $s = s(t)$ satisfying

$$\frac{ds}{dt} = |F'(t)| = \sqrt{F'(t)\cdot F'(t)} .$$

If F is given in terms of a coordinate system $(x^1,\ldots,x^n)$ by the component functions $x^i = f^i(t)$, then the tangent vector $F'(t)$ at $X = F(t) \in D$ has the expression

$$F'(t) = \Sigma_{i=1}^n \frac{df^i}{dt} \frac{\partial}{\partial x^i} = \Sigma_{i=1}^n \frac{dx^i}{dt} \frac{\partial}{\partial x^i} ,$$

so

$$(7) \qquad \left(\frac{ds}{dt}\right)^2 = \Sigma_{i,j=1}^n g_{ij}(X(t)) \frac{dx^i}{dt} \frac{dx^j}{dt} .$$

(In classical notation, the formula

$$ds^2 = \Sigma_{i,j=1}^n g_{ij}\, dx^i dx^j$$

appears; this is to interpreted as giving a formula of the type (7) for any particular choice of curve.) Formula (7) shows that a straight line in D need not be the shortest distance between two

points. The curves which satisfy the necessary condition to be of minimal length are called <u>geodesics</u>. (The situation is analogous to that found in determining the minimum value of a differentiable function $y = f(x)$ on an interval $a \leqq x \leqq b$. The methods of the calculus determine the points x at which the necessary condition for a relative minimum value of f occurs, but do not immediately exclude points giving a relative maximum or an inflection point; in any case, $f(a)$ or $f(b)$ may give the actual minimum value.) On the surface of a sphere, the arcs of great circles are the geodesics. In map making, a portion of the surface of the sphere must be projected into a plane sheet of paper. In using a map, one must think of the plane as having a riemannian metric rather than a flat metric: in general, the geodesics on the sphere do not give straight lines in the plane; if the projection is specially chosen so that the geodesics do map into straight lines in the plane, then some other aspect of metric geometry, such as area (= 2-dimensional volume) is distorted.

If a scalar product is given in each fibre T_X, then those constructions of Chapter IX which depend on a scalar product in T_X can be carried out at each $X \in D$. It remains only to verify that the results are differentiable as X varies. The computation of these operations in terms of bases is covered in the exercises of Chapter IX, §10; the formulas obtained there can be used in the present situation by taking the basis elements to be basis elements associated with a particular choice of coordinates in D.

For example, the isomorphism $T_X \longrightarrow T_X^*$, which also expresses the scalar product in T_X, may now be given as

$$(8) \qquad u \longrightarrow i(u)\gamma \ , \qquad u \in T_X \ ,$$

if we use the contraction operator of §4 as defined for tensors rather than forms. In terms of coordinates (Exercise IX, 10.15(a)) we have

$$u = \Sigma_{i=1}^n u^i \frac{\partial}{\partial x^i} \longrightarrow \omega = i(u)\gamma = \Sigma_{j=1}^n (\Sigma_{i=1}^n u^i g_{ij}) dx^j \ .$$

From this formula it is clear that this isomorphism, for varying X, sends a differentiable vector field on D into a differentiable differential form of degree 1 on D that is, a riemannian metric gives an isomorphism $\tau \longleftrightarrow A^1$, and also why this operation is sometimes called "lowering the indices" of u.

The scalar product in T_X^* induced from the scalar product on T_X may also be expressed by a (contravariant) tensor of order 2 which is symmetric and positive definite. In terms of a coordinate system $(x^1,\ldots,x^n)$, this tensor is expressed as $\Sigma_{i,j=1}^n g^{ij} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j}$, where $g^{ij} = dx^i \cdot dx^j$; if γ is given by (5), then, by Exercise IX, 10.15(b),

$$\Sigma_{k=1}^n g_{ik} g^{kj} = \delta_i^j \ , \qquad \Sigma_{k=1}^n g^{ik} g_{kj} = \delta_j^i \ .$$

It is left as an exercise to show that the coefficients g^{ij} are differentiable as X varies, and that $\det(g^{ij}) = 1/\det(g_{ij})$.

The scalar product in T_X (or T_X^*) induces a scalar product in $\Lambda^p T_X$ (or in $\Lambda^p T_X^*$) and it is easily verified that

the formulas expressing these scalar products are differentiable as X varies.

In $\Lambda^n T_X$ (or in $\Lambda^n T_X^*$) we have two elements of unit length, and it is necessary to choose one of them before using the $*$-operator of Definition IX, 9.13. The geometric ideas concerned in making the choice will be discussed next.

5.5. Proposition. Let V be a finite dimensional vector space. An equivalence relation (Definition I, 12.4) is defined in the set of bases for V by: $(A_1,\ldots,A_n) \equiv (B_1,\ldots,B_n)$ if and only if $A_1 \wedge \ldots \wedge A_n = \lambda B_1 \wedge \ldots \wedge B_n$ with $\lambda > 0$. The number of equivalence classes so determined is exactly 2.

Proof. If $A_1,\ldots,A_n$ is a basis for V, then $A_1 \wedge \ldots \wedge A_n \in \Lambda^n V$ is not equal to $\tilde{0}$. If $B_1,\ldots,B_n$ is another basis for V then, since $\dim \Lambda^n V = 1$,

$$A_1 \wedge \ldots \wedge A_n = \lambda B_1 \wedge \ldots \wedge B_n$$

with $\lambda \neq 0$; that is, $\lambda > 0$ or $\lambda < 0$. Thus a relation can be defined as above. It is left as an exercise to verify that the axioms given in Definition I, 12.4 are satisfied. For the last statement, we note that if $(A_1,\ldots,A_n) \not\equiv (B_1, B_2, B_3,\ldots,B_n)$, then $(A_1,\ldots,A_n) \equiv (B_2, B_1, B_3,\ldots,B_n)$.

5.6. Definition. An orientation in a finite dimensional vector space V is defined by a choice of one of the two equivalence classes in Proposition 5.5. The bases which lie in the selected class will be called positively oriented.

Remarks. An orientation in V induces an orientation

for any open $D \subset V$, as follows. A linear coordinate system on D will be called positively oriented if it is derived from a positively oriented basis for V. A general coordinate system on D will then be called positively oriented if the Jacobian determinant associated with the change of coordinates (from a positively oriented linear coordinate system) is positive. [This definition is possible because the determinant never vanishes (see Theorem 3.8, Remarks) and is therefore positive at all $X \in D$, or negative at all $X \in D$, if D is connected, or on each component of D otherwise.]

In terms of an arbitrary coordinate system $(x^1,\ldots,x^n)$, the n-vectors (or n-forms) of unit length are (Exercise IX, 10.15)

$$(9) \qquad \pm \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^1} \wedge \ldots \wedge \frac{\partial}{\partial x^n}, \qquad \pm \sqrt{g}\, dx^1 \wedge \ldots \wedge dx^n$$

respectively, where

$$(10) \qquad g = \left(\frac{\partial}{\partial x^1} \wedge \ldots \wedge \frac{\partial}{\partial x^n}\right) \cdot \left(\frac{\partial}{\partial x^1} \wedge \ldots \wedge \frac{\partial}{\partial x^n}\right) > 0$$

is the determinant of the matrix (g_{ij}) of the coefficients of γ in this coordinate system.

5.7. <u>Definition</u>. If an orientation of V is given, we shall select the elements of unit length in $\wedge^n T_X$ and $\wedge^n T_X{}^*$ by choosing the $+$ signs in (9) whenever the coordinate system is positively oriented.

<u>Proof</u>. In any coordinate system, we must clearly make the same choice of $+$ (or of $-$) in (9) for every $X \in D$, if D is connected, since otherwise the choice would be discontinuous

as X varies. To see that the choice is well-defined, we must check that the choice does not vary with a change of coordinates. This follows from the fact that under a change of coordinates we have, by (12) of §3,

$$d\tilde{x}^1 \wedge \ldots \wedge d\tilde{x}^n = \frac{\partial(\tilde{x}^1,\ldots,\tilde{x}^n)}{\partial(x^1,\ldots,x^n)}\, dx^1 \wedge \ldots \wedge dx^n ,$$

where the coefficient on the right is positive if and only if the two coordinate systems have the same orientation. From (10) we find also that

$$\sqrt{g} = \frac{\partial(\tilde{x}^1,\ldots,\tilde{x}^n)}{\partial(x^1,\ldots,x^n)}\sqrt{\tilde{g}} \tag{11}$$

if both coordinate systems have the same orientation.

Conversely, if D is connected, a choice of the n-form of unit length in $\wedge^n T_X{}^*$ for each $X \in D$, so as to give an element of A^n, determines an orientation in D (and in V) by saying that a coordinate system $(x^1,\ldots,x^n)$ is positively oriented if and only if the selected n-form is expressed by $\sqrt{g}\, dx^1 \wedge \ldots \wedge dx^n$ in terms of that coordinate system.

With the choice of n-form in $\wedge^n T_X{}^*$ determined as above, we may define the $*$-operator for forms by Definition IX, 9.13, for each $X \in D$. Then we have

5.8. <u>Theorem</u>. On an oriented riemannian space $D \subset V$, the $*$-operator is an automorphism

$$*: A \longrightarrow A$$

of A with

(12) $$*: A^p \longrightarrow A^{n-p}, \qquad p = 0, 1, \ldots, n,$$

where $n = \dim V$, and

(13) $$**\varphi = (-1)^{np+p}\varphi, \qquad \varphi \in A^p,$$

(14) $$\varphi \wedge *\psi = \psi \wedge *\varphi, \qquad \varphi, \psi \in A^p.$$

Proof. The above statements have all been given, for any fixed $X \in D$, in Chapter IX, and yield the corresponding statements about forms because their expressions in terms of a coordinate system have differentiable coefficients. The fact that $*$ is an endomorphism of A follows from the fact that $*$ is a linear transformation for fixed $X \in D$. Since no differentiation is involved, this statement is true if A is considered as a vector space with the constant functions as scalars, or if A is considered as a module with the elements of A^0 as scalars. The fact that $*$ is an automorphism follows from (13), which is derived from Exercise IX, 10.11. Formula (14) follows from Exercise IX, 10.10(b).

Note that $*$ does not give a homomorphism with respect to the exterior product.

For reference, we repeat the formula of Exercise IX, 10.15(e), for the case of a positively oriented coordinate system $(x^1, \ldots, x^n)$. If $\varphi \in A^p$ is given by

$$\varphi = \Sigma_{i_1 < \ldots < i_p} \varphi_{i_1 \ldots i_p} \, dx^{i_1} \wedge \ldots \wedge dx^{i_p},$$

and $\psi \in A^p$ is given by

$$\psi = \Sigma_{i_1<\ldots<i_p} \psi_{i_1\ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p} ,$$

then $\varphi \wedge *\psi \in A^n$ is given by

(15) $$\varphi \wedge *\psi = \mu \sqrt{g}\, dx^1 \wedge \ldots \wedge dx^n$$

where

$$\mu = \Sigma_{i_1<\ldots<i_p} \Sigma^n_{k_1=1} \cdots \Sigma^n_{k_p=1} \varphi_{i_1\ldots i_p} g^{i_1k_1} \cdots g^{i_pk_p} \psi_{k_1\ldots k_p} .$$

5.9. <u>Definition</u>. The <u>coderivative</u> $\delta\varphi$ of $\varphi \in A^p$ is defined by

$$\delta\varphi = (-1)^{np+n+1} *d*\varphi .$$

The coderivative $\delta\varphi$ of $\varphi \in A$ is defined to be the sum of the coderivatives of the components of φ of degree $p = 0, 1, \ldots, n$.

5.10. <u>Theorem</u>. On an oriented riemannian space $D \subset V$, the coderivative δ of Definition 5.9 is an endomorphism

$$\delta: A \longrightarrow A$$

of the vector space A such that

(16) $$\delta: A^p \longrightarrow A^{p-1} , \qquad p = 1, \ldots, n ,$$

(17) $$\delta: A^0 \longrightarrow \tilde{0} .$$

Further,

(18) $$\delta^2 = 0 ,$$

(19) $$*\delta d = d\delta * ,$$

(20) $$\delta d * = *d\delta .$$

<u>Proof</u>. The linearity of δ (over R) follows from the linearity of $*$ and of d, where A is considered as a vector space with the constant functions as scalars. To prove (16) note that, if $\varphi \in A^p$, then $*\varphi \in A^{n-p}$ by (12), $d*\varphi \in A^{n-p+1}$, and $\delta\varphi = \pm *d*\varphi \in A^{n-(n-p+1)}$. If $\varphi \in A^0$, then $d*\varphi = 0$ which gives (17).

For (18) we note that, if $\varphi \in A^p$, then (13) gives

$$**(d*\varphi) = (-1)^{n(n-p+1)+n-p+1} d*\varphi = (-1)^{(n+1)(n-p+1)} d*\varphi ,$$

so

$$\begin{aligned}\delta^2\varphi &= (-1)^{n(p-1)+n+1} *d*((-1)^{np+n+1} *d*\varphi) \\ &= (-1)^{n(p-1)+n+1}(-1)^{np+n+1}(-1)^{(n+1)(n-p+1)} *d(d*\varphi) = 0\end{aligned}$$

since $d^2 = 0$.

If $\varphi \in A^p$, then $d*d\varphi \in A^{n-p}$, and

$$\begin{aligned}*\delta d\varphi &= *((-1)^{n(p+1)+n+1} *d*)d\varphi \\ &= (-1)^{n(p+1)+n+1}(-1)^{n(n-p)+n-p} d*d\varphi \\ &= (-1)^{p+1} d*d\varphi ,\end{aligned}$$

using $(-1)^{n^2+n} = 1$ and $(-1)^{-p} = (-1)^p$, while

$$\begin{aligned}d\delta*\varphi &= d(-1)^{n(n-p)+n+1} *d*)*\varphi \\ &= (-1)^{n(n-p)+n+1}(-1)^{np+p} d*d\varphi = (-1)^{p+1} d*d\varphi\end{aligned}$$

This result proves the identity (19). The identity (20) is proved similarly and is left as an exercise.

Remarks. Although the properties of δ, as given in Theorem 5.10, are relatively simple, the explicit computation of $\delta\varphi$ in terms of coordinates is not so simple. As an example, we give the following (without proof): if the riemannian metric is flat, and if a coordinate system $(x^1,\ldots,x^n)$ in which $g_{ij} = \delta_{ij}$ is used, then for

$$\varphi = \Sigma_{i_1<\ldots<i_p} \varphi_{i_1\ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p}$$

we have

$$(21) \quad \delta\varphi = -\Sigma_{j_1<\ldots<j_{p-1}} (\Sigma_{k=1}^{n} \frac{\partial\varphi_{kj_1\ldots j_{p-1}}}{\partial x^k}) dx^{j_1} \wedge \ldots \wedge dx^{j_{p-1}} .$$

If the metric is not flat, or if special coordinates are not used even if the metric is flat, the coefficient in (21) is replaced by a more complicated expression involving the coefficients g_{ij} and their derivatives.

5.11. Definition. On an oriented riemannian space $D \subset V$, the Laplace-Beltrami operator Δ is defined by

$$(22) \qquad \Delta = d\delta + \delta d .$$

5.12. Theorem. The operator Δ of Definition 5.11 is an endomorphism

$$\Delta: A \longrightarrow A$$

of the vector space A such that

(23) $$\Delta: A^p \longrightarrow A^p , \qquad p = 0, 1, \ldots, n ,$$

where $n = \dim V$. Further,

(24) $$*\Delta = \Delta * ,$$

(25) $$d\Delta = \Delta d = d\delta d ,$$

(26) $$\delta\Delta = \Delta\delta = \delta d\delta .$$

These statements are easily derived from the properties of d, $*$, and δ.

Remarks. As in the case of δ, the computation of Δ in terms of coordinates involves the coefficients g_{ij} and their derivatives. In the case of a flat metric, with a "good" choice of coordinates, we have

(27) $$\Delta\varphi = -\Sigma_{i_1<\ldots<i_p}\left(\Sigma_{k=1}^{n} \frac{\partial^2\varphi_{i_1\ldots i_p}}{(\partial x^k)^2}\right)dx^{i_1} \wedge \ldots \wedge dx^{i_p} ,$$

and Δ is seen to be the usual laplacian operator, except for sign. In more general cases, the coefficient in (27) is replaced by

$$\Sigma_{j,k=1}^{n} g^{jk} \frac{\partial^2\varphi_{i_1\ldots i_p}}{\partial x^j \partial x^k} + \ldots$$

where the omitted terms involve first order derivatives of the coefficients of φ and the coefficients themselves and first and second order derivatives of the g_{ij}'s.

5.13. Definition. A differential form φ is called harmonic if $\Delta\varphi = 0$.

This definition includes the Definition VIII, 6.2 of a harmonic function in the case of a flat metric. The analogue of a harmonic vector field (Definition VIII, 6.1) would be a form φ which satisfies $d\varphi = 0$ and $\delta\varphi = 0$; such a form will also satisfy $\Delta\varphi = 0$, but the converse is not true in general.

§6. Exercises

1. [As a preliminary exercise, evaluate $\lim_{t \to +\infty} t^r e^{-t}$ for all real numbers r.]

(a) Let a be a real number, and let $h_a: R \longrightarrow R$ be defined by

$$h_a(x) = \begin{cases} \exp\left(-\dfrac{1}{x-a}\right), & x > a, \\ 0, & x \leqq a. \end{cases}$$

Show that h_a is a function of class C^∞.

(b) Let $a < b$ and let $h_{ab}: R \longrightarrow R$ be defined by $h_{ab}(x) = h_a(x) \cdot h_{-b}(-x)$. Show that h_{ab} is of class C^∞ and determine the support (Exercise X, 5.9) of h_{ab}.

(c) Let $a < b$ and define $g_{ab}: R \longrightarrow R$ by

$$g_{ab}(x) = \frac{\int_{-\infty}^{x} h_{ab}(t)dt}{\int_{-\infty}^{\infty} h_{ab}(t)dt}$$

Show that g_{ab} is of class C^∞, and that

(i) $$0 \leqq g_{ab}(x) \leqq 1 \qquad \text{for all } x \in R,$$

(ii) $\qquad g_{ab}(x) = 0 \,, \qquad\qquad x \leqq a \,,$

(iii) $\qquad g_{ab}(x) = 1 \,, \qquad\qquad x \geqq b \,.$

(d) Let $\tilde{a} < a < b < \tilde{b}$. Use the results of (c) to define a function $f\colon R \longrightarrow R$ which is of class C^∞ and satisfies

(i) $\qquad 0 \leqq f(x) \leqq 1 \qquad$ for all $x \in R$,

(ii) $\qquad f(x) = 1 \,, \qquad\qquad a \leqq x \leqq b \,,$

(iii) $\qquad f(x) = 0 \,, \qquad\qquad x \leqq \tilde{a}$ or $x \geqq \tilde{b}$.

(Note that $[a, b] \subset (\tilde{a}, \tilde{b})$ and that the support of f is contained in $[\tilde{a}, \tilde{b}]$.)

(e) Let $\tilde{a}^i < a^i < b^i < \tilde{b}^i$, $i = 1,\ldots,n$. Let $D \subset R^n$ and $\tilde{D} \subset R^n$ be defined by

$$D = \{X \mid X \in R^n \text{ and } a^i < x^i(X) < b^i,\ i = 1,\ldots,n\}$$

and

$$\tilde{D} = \{X \mid X \in R^n \text{ and } \tilde{a}^i < x^i(X) < \tilde{b}^i,\ i = 1,\ldots,n\} \,.$$

(Then $\bar{D} \subset \tilde{D}$.) Define a function $F\colon R^n \longrightarrow R$ of class C^∞ satisfying

(i) $\qquad 0 \leqq F(X) \leqq 1 \qquad$ for all $X \in R^n$,

(ii) $\qquad F(X) = 1 \,, \qquad\qquad X \in \bar{D} \,,$

(iii) $\qquad F(X) = 0 \,, \qquad\qquad X \notin \tilde{D} \,.$

[Hint: If $P^i: R^n \longrightarrow R$ is defined by $(x^1,\ldots,x^n) \longrightarrow x^i$ and $f: R \longrightarrow R$ is of class C^∞, then $f^i = (P^i)^* f: R^n \longrightarrow R$ is of class C^∞. Use a suitable product of such functions.]

(f) If $D_1,\ldots,D_k$, $\tilde{D}_1,\ldots,\tilde{D}_k$ are "rectangular" open sets in R^n, with $\bar{D}_\ell \subset \tilde{D}_\ell$, $\ell = 1,\ldots,k$, let

$$D = \cup_{\ell=1}^k D_\ell \ , \qquad \tilde{D} = \cup_{\ell=1}^k \tilde{D}_\ell \ .$$

Show that $\bar{D} \subset \tilde{D}$ and use the results of (e) to define a function $F: R^n \longrightarrow R$ of class C^∞ and satisfying (i), (ii), (iii) of (e).

(g) Let $D \subset R^n$ and $\tilde{D} \subset R^n$ be open sets, such that $\bar{D}$ is compact (Definition X, 2.13) and contained in $\tilde{D}$. Show that there exists a function $F: R^n \longrightarrow R$ satisfying (i), (ii), (iii) of (e). [Hint: Show that there exists a finite covering of $\bar{D}$ by rectangular open sets D_ℓ, with $\bar{D}_\ell \subset \tilde{D}$, that is, $\bar{D} \subset \cup_{\ell=1}^k D_\ell$, and that it is possible to choose $\tilde{D}_\ell$ with $\bar{D}_\ell \subset \tilde{D}_\ell \subset \tilde{D}$, $\ell = 1,\ldots,k$.]

2. Let $F: R \longrightarrow R$ be defined by $\tilde{x}(x) = x^3$. Show that F is of class C^∞ and bijective, but that F^{-1} fails to be of class C^∞. What hypothesis of Theorem 3.8 is not satisfied?

3. (Spherical coordinates). Let $D \subset R^3$ be the set of all points of R^3 except the points $(x^1, 0, x^3)$, $x^1 \geq 0$, $-\infty < x^3 < \infty$, and write $x = x^1$, $y = x^2$, $z = x^3$ for the coordinates of points of D. In $\tilde{D} \subset R^3$, write $r = \tilde{x}^1$, $\theta = \tilde{x}^2$, $\varphi = \tilde{x}^3$ where $\tilde{D}$ is defined by $0 < r < \infty$, $0 < \theta < \pi$, $0 < \varphi < 2\pi$. Finally, let $F: \tilde{D} \longrightarrow D$ be defined by the component functions

$$x = r \sin\theta \cos\varphi ,$$
$$y = r \sin\theta \sin\varphi ,$$
$$z = r \cos\theta .$$

(a) Show that F is bijective and has positive Jacobian determinant on $\tilde{D}$. (If we write $r = \tilde{x}^1$, $\varphi = \tilde{x}^2$, $\theta = \tilde{x}^3$, then this determinant is always negative.) Compute $F_*\frac{\partial}{\partial r}$, $F_*\frac{\partial}{\partial \theta}$, $F_*\frac{\partial}{\partial \varphi}$, F^*dx, F^*dy, F^*dz, $F^*(dx \wedge dy \wedge dz)$, and $F^*\gamma$, where γ defines the euclidean metric on D.

(b) Show that F can be extended to include the boundary points of D for which $\varphi = 0$, and the boundary points of $\tilde{D}$ given by $(x, 0, z)$ with $x > 0$, so that the extended map is bijective. Discuss the behavior of F at other boundary points of D.

(c) Show that, if the coordinates on $\tilde{D}$ are introduced as coordinates on D, any associated basis in the new coordinate system is orthogonal, but not orthonormal except at certain points $X \in D$.

4. (Cylindrical coordinates). The component functions of F, analogous to those in Exercise 3, are

$$x = r \cos\theta ,$$
$$y = r \sin\theta ,$$
$$z = z .$$

Give suitable choices of $D \subset R^3$ and $\tilde{D} \subset R^3$ so that $F: \tilde{D} \longrightarrow D$ is bijective, where D and $\tilde{D}$ are open sets, and work out the

properties of F, following the general pattern of Exercise 3.

5. Let $D \subset V$ be open, and let $(x^1, \ldots, x^n)$ be a system of coordinates on D.

(a) If $u, v \in \tau(D)$, work out the formula for $[u, v]$ (see §1.10) in terms of coordinates.

(b) If $u, v \in \tau(D)$ and $\omega \in A^1(D)$, show that

$$< u \wedge v, d\omega > = u \cdot < v, \omega > - v \cdot < u, \omega > - < [u, v], \omega >$$

by computing both sides of this formula in terms of coordinates.

6. Let ω be a differential form of degree 1. A non-vanishing function f is called an <u>integrating factor</u> of ω if $f\omega$ is closed. Show that the existence of an integrating factor of ω implies $\omega \wedge d\omega = 0$.

7. Let ω be a differential form of degree 1 on R^n. Verify the formula

$$kd\omega + dk\omega = \omega$$

of Proposition 4.12 by computing in terms of the euclidean coordinates, when the operator k is expressed in terms of coordinates by (25) of §4.

8. Let $D \subset V$ be an oriented riemannian space, $\dim V = 3$, and let (x^1, x^2, x^3) be a positively oriented coordinate system on D. Compute $*1$, $*dx^i$, $i = 1, 2, 3$, $*(dx^\ell \wedge dx^m)$, $\ell < m$, and $*(dx^1 \wedge dx^2 \wedge dx^3)$.

If $\omega = \Sigma_{i=1}^3 \omega_i dx^i$, show that

$$\delta\omega = - \Sigma_{i,\ k=1}^{3} \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i} (\sqrt{g}\, g^{ik}\, \omega_k) .$$

Does this type formula hold if $\dim V > 3$? Use the above result to give a formula for Δf in terms of coordinates, if f is a function D.

Work out the computation for $\delta\varphi$ if φ is a differential form of degree 2, and use this result to compute $\Delta\omega$.

9. Verify that

$$\frac{\partial}{\partial x^i} \log g = \Sigma_{\ell,m=1}^{3} g^{\ell m} \frac{\partial g_{\ell m}}{\partial x^i} , \qquad i = 1, 2, 3 ,$$

and use this result to eliminate the expression $\sqrt{g}$ from the formulas obtained in Exercise 8.

10. For suitable $D \subset R^3$, taken with the euclidean metric, give the formulas obtained in Exercise 8 for the special case of spherical coordinates (Exercise 3), and the special case of cylindrical coordinates (Exercise 4).

11. On an oriented riemannian space $D \subset V$, we have the isomorphism $\tau \longleftrightarrow A^1$, that is, $v \longleftrightarrow \omega$, obtained by raising or lowering indices. For a function f on D, define ∇f by the condition $\nabla f \longleftrightarrow df$. For a vector field $v \longleftrightarrow \omega$, define $\operatorname{div} v$ by $\operatorname{div} v = - \delta\omega = *d*\omega$ and, if $\dim V = 3$, $\operatorname{curl} v$ by $\operatorname{curl} v \longleftrightarrow *d\omega$. Show that the above definitions are generalizations of those given in Theorem VII, 3.2, Definition VIII, 4.1 and Definition VIII, 4.4, by considering the special case that the riemannian metric on D is derived from a choice of scalar product on V.

XII. INTEGRAL CALCULUS OF FORMS

§1. Introduction

Throughout this chapter, V will denote an n-dimensional vector space over the reals and D will denote an open set in V.

The integral of a differential form of degree q over a finite singular differentiable chain in D of dimension q, $0 \leq q \leq n$, will be defined and the fundamental theorem in the integral calculus of differential forms, namely Stokes' theorem, will be proved. Some homological applications of Stokes' theorem will be pointed out. Finally a riemannian metric and an orientation will be assumed given on V and, for domains D of suitably restricted type, Green's formulas for differential forms will be derived from Stokes' theorem.

These formulas contain most of the multiple integral formulas common in calculus.

§2. Standard simplexes

Denote by $t^1, t^2, \ldots, t^q$ the orthogonal coordinates of euclidean space R^q.

2.1. Definition. The standard q-simplex s_q $(q > 0)$ is the closed set of points in R^q whose coordinates satisfy the inequalities

$$t^1 \geqq 0,\ t^2 \geqq 0,\ \ldots,\ t^q \geqq 0$$

and

$$t^1 + t^2 + \ldots + t^q \leq 1 \quad .$$

The <u>vertices</u> of s_q are the $q + 1$ points $E_o, E_1, \ldots, E_q$, where E_o is the origin of R^q and E_i, $i = 1, \ldots, q$, is the point whose i^{th} coordinate is equal to 1 and whose remaining coordinates are equal to zero. Thus, s_q is the convex hull (Exercise X, 5.6) of the points $E_o, E_1, \ldots, E_q$. The ith <u>face</u> of s_q, denoted by $F^i s_q$, is the face opposite the vertex E_i, that is, the convex hull of the points $E_o, \ldots, E_{i-1}, E_{i+1}, \ldots, E_q$. The <u>standard 0-simplex</u> is the single point E_o.

We obtain the so-called "barycentric" coordinates on s_q $(q > 0)$ by introducing an additional coordinate

$$t^o = 1 - \Sigma_{i=1}^q t^i \quad .$$

The points of s_q then have the barycentric coordinates $(t^o, t^1, \ldots, t^q)$ where the t^i, $i = 0, 1, \ldots, q$, are non-negative and satisfy

$$t^o + t^1 + \ldots + t^q = 1 \quad .$$

The point $(t^o, t^1, \ldots, t^q)$ may be regarded as the center of mass of $q + 1$ particles, of masses $t^o, t^1, \ldots, t^q$ respectively, situated at the vertices $E_o, E_1, \ldots, E_q$. The ith face $F^i s_q$ is the closed subset of s_q composed of the points whose i^{th} barycentric coordinate t^i vanishes.

We next define what is meant by a differentiable map or function on s_q $(q > 0)$. Previous definitions concerning the notion

of differentiability have covered only the case of maps defined on an open set of a vector space (whereas s_q is a compact subset of R^q). Thus, if $F: s_q \longrightarrow W$, all earlier definitions apply only to the function $\tilde{F}$ obtained by restricting F to the interior of s_q . However, the notion of a continuous function (class C^0) on s_q has already been defined (Definition X, 1.22).

2.2. Definition. A function $F: s_q \longrightarrow W$, where W is a vector space, is said to be of class C^1 if $\tilde{F}$ is of class C^1 and if, for each $Y \in R^q$,

$$\lim_{X \longrightarrow A} \tilde{F}'(X, Y) \tag{1}$$

exists, for each A in the topological boundary bs_q of s_q, and the function $\mathcal{D}_Y F: s_q \longrightarrow W$, obtained by defining $\mathcal{D}_Y F(A) = F'(A, Y)$ to be the limit in (1), is continuous on s_q.

Remarks. For certain directions Y (which point "into" s_q at A), the difference quotient

$$\frac{F(A + hY) - F(A)}{h} \tag{2}$$

will be defined for sufficiently small positive values of h. If this difference quotient has a limit as h tends to zero, and if $F'(A, Y)$ exists as defined above, these two values will coincide since, by the mean value theorem, we have

$$\frac{F(A + hY) - F(A)}{h} = \tilde{F}'(A + \theta hY, Y) \tag{3}$$

for some θ (depending on h) with $0 < \theta < 1$. Note that the

existence of the limit in (1) is a stronger assumption than the existence of a limit in (3), since (3) refers only to points $X = A + \theta hY$. The difference quotient (2) will also be defined for certain directions Y which point "along" the boundary of s_q. If F is of class C^1, in the sense of Definition 2.2, it can be shown that these directions also give limits compatible with Definition 2.2. Finally, it can be shown that a map which is of class C^1 in the sense of Definition 2.2 is also of class C^0.

2.3. Definition (cf. Definition XI, 1.1). The map or function $F: s_q \longrightarrow W$ is of class C^k, $k = 1, 2, \ldots$, if for any set of vectors $Y_1, \ldots, Y_k$ in V the function $\mathfrak{D}_{Y_1} \mathfrak{D}_{Y_2} \ldots \mathfrak{D}_{Y_k} F: s_q \longrightarrow W$ is defined (in the sense of Definition 2.2) and continuous. F is of class C^∞ if it is of class C^k for all non-negative integers k.

We again have

2.4. Proposition (cf. Proposition XI, 1.2). A function $F: s_q \longrightarrow W$ is of class C^k, $k \geq 1$, if and only if $\mathfrak{D}_Y F$ is of class C^{k-1} for each $Y \in V$. A function which of class C^k is of class C^ℓ for all non-negative $\ell \leq k$.

2.5. Proposition (cf. Proposition XI, 1.5). If W is finite dimensional, then $F: s_q \longrightarrow W$ is of class C^k if and only if the component functions of F, with respect to any basis for W, are of class C^k.

With $F'(X, Y)$ defined as in Definition 2.2, it is clear that the usual properties are retained, such as the fact that

$F'(X, Y)$ is linear in Y (since the limit of a sum is the sum of the limits). However, the full strength of Definition 2.3 is shown by the following theorem (the proof of which will be omitted).

2.6. Theorem. A function $f: s_q \longrightarrow R$ is of class C^k, $k = 0, 1, \ldots, \infty$, if and only if there exists a function $g: D \longrightarrow R$ which is of class C^k, where $D \subset R^q$ is an open set containing s_q, such that the restriction of g to s_q coincides with f.

2.7. Definition. A function g having the property described in Theorem 2.6 will be called an extension of f, and f will be called extendable.

Remark. If f is extendable, an extension of f is not uniquely determined, and the choice of g or of D can be varied. For example, if $s_q \subset \tilde{D} \subset D$, where $\tilde{D}$ is open, then the restriction of g to $\tilde{D}$ is also an extension of f.

2.8. Proposition. A function $F: s_q \longrightarrow W$, where W is finite dimensional, is of class C^k, $k = 0, 1, \ldots, \infty$, if and only if there exists a function $G: D \longrightarrow W$ which is of class C^k, where $D \subset R^q$ is an open set containing s_q, such that the restriction of G to s_q coincides with F.

Proof. To show the existence of G, take a basis for W and apply Theorem 2.6 to the component functions f^i of F. Then take D to be the intersection of the finite number of open sets D_i, where f^i has an extension g^i on D_i, and let G be defined by the component functions g^i, restricted to D.

Conversely, if F is extendable, and the extension G

of F is of class C^k, then the component functions of G relative to a basis for W are of class C^k and give extensions of the component functions of F, which are therefore of class C^k.

Thus, for example, a vector field on s_q of class C^∞ is a section u of $s_q \times R^q$ over s_q which is the restriction to s_q of some function $v: D \longrightarrow D \times R^q \in \tau(D)$, where D is an open set containing s_q.

Since a map $F: s_q \longrightarrow W$ of class C^∞ is extendable if W is finite dimensional, a function F^* from differential forms of class C^∞ on W to differential forms of class C^∞ on s_q can be defined by restricting the function G^* induced by an extension G of F. Forms $F^*\varphi$ on s_q are thus extendable by construction, and therefore of class C^∞. It is left as an exercise to verify that F^* is well-defined, that is, independent of the choice of G, by showing that an explicit formula for $F^*\varphi$ involves only values uniquely determined by F. Thus, G^* serves only to guarantee the existence and behavior of F^*.

§3. Singular differentiable chains; singular homology

For simplicity, we shall confine ourselves to the case C^∞ (see, however, the remark following Definition 3.9). Consequently, we shall suppose, throughout the remainder of this chapter, that all maps and differential forms, etc. are differentiable of class C^∞ and that "differentiable" means "differentiable of class C^∞".

Let D be an open set of the n-dimensional vector space V.

3.1. Definition. A singular (differentiable) q-simplex σ_q of D is a pair (s_q, F) where s_q is the standard q-simplex and $F: s_q \longrightarrow D$ is a differentiable map.

We remark again that the map F may be extended (non-uniquely, of course) to a differentiable map of some neighborhood in R^q of s_q into D. Two q-simplexes $\sigma_q = (s_q, F)$ and $\sigma'_q = (s_q, F')$ are the same if and only if F and F' coincide.

3.2. Definition. A singular (differentiable) q-chain c_q of D is a function from all singular (differentiable) q-simplexes σ_q of D to the real numbers which is zero except on a finite number of the σ_q.

3.3. Definition. The support of a singular (differentiable) q-simplex $\sigma_q = (s_q, F)$, denoted supp σ_q, is the compact set $F(s_q) \subset D$. The support of a q-chain, denoted supp c_q, is the union of the supports of all the simplexes σ_q on which the function from simplexes to the reals does not vanish.

We note that the support of a chain c_q, the union of finitely many compact sets, is always compact.

3.4. Definition. The support of a differential form φ on D, denoted supp φ, is the smallest closed subset of D outside of which the differential form vanishes identically. A differential form is said to have compact support if its support is a compact subset of D.

The support of a form, in contrast to that of a chain, is not generally compact.

Given an arbitrary compact subset K of D we recall

(Exercise XI, 6.1) that there exists a differentiable function with compact support (relative to D) which is identically equal to 1 on K. Multiplying an arbitrary differential form on D by such a function, we obtain a differential form, with compact support, which coincides with the given differential form on K.

The use of the same term "support" in these two cases is based on the fact that, in computations, a chain or a differential form gives no contribution at points outside its support.

The q-chains on D form an infinite dimensional real vector space $C_q = C_q(D)$, $q = 0, 1, \ldots$. Thus we obtain a graded vector space $C = \{C_0, C_1, \ldots\}$, with C_q as the entry in dimension q.

As in Chapter IX, §1, we denote by σ_q the q-chain which has the value one on σ_q and zero on all other q-simplexes. Then a chain c_q is expressed as a finite sum

$$c_q = \Sigma\, a_i \sigma_q^{(i)} \tag{1}$$

where the real number a_i is the value of the function c_q on $\sigma_q^{(i)}$. Given two chains c_q, c_q', we can arrange, by adding to each a finite number of terms with zero coefficients, that they are of the form

$$c_q = \Sigma\, a_i \sigma_q^{(i)}, \qquad c_q' = \Sigma\, a_i' \sigma_q^{(i)}, \tag{2}$$

where the same simplexes $\sigma_q^{(i)}$ occur in each. Then their sum $c_q + c_q'$ is expressed by

$$c_q + c_q' = \Sigma(a_i + a_i')\sigma_q^{(i)} ,$$

and the product of a q-chain (1) by a scalar $a \in R$ is expressed by

$$ac_q = \Sigma(aa_i)\sigma_q^{(i)} .$$

In the case $D = R^q$, there is a distinguished q-simplex defined by $\sigma_q = (s_q, F)$ where F is the identity map. In this case we drop F and write simply $\sigma_q = s_q$. Corresponding to each face $F^j s_q$, if $q > 0$, there is a distinguished (q-1)-simplex s_{q-1}^j, with $\operatorname{supp} s_{q-1}^j = F^j s_q$, defined by the pair (s_{q-1}, F_q^j) where the injective linear map $F_q^j: s_{q-1} \longrightarrow R^q$ is defined as follows. If $q > 1$, let $\tau^1, \ldots, \tau^{q-1}$ denote the orthogonal coordinates of the euclidean space R^{q-1}. Then F_q^j, $j = 1, \ldots, q$, maps a point $(\tau^1, \ldots, \tau^{q-1})$ of s_{q-1} into the point in R^q whose coordinates are

$$(3) \qquad t^1 = \tau^1, \ldots, t^{j-1} = \tau^{j-1}, t^j = 0, t^{j+1} = \tau^j, \ldots, t^q = \tau^{q-1} ;$$

F_q^o is defined in the same way by

$$(4) \qquad t^1 = 1 - \tau^1 - \ldots - \tau^{q-1}, t^2 = \tau^1, \ldots, t^q = \tau^{q-1} .$$

If $q = 1$, then $s_{q-1} = s_o$ and the faces of $s_q = s_1$ are points, so $F_1^o: s_o \longrightarrow F^o s_1$, $F_1^1: s_o \longrightarrow F^1 s_1$.

3.5. <u>Definition</u>. The <u>boundary</u> ∂s_q <u>of the standard q-simplex</u> s_q, $q > 0$, is the (q-1)-chain in R^q defined by

$$\partial s_q = \Sigma_{j=o}^q (-1)^j s_{q-1}^j .$$

Remark. Clearly, supp ∂s_q coincides with the topological boundary bs_q of s_q, which is the union of the faces of s_q. However, in ∂s_q these faces are interpreted as (the supports of) $(q-1)$-simplexes in R^q and assigned an "orientation".

3.6. Definition. The boundary $\partial\sigma_q$ of a singular (differentiable) q-simplex $\sigma_q = (s_q, F)$ of D, $q > 0$, is the $(q-1)$-chain

$$\partial\sigma_q = \Sigma_{j=0}^{q}(-1)^j\sigma_{q-1}^j ,$$

where $\sigma_{q-1}^j = (s_{q-1}, FF_q^j)$. The boundary of a q-chain $c_q = \Sigma\, a_i\sigma_q^{(i)}$, $q > 0$, is the $(q-1)$-chain

$$\partial c_q = \Sigma\, a_i\partial\sigma_q^{(i)} .$$

Remark. In particular cases, the formula for ∂c_q may need to be rewritten, by combining terms, in order to give an expression in terms of distinct $(q-1)$-simplexes.

3.7. Proposition. The boundary operator ∂ is an endomorphism

$$\partial: \quad C \longrightarrow C$$

of dimension -1 of the graded vector space $C = \{C_0, C_1, \ldots\}$, that is,

$$\partial: C_q \longrightarrow C_{q-1} , \qquad q > 1 ,$$

and is linear. Further,

(4) $$\partial^2 = 0 \quad .$$

The identity (4) follows from $\partial^2\sigma_q$ for any q-simplex σ_q. This is a simple calculation which is left as an exercise.

3.8. Definition. A chain c_q is called a cycle if $q = 0$ or if $c_q \in \ker \partial$, that is, if $\partial c_q = 0$; c_q is called a boundary if $c_q \in \operatorname{im} \partial$, that is, if there exists a $(q+1)$-chain c_{q+1} satisfying $\partial c_{q+1} = c_q$.

The q-cycles constitute a linear subspace Z_q of C_q; the q-boundaries constitute a linear subspace B_q of Z_q.

3.9. Definition. The (finite or infinite dimensional) vector space $H_q = Z_q/B_q$ is called the q-dimensional singular homology of D, or simply the q-homology of D. Its elements are called homology classes and two cycles in the same homology class are said to be homologous, i.e. they differ by a boundary. The graded vector space $H_* = \{H_0, H_1, \ldots\}$ is called the singular homology of D.

Remarks. Simplexes $\sigma_q = (s_q, F)$ with $F: s_q \longrightarrow D$ merely continuous (class C^0) are called, in algebraic topology, singular simplexes, and the corresponding chains, singular chains. It is clear that the definitions given above can equally well be stated for the continuous case. The resulting homology, called the singular homology, is isomorphic, by a classical theorem of topology, to the homology defined above in terms of differentiable maps. For this reason, the qualification "differentiable" has been omitted in Definition 3.9.

3.10. Geometric interpretation. A geometric picture of the above notions is obtained by considering supp σ_q as the "picture" of a q-simplex $\sigma_q = (s_q, F)$. A positive coefficient is interpreted as a constant weighting factor and is not expressed in the picture. For $q > 0$, a coefficient -1 is pictured as reversing the orientation induced by the assigned orientation in the standard q-simplex (Fig. 1). Thus,

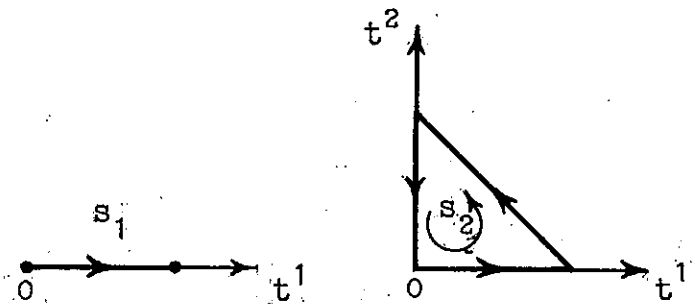


Fig. 1

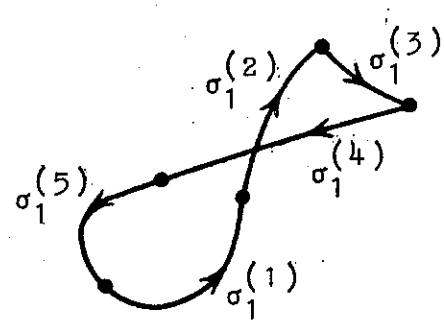


Fig. 2

a 0-simplex gives a point of D, a 0-chain with all coefficients $+1$ gives a finite set of points of D. A 1-simplex of D is a smooth curve in D. A 1-chain $c_1 = \Sigma\, \sigma_1^{(i)}$ of D is a finite collection of such curves. It is a 1-cycle (Fig. 2) if the initial point of $\sigma_1^{(i)}$ coincides with the terminal point of $\sigma_1^{(i-1)}$, $i > 1$, and if the terminal point of the last 1-simplex of c_1 coincides with the initial point of $\sigma_1^{(1)}$. (The total is then a closed piecewise differentiable curve in D, with an assigned sense of direction.) In Fig. 3, the orientations shown are those of the 2-simplexes $(-1)^j s_2^j$, the oriented faces of s_3. Note that where two faces meet, the induced orientations are opposite; this is the

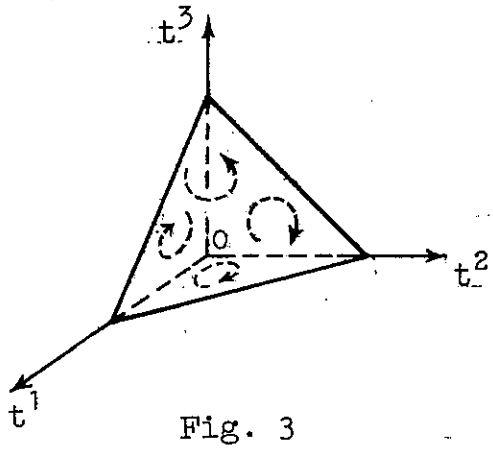


Fig. 3

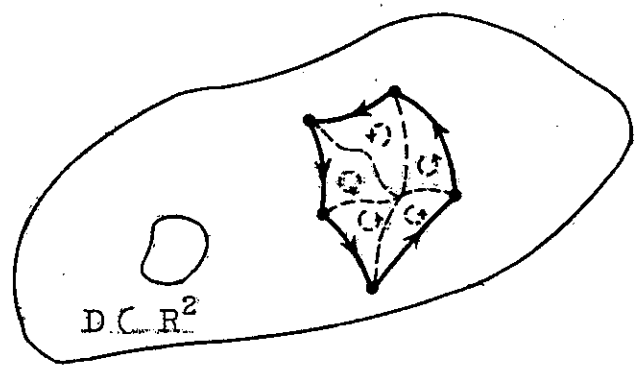


Fig. 4

geometric expression of the identity $\partial^2 = 0$.

Two 0-simplexes (points) of D are homologous if they can be joined by a piecewise differentiable curve in D. A q-cycle is "homologous to zero" if it lies in the trivial homology class of H_q, that is, if it is the boundary of a (q+1)-chain of D (Fig. 4).

The geometric picture must be used with caution whenever supports coincide. For example, the support of a 1-simplex which gives a constant curve in D cannot be distinguished from the support of a 0-simplex; two q-simplexes $\sigma_q = (s_q, F)$ and $\sigma'_q = (s_q, F')$ may have the same supports and the same orientation (that is, the same geometric picture) but they are not equal unless $F = F'$.

3.11. <u>Conical operators</u>. The fact that the boundary operator ∂ on chains and the exterior derivative d are dual operators will be considered in detail in §4 and §6. Here we mention another aspect of this duality.

In Chapter XI, §4, we considered the special choices of $D \subset V$ for which there exists a differentiable map $S: D \times I \longrightarrow D$,

where I is the unit interval, such that $S(X, 1) = X$, $S(X, 0) = X_0$ where X_0 is a fixed point of D. That is, D is differentiably contractible to a point. Corresponding to such a map S, we constructed an operator

$$k: A^p(D) \longrightarrow A^{p-1}(D) , \qquad p = 1, 2, \ldots,$$

such that

$$(5) \qquad kd\varphi + dk\varphi = \varphi , \qquad \varphi \in A^p, \ p > 0 ,$$

$$(6) \qquad kdf = f - f(X_0) , \qquad f \in A^0 ,$$

with the corollary that, for such a D, every closed form of positive degree is exact, and every closed function is a constant.

For the same D and S we can construct an operator

$$K: C_q(D) \longrightarrow C_{q+1}(D) , \qquad q = 0, 1, \ldots ,$$

such that

$$(7) \qquad K\partial c_q + \partial K c_q = c_q , \qquad c_q \in C_q(D), \ q > 0 ,$$

$$(8) \qquad \partial K\sigma_0 = \sigma_0 - (X_0) , \qquad \sigma_0 \in C_0(D) ,$$

where (X_0) denotes the 0-simplex of D whose support is X_0. As a corollary we have that, for such a D, every cycle of positive dimension is a boundary, and every 0-simplex is homologous to (X_0).

We make the definitions for the case of an arbitrary q-simplex $\sigma_q = (s_q, F)$ in D. The operator K is then extended to an arbitrary chain $c_q = \Sigma\, a_i \sigma_q^{(i)}$ by defining $Kc_q = \Sigma\, a_i K\sigma_q^{(i)}$. First we define a $(q+1)$-simplex $\tilde{\sigma}_{q+1} = (s_{q+1}, \tilde{F})$ on $D \times I$ as follows. Let $t^0, t^1, \ldots, t^{q+1}$ be the barycentric coordinates on

s_{q+1}. If $q > 0$, define

$$\tilde{F}(t^o, t^1, \ldots, t^{q+1}) = \begin{cases} (F(\frac{t^1}{1-t^o}, \ldots, \frac{t^{q+1}}{1-t^o}), 1-t^o), & t_o < 1, \\ (F(0, \ldots, 0), 0), & t_o = 1. \end{cases}$$

Then supp $\tilde{\sigma}_{q+1}$ is a "cone" in $D \times I$ with vertex $(F(0, \ldots, 0), 0)$ — the image of E_o of s_{q+1} — and base the points $(X, 1)$, $X \in$ supp σ_q — the image of $F^o s_{q+1}$. (Note that supp $\tilde{\sigma}_{q+1}$ is an ordinary cone if the given F is an affine transformation.) If $q = 0$, we take

$$\tilde{F}(t^o, t^1) = (\text{supp } \sigma_o, 1 - t_o) .$$

Then we define $K\sigma_q$ by

$$K\sigma_q = (s_{q+1}, S\tilde{F})^{\bullet} .$$

The computation of the identities (7) for the case $c_q = \sigma_q$ is left as an exercise. The identity (8) is obvious.

§4. Integrals of forms over chains

We begin by defining the integral of a q-form over the standard q-simplex s_q, $q > 0$. Let E denote an arbitrary point of R^q, with coordinates $(t^1, \ldots, t^q)$. (We reserve the usual notation X for points in $D(V)$. A (differentiable) q-form α on s_q is of the form

$$\alpha = \alpha_{12\ldots q}(E)dt^1 \wedge \ldots \wedge dt^q$$

where $\alpha_{12\ldots q}(E) = \alpha_{12\ldots q}(t^1, \ldots, t^q)$ is a (differentiable) function on s_q. Note that $dt^1 \wedge \ldots \wedge dt^q$ is the q-form of "length" $+1$ on R^q, relative to which the fixed coordinate system on R^q is positively oriented.

4.1. <u>Definition</u>. We define

$$(1) \qquad \int_{s_q} \alpha = \int_{s_q} \alpha_{12\ldots q}(E) dt^1 dt^2 \ldots dt^q .$$

The integral on the right of (1) can be evaluated as an iterated integral, for example, as

$$\int_{t^1=0}^{1} \int_{t^2=0}^{1-t^1} \ldots \int_{t^q=0}^{1-t^1-\ldots-t^{q-1}} \alpha_{12\ldots q}(t^1, \ldots, t^q) dt^1 \ldots dt^q ,$$

whenever $\alpha_{12\ldots q}$ is a continuous function on s_q.

4.2. <u>Definition</u>. Let $\sigma_q = (s_q, F)$ be a singular differentiable q-simplex on D, where D is an open set in a finite dimensional vector space V, and let φ be a differential form on D of degree q. If $q > 0$, we define

$$(2) \qquad \int_{\sigma_q} \varphi = \int_{s_q} F^* \varphi .$$

If $q = 0$, we define

$$\int_{\sigma_0} \varphi = F^* \varphi(E_0) = \varphi(F(E_0))$$

where $s_0 = E_0$.

In terms of a coordinate system $(x^1, \ldots, x^n)$ on D, we have

$$\varphi = \Sigma_{i_1<\ldots<i_q} \varphi_{i_1\ldots i_q}(X)dx^{i_1}\wedge\ldots\wedge dx^{i_q}$$

and

$$F^*\varphi = \Sigma_{i_1<\ldots<i_q} \varphi_{i_1\ldots i_q}(F(E))\frac{\partial(f^{i_1},\ldots,f^{i_q})}{\partial(t^1,\ldots,t^q)}\, dt^1\wedge dt^2\wedge\ldots\wedge dt^q$$

Here $(f^1(E), \ldots, f^n(E))$ are the coordinates of the point $X = F(E) \in D$, and

$$\frac{\partial(f^{i_1},\ldots,f^{i_q})}{\partial(t^1,\ldots,t^q)}$$

is the Jacobian determinant.

<u>Remark</u>. If $D = R^q$, then $\int_{\text{supp } \sigma_q} \varphi$ can be given a meaning, the analogue of (1). If F is injective with non-vanishing Jacobian determinant, then it is clear, from the above expression for $F^*\varphi$, that (2) is the usual formula for transformation of variables, by means of F, in evaluating the integral. In general, however, the left side of (2) has no meaning except that assigned by (2), and no assumption is made about F other than that it be differentiable.

4.3. <u>Definition</u>. The integral of a differential form φ of degree q over a singular differentiable q-chain $c_q = \Sigma\, a_i \sigma_q^{(i)}$ is defined by

$$\int_{c_q} \varphi = \Sigma_i\, a_i \int_{\sigma_q^{(i)}} \varphi \quad . \tag{3}$$

We remark that (3) is defined if φ is a q-form which is given only on a neighborhood in D of supp c_q. However, we suppose (for simplicity) that all forms are defined on the whole of D.

We are now able to state the fundamental theorem in the integral calculus of differential forms, namely:

4.4. Theorem (Stokes' formula). Let c_q be a singular differentiable q-chain on D, $q > 0$, φ a differential form of degree $q - 1$ on D. Then

$$\int_{c_q} d\varphi = \int_{\partial c_q} \varphi .$$

Proof. It is clearly sufficient to prove Stokes' formula for the case when $c_q = \sigma_q = (s_q, F)$ is a differentiable q-simplex of D. We divide the proof into two parts.

(i) Stokes' formula for the standard q-simplex s_q. Given a differentiable (q-1)-form β on s_q, we prove

$$\int_{s_q} d\beta = \int_{\partial s_q} \beta \tag{4}$$

where $\partial s_q = \Sigma_{j=0}^{q}(-1)^j s_{q-1}^j$. The case $q = 1$ of (4) is a familiar formula from elementary calculus. In this case, β is a function of one variable and we have

$$\int_{s_1} d\beta = \int_0^1 \frac{\partial\beta}{\partial t^1} dt^1 = \beta(1) - \beta(0) = \int_{s_1^0} \beta - \int_{s_1^1} \beta = \int_{\partial s_1} \beta .$$

For $q \geqq 2$, we may write

$$\beta = \Sigma_{i=1}^{q} \beta_{1\ldots(i-1)(i+1)\ldots q} dt^1 \wedge \ldots \wedge dt^{i-1} \wedge dt^{i+1} \wedge \ldots \wedge dt^q = \Sigma_{i=1}^{q} \beta_{(i)} .$$

Then

$$d\beta = \Sigma_{i=1}^{q} (-1)^{i-1} \frac{\partial\beta_{1\ldots(i-1)(i+1)\ldots q}}{\partial t^i} dt^1 \wedge \ldots \wedge dt^q = \Sigma_{i=1}^{q} d\beta_{(i)} .$$

Since integration is a linear operation, (4) follows from the

formulas

$$(5) \qquad \int_{s_q} d\beta_{(i)} = \int_{s^o_{q-1}} \beta_{(i)} + (-1)^i \int_{s^i_{q-1}} \beta_{(i)}$$

and

$$(6) \qquad 0 = (-1)^j \int_{s^j_{q-1}} \beta_{(i)}, \qquad j \neq i,\ j = 1,\ldots,q .$$

To prove (5), we evaluate the left side of (5) as an iterated integral, integrating first with respect to t^i, from $t^i = 0$ to $t^i = 1 - t^1 - \ldots - t^{i-1} - t^{i+1} - \ldots - t^q$. After the first integration, the new integrand is

$$(7) \qquad \begin{aligned} &(-1)^{i-1}\beta_{1\ldots(i-1)(i+1)\ldots q}(t^1,\ldots,t^{i-1}, \zeta, t^{i+1},\ldots,t^q) \\ &+ (-1)^i\beta_{1\ldots(i-1)(i+1)\ldots q}(t^1,\ldots,t^{i-1}, 0, t^{i+1},\ldots,t^q), \\ &\qquad \zeta = 1 - t^1-\ldots-t^{i-1} - t^{i+1}-\ldots-t^q, \end{aligned}$$

still to be integrated with respect to $t^1,\ldots,t^{i-1}, t^{i+1},\ldots,t^q$ (over $F^i s_q$). To evaluate the integral resulting from the second term of (7), take (3) of §3, $j = i$, as a change of variables. The resulting integral is then seen to coincide with the integral $(-1)^i \int_{s_{q-1}} (F^i_q)^* \beta_{(i)} = (-1)^i \int_{s^i_{q-1}} \beta_{(i)}$. To evaluate the integral resulting from the first term of (7), take (4) of §3 as a change of variables. In this case the Jacobian determinant of the change of variables is $(-1)^{i-1}$, and the resulting integral is seen to coincide with $\int_{s_{q-1}} (F^o_q)^* \beta_{(i)} = \int_{s^o_{q-1}} \beta_{(i)}$.

To prove (6), it is sufficient to verify that $(F_q^j)^* \beta_{(i)} = 0$, $j \neq i$, $j = 1, \ldots, q$.

(ii) Stokes' formula for an arbitrary singular differentiable q-simplex $\sigma_q = (s_q, F)$ on D. Let φ be a differential form of degree q on D. Then by Theorem XI, 4.3, and part (i):

$$\int_{\sigma_q} d\varphi = \int_{s_q} F^*(d\varphi) = \int_{s_q} d(F^*\varphi) = \int_{\partial s_q} F^*\varphi = \int_{\partial \sigma_q} \varphi ,$$

q.e.d.

§5. Exercises

1. Verify that $\partial^2 = 0$ by direct calculation.

2. Verify the identity (7) of §3 for the case $c_q = \sigma_q$, $q > 0$.

3. Justify the use of Theorem XI, 4.3 in proving Theorem 4.4.

4. Show, by means of Stokes' formula, that $d^2 = 0$ implies $\partial^2 = 0$ and conversely.

5. Take $D = R^n$ and examine Stokes' formula in the following special cases: (a) $n = 2$, $q = 2$, $\sigma_2 = s_2$; (b) $n = 3$, $q = 2$, $\sigma_2 = (s_2, F)$, $F: s_2 \longrightarrow R^3$. Relate each of these cases to classical formulas of calculus.

6. For $D \subset V$ of the type considered in 3.11, assume the result that a differential form of degree q is uniquely determined by the values $\int_{c_q} \varphi$, $c_q \in C_q(D)$. Corresponding to an operator $K: C_q(D) \longrightarrow C_{q+1}(D)$ satisfying (7) and (8) of §3,

define an operator $k: A^q(D) \longrightarrow A^{q-1}(D)$, $q > 0$, as follows: for $\varphi \in A^q$, $k\varphi$ is the $(q-1)$-form determined by $\int_{c_{q-1}} k\varphi = \int_{Kc_{q-1}} \varphi$. Use Stokes' formula to show that k satisfies (5) and (6) of §3.

§6. Cohomology; de Rham theorem

Let D be an open set of V and denote (as in Chapter XI) by $A^q = A^q(D)$ the space of q-forms on D. Let Z^q be the subspace of closed forms (cocycles), B^q the subspace of Z^q composed of exact forms (coboundaries).

6.1. Definition. The vector space $H^q = Z^q/B^q$ is called the q-dimensional (de Rham) cohomology of D, or simply the q-cohomology of D. Its elements are called cohomology classes and two cocycles (closed forms) in the same cohomology class are said to be cohomologous, i.e. they differ by a coboundary (exact form). The graded vector space $H^* = \{H^0, H^1, \ldots\}$ is called the (de Rham) cohomology of D.

Remark. $H^q = \tilde{0}$ for $q > \dim V$, since then the only differential form in A^q is 0.

We drop the subscript q on chains c_q and assume that the degree of a form φ on $D \subset V$ is equal to the dimension of the chain over which it is being integrated.

6.2. Definition. The integral of a closed form φ $(d\varphi = 0)$ over a cycle c $(\partial c = 0)$ is called the period of φ over c.

6.3. Proposition. The period of a closed form on a cycle depends only on the cohomology class of the form and on the homology

class of the cycle.

Proof. Suppose that c and c' are homologous cycles, i.e. $c - c' = \partial c''$, and let φ be a closed form. Then, by Stokes' formula,

$$\int_c \varphi - \int_{c'} \varphi = \int_{\partial c''} \varphi = \int_{c''} d\varphi = 0 .$$

Dually, if φ and φ' are cohomologous closed forms, i.e. $\varphi - \varphi' = d\psi$, and if c is a cycle, then

$$\int_c \varphi - \int_c \varphi' = \int_c d\psi = \int_{\partial c} \psi = 0 .$$

6.4. Proposition. The integration of forms over chains induces a linear transformation

$$(1) \qquad H_q \otimes H^q \longrightarrow R ,$$

$q = 0, 1, \ldots$.

Proof. To a pair consisting of a class in H_q and a class in H^q, assign the real number obtained by integrating a closed form in the given cohomology class over a cycle in the given homology class. By Proposition 6.3, the resulting function is well-defined, that is, independent of the particular choice of representative of either class. It is left as an exercise to verify that this function is in $L(H_q, H^q; R)$ and therefore, by Theorem IX, 2.4, induces a function in $L(H_q \otimes H^q, R)$.

6.5. Theorem (de Rham). The linear transformation $H_q \otimes H^q \longrightarrow R$ of Proposition 6.4 establishes H^q as the dual of H_q.

6.6. <u>Corollary</u>. $H_q = \tilde{0}$ for $q > \dim V$.

Theorem 6.5 is a deep theorem, and the proof will be omitted! Clearly, for a fixed choice of a class in H^q, (1) gives an element of $L(H_q, R) = (H_q)^*$. It is necessary to show that every $T \in L(H_q, R)$ can be obtained in this way, and that two distinct classes in H^q cannot induce the same element of $L(H_q, R)$. If H_q is finite dimensional, then it has a basis and an arbitrary $T \in L(H_q, R)$ is determined by its values on this basis. If T is induced by a class in H^q, then any closed differential form in this class must have these values as periods over cycles representing the basis. The proof (in this special case) then consists in showing that there exists a closed form having these values as periods, and that any two closed forms having these periods must be cohomologous.

§7. <u>Exercises</u>

1. Consider R^2 with coordinates x, y and take D to be the annulus $0 < 1/4 < x^2 + y^2 < 4$. Construct a 1-cycle c which is not homologous to zero and whose support coincides with the circle $x^2 + y^2 = 1$. Find a closed 1-form φ on D whose period on c is equal to 1. Construct a 1-cycle c' whose support is the circle $x^2 + y^2 = 1$ and which is homologous to zero.

2. Consider R^3 with coordinates x, y, z. From the open torus obtained by rotating around the z-axis the open disk $(x-1)^2 + z^2 < \frac{1}{4}$ in the plane $y = 0$, remove the closed torus obtained by rotating about the z-axis the closed disk

$(x-1)^2 + z^2 \leqq \frac{1}{16}$ in the plane $y = 0$, and take D to be the resulting difference domain. Find two 1-cycles c and c' which represent a basis for H_1. (Give geometric arguments to show that c and c' are not homologous to each other or to zero, and that any 1-cycle of D is of the form $ac + a'c' + c''$, where $a, a' \in R$ and c'' is homologous to zero.) Find a closed 1-form φ on D which has prescribed periods $r, r' \in R$ on c and c', respectively. Describe the closed 1-forms on D which are cohomologous to zero.

3. Construct a domain $D \subset R^3$ for which $\dim H_1 = \infty$.

§8. Green's formulas

In Definition 2.1, a $(q-1)$-dimensional face of the standard simplex s_q in R^q was defined to be the convex hull of a set of points consisting of q of the vertices of s_q. More generally, an r-dimensional face of s_q, $0 \leqq r \leqq q-1$, is the convex hull of a set of points consisting of $r + 1$ vertices of s_q. In this context, a vertex of s_q is a 0-dimensional face of s_q.

V will denote a finite dimensional vector space in which a fixed orientation has been chosen.

8.1. Definition. A differentiable n-chain c of V, $n = \dim V$, will be called regular if it satisfies the following conditions:

(i) $c = \sigma_n^{(1)} + \sigma_n^{(2)} + \ldots$, i.e. all non-zero coefficients in the chain are equal to 1.

(ii) For each n-simplex $\sigma_n = (s_n, F)$ of c,

$F: s_n \longrightarrow F(s_n) \subset V$ is a bidifferentiable map whose Jacobian determinant, relative to any positively oriented coordinate system on V, is positive at each point of s_n.

(iii) The intersection of the supports of two (distinct) n-simplexes of c is either empty or coincides with the support of a q-dimensional face of each, where $0 \leqq q \leqq n - 1$, and the intersection of the supports of two (distinct) (n-1)-simplexes of ∂c is either empty or coincides with the support of a q-dimensional face of each, where $0 \leqq q \leqq n - 2$.

(iv) Whenever the intersection of the supports of two (distinct) n-simplexes of c coincides with the support of an (n-1)-dimensional face of each, the corresponding (n-1)-simplexes which occur in computing ∂c cancel exactly.

8.2. <u>Definition</u>. An open set $D \subset V$ will be called a finite domain with (piecewise differentiable) boundary, or simply a <u>finite domain</u>, if there exists a regular n-chain c of V whose support coincides with the closure of D and whose boundary ∂c has support coinciding with the boundary bD of D.

We remark that the word "domain" usually means a <u>connected</u> open set; however we do not require that a finite domain be connected. Definition 8.2 implies, however, that a finite domain can have at most a finite number of connected components, and that its closure is compact.

8.3. <u>Definition</u>. Let D be a finite domain, $D \subset V$. A regular n-chain of V satisfying the conditions of Definition 8.2 with respect to D will be called an <u>associated</u> (regular)

n-chain.

Remark. A finite domain D has many different associated regular n-chains. We shall be interested only in properties of D which are independent of the choice of the regular n-chain associated with D. For example, the identification of bD with $\operatorname{supp} \partial c$, where c is a regular n-chain associated with D, induces an orientation of bD. The conditions of Definition 8.1 ensure that the orientation is the same for all choices of c.

8.4. Theorem. If D is a finite domain, then H_q (and therefore H^q) is finite dimensional.

The proof will be omitted.

8.5. Definition. If D is a finite domain, we denote by $\mathfrak{a}^p(D)$ the linear subspace of $A^p(D)$ consisting of those $\varphi \in A^p(D)$ which satisfy the following condition: for each n-simplex (s_n, F) of a regular n-chain c associated with D, $F^*\varphi$ is of class C^∞ on s_n.

Remarks. A form $\varphi \in \mathfrak{a}^p(D)$ is thus differentiable in D, and also in the closure of D except possibly at certain boundary points. An exceptional boundary point must lie in the supports of at least two distinct n-simplexes of a regular n-chain associated with D and, in fact, in the support of a q-dimensional face of each, where $0 \leq q \leq n - 2$. Further, any coefficient of φ or derivative of a coefficient must have a limit at such a point from within each n-simplex, so any discontinuity must be a simple "jump" discontinuity.

A condition analogous to that of Definition 8.5 can be

used to define what is meant by a form of class C^k, $k < \infty$, on a finite domain.

For $\varphi \in \mathcal{A}^n(D)$, $\psi \in \mathcal{A}^{n-1}(D)$, the integrals

$$(1) \qquad \int_c \varphi , \qquad \int_{\partial c} \psi$$

are defined for any regular n-chain c associated with D.

8.6. Proposition. Let D be a finite domain, and let $\varphi \in \mathcal{A}^n(D)$, $\psi \in \mathcal{A}^{n-1}(D)$. Then the integrals (1) are independent of the choice of the associated n-chain c.

In view of Proposition 8.6 we state

8.7. Definition. If D is a finite domain, we define

$$\int_D \varphi = \int_c \varphi , \qquad \int_{bD} \psi = \int_{\partial c} \psi ,$$

where c is an associated n-chain, $\varphi \in \mathcal{A}^n(D)$, $\psi \in \mathcal{A}^{n-1}(D)$.

Proposition 8.6 is an expression of the basic facts of integral calculus, that the multiple integral is the sum of the integrals over subdomains of the domain of integration and that the evaluation of the integral is independent of the choice of the parameters of integration so long as all parameter changes are bidifferentiable transformations. The proof is left as an exercise.

Assume now that a riemannian metric is given on V in addition to an orientation. If V is R^n, we shall always suppose that the metric is the euclidean one. Then (Theorem XI, 5.8) we have the isomorphisms

$$*: A^p \longrightarrow A^{n-p}, \qquad 0 \leq p \leq n .$$

It is easily verified that

$$*: \quad \mathcal{A}^p(D) \longrightarrow \mathcal{A}^{n-p}(D) ,$$

and that this correspondence is also an isomorphism, for each $p = 0, 1, \ldots, n$.

8.8. Definition. Let D be a finite domain. Given $\varphi, \psi \in \mathcal{A}^p(D)$ we define the scalar product (φ, ψ) of φ and ψ to be

$$(2) \qquad (\varphi, \psi) = \int_D \varphi \wedge * \psi ,$$

and we write $\|\varphi\| = \sqrt{(\varphi, \varphi)}$ (norm of φ).

8.9. Proposition. The formula (2) does indeed define a scalar product in $\mathcal{A}^p(D)$; that is, (φ, ψ) is a symmetric bilinear function over the real numbers R, and $\|\varphi\| \geq 0$, with $\|\varphi\| = 0$ if and only if $\varphi = 0$. Moreover, $(*\varphi, *\psi) = (\varphi, \psi)$.

The proof is left as an easy exercise.

8.10. Theorem (Green's formulas). Let D be a finite domain, $D \subset V$, on which is given a riemannian metric. Then, for $\varphi \in \mathcal{A}^{p-1}(D)$, $\psi \in \mathcal{A}^p(D)$,

$$\text{(i)} \qquad (d\varphi, \psi) - (\varphi, \delta\psi) = \int_{bD} \varphi \wedge *\psi ;$$

and, for $\varphi, \psi \in \mathcal{A}^p(D)$,

$$\text{(ii)} \qquad (d\delta\varphi, \psi) - (\delta\varphi, \delta\psi) = \int_{bD} \delta\varphi \wedge *\psi ,$$

(iii) $$(d\varphi, d\psi) - (\varphi, \delta d\psi) = \int_{bD} \varphi \wedge *d\psi ,$$

(iv) $$(\Delta\varphi, \psi) - (d\varphi, d\psi) - (\delta\varphi, \delta\psi) = \int_{bD} (\delta\varphi \wedge *\psi - \psi \wedge *d\varphi) ,$$

(v) $$(\Delta\varphi, \psi) - (\varphi, \Delta\psi) = \int_{bD} (\delta\varphi \wedge *\psi - \delta\psi \wedge *\varphi + \varphi \wedge *d\psi - \psi \wedge *d\varphi) ,$$

where Δ is the laplacian for differential forms: $\Delta = d\delta + \delta d$ (Definition XI, 5.11).

<u>Proof</u>. Let $\varphi \in \mathcal{A}^{p-1}(D)$, $\psi \in \mathcal{A}^{p}(D)$. Then $\varphi \wedge *\psi \in \mathcal{A}^{n-1}(D)$, where $\dim V = n$, and we have

(3) $$d(\varphi \wedge *\psi) = d\varphi \wedge *\psi + (-1)^{p-1}\varphi \wedge d(*\psi) = d\varphi \wedge *\psi - \varphi \wedge *\delta\psi .$$

In fact, by (Definition XI, 5.9),

$$\delta\psi = (-1)^{np+n+1} *d*\psi ;$$

then, since $\delta\psi \in \mathcal{A}^{p-1}(D)$, and $d*\psi \in \mathcal{A}^{n-p+1}(D)$,

$$*\delta\psi = (-1)^{np+n+1} **d*\psi = (-1)^{np+n+1}(-1)^{n(n-p+1)+n-p+1} d*\psi = (-1)^{p} d*\psi .$$

Applying Stokes' formula (Theorem 4.4) to $\varphi \wedge *\psi$ and using formula (3), we obtain

$$(d\varphi, \psi) - (\varphi, \delta\psi) = \int_{D} d(\varphi \wedge *\psi) = \int_{bD} \varphi \wedge *\psi$$

which is formula (i) of the theorem. This is the fundamental formula from which the remaining four formulas are immediately derived.

In fact, replacing φ in (i) by $\delta\varphi$, $\varphi \in \mathcal{A}^{p}(D)$, we obtain (ii). Similarly, replacing ψ in (i) by $d\psi$, $\psi \in \mathcal{A}^{p-1}(D)$,

we obtain (iii) (with p replaced by $p - 1$). By subtraction of (iii) with φ and ψ interchanged, from (ii), formula (iv) follows. Finally subtracting from (iv) the formula obtained from it by interchanging φ with ψ, we obtain (v).

Remark. Definitions 8.8 and 8.9, and the formulas of Theorem 8.10 remain valid under much weaker assumptions on the differential forms involved.

8.11. Corollary. If either φ or ψ has a compact support relative to D, then

$$(d\varphi, \psi) = (\varphi, \delta\psi) ,$$

that is, δ is the adjoint of d in the sense of Definition V, 5.1, and

$$(\Delta\varphi, \psi) = (\varphi, \Delta\psi) ,$$

that is, Δ is self-adjoint. Moreover,

$$(\Delta\varphi, \psi) = (d\varphi, d\psi) + (\delta\varphi, \delta\psi) . \tag{4}$$

Proof. If $\varphi \in A^p(D)$ has compact support relative to D, then φ is automatically in $\mathcal{A}^p(D)$ and, moreover, has zero boundary values. Thus the integrals over bD, in (i), (v), and (iv) in Theorem 8.10, vanish.

Remark. It has been noted (Chapter XI, §5) that the laplacian operator Δ used in theoretical work differs in sign from the "ordinary" laplacian. One reason for this choice is that, for $\varphi = \psi$, we then have

$$(4') \qquad (\Delta\varphi, \varphi) = (d\varphi, d\varphi) + (\delta\varphi, \delta\varphi) \geqq 0 ,$$

that is, Δ is a "positive" operator.

The boundary integrals occurring in the formulas of Theorem 8.10 are evaluated on an $(n-1)$-chain ∂c, where c is a regular n-chain associated with D. Thus, if $\chi \in \mathfrak{a}^{n-1}(D)$, then $\int_{\partial c}\chi$ is expressed as a sum of integrals of the form

$$\pm \int_{s_{n-1}} (F F^j_q)^* \chi = \pm \int_{s_{n-1}} \mu d\tau^1 \ldots d\tau^{n-1}$$

where

$$(5) \qquad \begin{aligned} \mu &= \left\langle \frac{\partial}{\partial\tau^1} \wedge \ldots \wedge \frac{\partial}{\partial\tau^{n-1}}, (F F^j_q)^* \chi \right\rangle \\ &= \left\langle (F F^j_q)_* \left(\frac{\partial}{\partial\tau^1} \wedge \ldots \wedge \frac{\partial}{\partial\tau^{n-1}}\right), \chi \right\rangle . \end{aligned}$$

If $X \in bD$ is not an exceptional point, that is, if c can be chosen so that X is an interior point of some $(n-1)$-simplex σ_{n-1} in ∂c, then the vectors $(F F^j_q)_* \frac{\partial}{\partial\tau^\ell}$, $\ell = 1, \ldots, n-1$, span a linear subspace U of T_X, $\dim U = n-1$, where T_X denotes the tangent space to V at X. Here U does not depend on the choice of σ_{n-1}, and may be called the tangent space to bD at X.

Using the scalar product in T_X determined by the riemannian metric, we have

$$T_X = U \oplus U^\perp$$

where $\dim U^\perp = 1$, and $U^\perp$ is the normal line to bD at X. By Theorem IX, 7.6 and Lemma IX, 7.10, we have, for $p > 0$,

$$\wedge^p T_X = \wedge^p U \oplus \wedge^{p-1} U \otimes U^\perp$$

and, by duality,

(6) $$\Lambda^p T_X^* = \Lambda^p U^* \oplus \Lambda^{p-1} U^* \otimes (U^\perp)^* .$$

We write this decomposition as

$$\varphi = t\varphi + n\varphi , \qquad \varphi \in \Lambda^p T_X^* .$$

It is easily verified that

$$*: U \longrightarrow U^\perp , \qquad *: U^\perp \longrightarrow U ,$$

etc., from which we obtain

(7) $$*n = t* , \qquad n* = *t ,$$

using (14) of Chapter IX, §9 as the definition of $*$ on forms. Moreover, it is clear that

(8) $$t(\varphi \wedge \psi) = t\varphi \wedge t\psi .$$

By (5), only $t\chi$ can contribute to $\int_{\partial c} \chi$, $\chi \in \alpha^{n-1}(D)$. However, if $\chi = \varphi \wedge *d\psi$, for example,

$$t\chi = t\varphi \wedge t*d\psi = t\varphi \wedge *nd\psi ,$$

etc. With $tf = f$, $nf = 0$, all statements include the case $p = 0$.

A full discussion of the above decomposition would require showing that the decomposition varies "differentiably" as X varies on bD (so long as X stays away from exceptional boundary points), including the fact that the decomposition can be differentiably extended (locally) into D in a neighborhood of each "good" boundary point.

§9. Potential theory on euclidean domains

Let $V = R^n$, $n \geq 2$, with the euclidean metric, and let D be an open set of R^n. If we compute in terms of the euclidean coordinates $(x^1,\ldots,x^n)$ of R^n, then, for any $\varphi \in A^p(D)$, with

$$(1) \qquad \varphi = \Sigma_{i_1<\ldots<i_p} \varphi_{i_1\ldots i_p}\, dx^{i_1} \wedge \ldots \wedge dx^{i_p} ,$$

the formula (27) of Chapter XI, §5, may be expressed as

$$(2) \qquad \Delta\varphi = \Sigma_{i_1<\ldots<i_p} (\Delta\varphi)_{i_1\ldots i_p}\, dx^{i_1} \wedge \ldots \wedge dx^{i_p}$$

where

$$(3) \qquad (\Delta\varphi)_{i_1\ldots i_p} = \Delta\varphi_{i_1\ldots i_p} = -\Sigma_{k=1}^{n} \frac{\partial^2\varphi_{i_1\ldots i_p}}{(\partial x^k)^2} .$$

Thus, the form φ is harmonic $(\Delta\varphi = 0)$ if and only if each coefficient of φ in the expression (1) is a harmonic function. Thus, the theory of harmonic forms, in this case, could be expressed entirely in terms of harmonic functions.

In the case of a general riemannian metric, however, the formula expressing $(\Delta\varphi)_{i_1\ldots i_p}$ involves the other components of φ as well, etc., and this simplification is not possible. We shall develop the theory for the euclidean case by methods which have analogues in the more general case.

9.1. Definition. Let W be a vector space, V a finite dimensional vector space, and D an open set of V. A differentiable section of $D \times \{W \otimes \Lambda^p V^*\}$ will be called a differential form of degree p with values in W or, shortly, a W-valued p-form.

9.2. <u>Definition</u>. Let V_1, V_2 be finite dimensional vector spaces, and D_1, D_2 open sets in V_1, V_2, respectively. A differentiable section of $\{D_1 \times D_2\} \times \{\Lambda^p V_1^* \otimes \Lambda^q V_2^*\}$ will be called a <u>double form</u> of type (p, q) on $D_1 \times D_2$.

A double form of type (p, q) may be regarded either as a p-form on D_1 with values in (the vector space of) q-forms over D_2, or as a q-form on D_2 with values in the p-forms over D_1.

Let $X_1 \in D_1$, $X_2 \in D_2$ have the coordinates $X_1 = (x_1^1, \ldots, x_1^m)$, $X_2 = (x_2^1, \ldots, x_2^n)$ respectively. Then a double form $\gamma = \gamma(X_1, X_2)$ can be expressed as

$$\gamma(X_1, X_2) = \sum_{\substack{i_1 < \ldots < i_p \\ j_1 < \ldots < j_q}} \gamma_{i_1 \ldots i_p, j_1 \ldots j_q}(X_1, X_2)(dx_1^{i_1} \wedge \ldots \wedge dx_1^{i_p}) \circledcirc (dx_2^{j_1} \wedge \ldots \wedge dx_2^{j_q}) .$$

The symbol $\circledcirc$, indicating a "symmetric" product, reminds that the exterior product of double forms is to be computed by the rule

$$(dx_1^i \circledcirc dx_2^j) \wedge (dx_1^\ell \circledcirc dx_2^s) = (dx_1^i \wedge dx_1^\ell) \circledcirc (dx_2^j \wedge dx_2^s) .$$

Thus if γ is a double form of type (p, q), and γ' a double form of type (p', q'), then

$$\gamma \wedge \gamma' = (-1)^{pp'+qq'} \gamma' \wedge \gamma . \tag{4}$$

The case where $V_1 = V_2 = V$, $D_1 = D_2 = D$, and $p = q$ is of particular importance. Let $X, Y \in D$ be variable points with coordinates $(x^1, \ldots, x^n)$, $(y^1, \ldots, y^n)$ respectively. A

double form $\gamma = \gamma(X, Y)$ of type (p, p) on $D \times D$ will be called simply a <u>double form of degree p on D</u>; it has the form

$$\gamma(X, Y) = \Sigma_{\substack{i_1<\ldots<i_p \\ j_1<\ldots<j_p}} \gamma_{i_1\ldots i_p, j_1\ldots j_p}(X, Y)(dx^{i_1} \wedge \ldots \wedge dx^{i_p}) \circledcirc (dy^{j_1} \wedge \ldots \wedge dy^{j_p}) .$$

9.3. <u>Definition</u>. A double form $\gamma(X, Y)$ on D will be said to be <u>symmetric</u> if $\gamma(X, Y) = \gamma(Y, X)$.

9.4. <u>Definition</u>. For the remainder of this section, $\gamma = \gamma(X, Y)$ will denote the double form on R^n defined, for $X \neq Y$, as follows. Let

$$(5) \qquad f(X, Y) = \begin{cases} \dfrac{1}{(n-2)v_{n-1}} \dfrac{1}{|X - Y|^{n-2}} , & n > 2 , \\ \dfrac{1}{2\pi} \log \dfrac{1}{|X - Y|} , & n = 2 , \end{cases}$$

where the constant v_{n-1} denotes the volume of the unit $(n-1)$-sphere in R^n (cf. Exercise 11.1). Then

$$\gamma(X, Y) = \Sigma_{p=1}^{n} \gamma^p(X, Y) ,$$

where

$$(6) \quad \gamma^p(X,Y) = \Sigma_{i_1<\ldots<i_p} f(X,Y)(dx^{i_1} \wedge \ldots \wedge dx^{i_p}) \circledcirc (dy^{i_1} \wedge \ldots \wedge dy^{i_p}) .$$

<u>Remarks</u>. Thus $\gamma^p(X, Y)$ is a symmetric double form of degree p on R^n which is singular along the diagonal of $R^n \times R^n$. The double form γ is a so-called "fundamental solution" of the Laplace equation $\Delta\gamma = 0$ (typical singularity at $X = Y$, and (9), $X \neq Y$). Fundamental solutions are known to exist for the

laplacian Δ defined by an arbitrary riemannian metric and for much more general elliptic differential operators.

We shall denote by $d_X\gamma^p(X, Y)$, $\delta_X\gamma^p(X, Y)$ and $\Delta_X\gamma^p(X, Y)$ the double forms on $R^n \times R^n$, of types $(p+1, p)$, $(p-1, p)$ and (p, p) respectively, which are obtained from $\gamma^p(X, Y)$ by regarding it as a differential form $\lambda^p(X)$, $X \in D$, with values in the differential forms of degree p on D, and operating on it by d, δ and Δ respectively.

9.5. <u>Lemma</u>. We have the formulas,

$$d_X\gamma^p(X, Y) = \delta_Y\gamma^{p+1}(X, Y) , \tag{7}$$

$$\delta_X\gamma^{p+1}(X, Y) = d_Y\gamma^p(X, Y) , \tag{8}$$

$$\Delta_X\gamma^p(X, Y) = \Delta_Y\gamma^p(X, Y) = 0 , \qquad X \neq Y . \tag{9}$$

<u>Proof</u>. Since $\gamma(X, Y) = \gamma(Y, X)$ is symmetric, the formula (8) is obtained from (7) by interchanging X and Y. Formula (9) is a consequence of (7) and (8), and (3) together with the fact, easily verified, that $\Delta_X f(X, Y) = 0$, $X \neq Y$. We shall give the proof of (7) for the case $n > 2$, leaving the case $n = 2$ as an exercise. We have

$$d_X f(X, Y) = -\frac{1}{v_{n-1}} \frac{1}{|X - Y|^n} \Sigma_{i=1}^n (x^i - y^i)dx^i ,$$

$$d_X\gamma^p(X, Y) =$$
$$- (n-2)\frac{f(X,Y)}{|X-Y|^2}\Sigma_{\ell=1}^{p+1}(-1)^{\ell-1}(x^{i_\ell}-y^{i_\ell})\Sigma_{i_1<\ldots<i_{p+1}}(dx^{i_1}\wedge\ldots\wedge dx^{i_{p+1}}) \circledcirc$$
$$(dy^{i_1}\wedge\ldots\wedge dy^{i_{\ell-1}}\wedge dy^{i_{\ell+1}}\wedge\ldots\wedge dy^{i_{p+1}}) .$$

Formula (7) follows from the fact that we find exactly the same expression for $\delta_Y \gamma^{p+1}(X, Y)$. To compute $\delta_Y \gamma^{p+1}(X, Y)$, we use the formula (21) of Chapter XI, §5 which states that, if φ is given by (1), then

$$(10) \qquad \delta\varphi = \Sigma_{i_1<\ldots<i_{p-1}} (\delta\varphi)_{i_1\ldots i_{p-1}} dx^{i_1} \wedge \ldots \wedge dx^{i_{p-1}}$$

where

$$(11) \qquad (\delta\varphi)_{i_1\ldots i_{p-1}} = - \Sigma_{\ell=1}^{n} \frac{\partial\varphi_{\ell i_1\ldots i_{p-1}}}{\partial x^\ell} .$$

We suppose henceforth that $D \subset R^n$ is a finite domain.

9.6. <u>Definition</u>. For $\varphi \in \mathcal{A}^p(D)$, we define $(\Gamma\varphi)(Y)$ by

$$(12) \qquad \begin{aligned} (\Gamma\varphi)(Y) &= (\varphi(X), \gamma(X, Y)) = (\varphi(X), \gamma^p(X, Y)) \\ &= \int_D \varphi(X) \wedge *_X \gamma^p(X, Y) . \end{aligned}$$

<u>Remark</u>. If $Y \in D$, then the integrand has a singularity at $X = Y$, and we have

$$\begin{aligned} (\varphi(X), \gamma(X, Y)) &= \lim_{\varepsilon \to 0} (\varphi(X), \gamma(X, Y))_{D-B_\varepsilon} \\ &= \lim_{\varepsilon \to 0} \int_{D-B_\varepsilon} \varphi(X) \wedge *_X \gamma^p(X, Y) , \end{aligned}$$

where B_ε is the ball of radius $\varepsilon > 0$ centered at Y, and $D - B_\varepsilon$ denotes the set consisting of the points in D but not in B_ε where ε is sufficiently small that $\overline{B}_\varepsilon \subset D$. Then $D - B_\varepsilon$ is also a finite domain.

The boundary bB_ε of B_ε will be denoted by S_ε and is an $(n-1)$-sphere about Y with radius ε. Then

$b(D - B_\varepsilon) = bD - S_\varepsilon$. Thus the notation is consistent with the fact that, as part of the boundary of $D - B_\varepsilon$, the $(n-1)$-sphere S_ε is taken with the opposite orientation from that induced on it as bB_ε .

9.7. Proposition. Let $\varphi \in \mathcal{A}^p(D)$. Then, if $Y \in D$, we have

$$\varphi(Y) = (\Gamma\Delta\varphi)(Y)$$

$$(13) \qquad - \int_{bD} \{\delta\varphi(X) \wedge *_X\gamma(X, Y) - \gamma(X, Y) \wedge *d\varphi(X) - \delta_X\gamma(X, Y) \wedge *\varphi(X) + \varphi(X) \wedge *_X d_X\gamma(X, Y)\} .$$

If Y lies in the exterior of D, then the right side of (13) is equal to zero.

Remark. Proposition 9.7 remains true for an arbitrary riemannian metric, provided that $\gamma(X, Y)$ is a fundamental solution for the corresponding laplacian.

Proof. Suppose $Y \in D$ and consider the integral

$$(\Delta\varphi(X), \gamma(X, Y))_{D-B_\varepsilon} = \int_{D-B_\varepsilon} \Delta\varphi(X) \wedge *_X\gamma(X, Y) .$$

By Theorem 8.10, (v), for the finite domain $D - B_\varepsilon$, we have

$$(\Delta\varphi, \gamma(X, Y))_{D-B_\varepsilon} = (\varphi(X), \Delta_X\gamma(X, Y))_{D-B_\varepsilon} + \int_{bD-S_\varepsilon} \{\delta\varphi(X) \wedge *_X\gamma(X, Y) - \gamma(X, Y) \wedge *d\varphi(X) - \delta_X\gamma(X, Y) \wedge *\varphi(X) + \varphi(X) \wedge *_X d_X\gamma(X, Y)\} .$$

Here, $(\varphi(X), \Delta_X\gamma(X, Y))_{D-B_\varepsilon} = 0$, since $\Delta_X\gamma(X, Y) = 0$ for

$X \neq Y$, by Lemma 9.5. Further,

$$\lim_{\varepsilon \to 0} (\Delta\varphi(X), \gamma(X, Y))_{D-B_\varepsilon} = (\Gamma\Delta\varphi)(Y) .$$

Hence, to prove formula (13), it is sufficient to show that

$$(14) \quad \varphi(Y) = -\lim_{\varepsilon \to 0} \int_{S_\varepsilon} \{\varphi(X) \wedge *_X d_X \gamma(X, Y) - \delta_X \gamma(X, Y) \wedge *\varphi(X)\}$$

and

$$(15) \quad 0 = \lim_{\varepsilon \to 0} \int_{S_\varepsilon} \{\delta\varphi(X) \wedge *_X \gamma(X, Y) - \gamma(X, Y) \wedge *d\varphi(X)\} .$$

We give the proofs for the case $n > 2$.

The result (15) follows from the fact that, for $X \in S$,

$$f(X, Y) = \frac{1}{(n-2)v_{n-1}} \frac{1}{\varepsilon^{n-2}} ,$$

while the volume element in S_ε is $\varepsilon^{n-1} d\Omega$ (cf. Exercise 11.1).

For (14), let the coordinates $(x^1, \ldots, x^n)$ of integration be replaced by new coordinates

$$u^i = \frac{1}{\varepsilon} (x^i - y^i) , \qquad i = 1, \ldots, n .$$

Then, for $X \in S_\varepsilon$, the point U with coordinates $(u^1, \ldots, u^n)$ lies on the unit $(n-1)$-sphere $S_1 = S$, and the integral on the right side in (14), for fixed $\varepsilon > 0$, becomes

$$\int_S \{\varphi(Y + \varepsilon U) \wedge *_U d_U \gamma(U, \tilde{0}) - \delta_U \gamma(U, \tilde{0}) \wedge *_U \varphi(Y + \varepsilon U)\} .$$

Letting $\varepsilon \to 0$, we see that it is sufficient to show that

$$(14') \quad \varphi(Y) = -\int_S \{\psi(U) \wedge *_U d_U \gamma(U, \tilde{0}) - \delta_U \gamma(U, \tilde{0}) \wedge *\psi(U)\} ,$$

where

$$\psi(U) = \lim_{\varepsilon \to 0} \varphi(Y + \varepsilon U)$$
$$= \Sigma_{i_1 < \ldots < i_p} (\varphi(Y))_{i_1 \ldots i_p} du^{i_1} \wedge \ldots \wedge du^{i_p} .$$

Since $\int_S d\Omega(U) = v_{n-1}$, where $d\Omega$ is the volume element in S, it is sufficient to show that the tangential component of the integrand on the right in (14') is equal to

$$(16) \qquad - \frac{1}{v_{n-1}} \Sigma_{i_1 < \ldots < i_p} (\varphi(Y))_{i_1 \ldots i_p} d\Omega(U) \circledcirc dy^{i_1} \wedge \ldots \wedge dy^{i_p} .$$

The computations leading to this result are given in Exercises 11.2 and 11.3.

If Y lies in the exterior of D, hence at a positive distance from the closure of D, then $\gamma(X, Y) \in \mathcal{A}(D)$, and it is not necessary to excise the ball B_ε. Then Theorem 8.10, (v) may be applied directly to the integral $(\Delta\varphi(X), \gamma(X, Y))_D = (\Gamma\Delta\varphi)(Y)$ to show that the right side of (13) vanishes in this case.

<u>Remark</u>. If Y lies in the boundary of D, but is not an exceptional point, then a similar argument shows that the right hand member of (13) has the value $\varphi(Y)/2$.

9.8. <u>Proposition</u>. If $\varphi \in \mathcal{A}^p(D)$, then $\Gamma\varphi \in \mathcal{A}^p(D)$. That is, Γ is a linear (over R) transformation

$$(17) \qquad \Gamma: \mathcal{A}^p(D) \longrightarrow \mathcal{A}^p(D) , \qquad p = 0, 1, \ldots, n .$$

<u>Proof</u>. We shall sketch a proof, leaving the details as an exercise. We note first that $|X - Y|^{-2+\varepsilon}\gamma(X, Y)$ is of class

C^0 for any $\varepsilon > 0$, so $(\varphi(X), |X - Y|^{-2+\varepsilon}\gamma(X, Y))$ defines a differential form which is of class C^0 in the closure of D.

Secondly, applying $\partial/\partial y^k$ to (the coefficients of) the form $(\Gamma\varphi)(Y)$, where Y has coordinates $(y^1,\ldots,y^n)$, we infer from a standard theorem of calculus that the differentiation commutes with the integration, namely

$$\begin{aligned}(\frac{\partial}{\partial y^k}\Gamma\varphi)(Y) &= (\varphi(X), \frac{\partial}{\partial y^k}\gamma(X, Y)) \\ &= -(\varphi(X), \frac{\partial}{\partial x^k}\gamma(X, Y))\end{aligned}$$

since

$$\frac{\partial}{\partial y^k}\gamma(X, Y) = -\frac{\partial}{\partial x^k}\gamma(X, Y) .$$

Integrating by parts, we see that $\frac{\partial}{\partial y^k}\Gamma\varphi$ is equal to $(\frac{\partial}{\partial x^k}\varphi(X), \gamma(X, Y))$ plus an integral over the boundary which is differentiable of class C^∞ in the interior of D and of class C^0 in the closure of D. It then follows that

$$\frac{\partial^2}{\partial y^k \partial y^\ell}\Gamma\varphi$$

is of class C^0 in D and, integrating by parts in the boundary integral (no further boundary terms appearing because bD is compact), we infer that $\frac{\partial^2}{\partial y^k \partial y^\ell}\Gamma\varphi$ is of class C^0 in the closure of D, etc.

9.9. <u>Theorem</u>. If $\varphi \in \mathcal{A}^p(D)$, then

$$\Delta\Gamma\varphi = \varphi .$$

In particular, the linear transformation (17) is injective.

Proof. Let $\varphi \in \mathcal{A}^p(D)$, and let ψ be an arbitrary p-form with compact support relative to D. Then, by (13), since the boundary integral vanishes, we have

$$\psi = \Gamma\Delta\psi .$$

Since $\gamma(X, Y) = \gamma(Y, X)$, the operator Γ is (formally) self-adjoint, that is,

$$(\Gamma\lambda, \eta) = (\lambda, \Gamma\eta) , \qquad \lambda, \eta \in \mathcal{A}^p(D) .$$

Thus, since ψ has compact support, we have $(\Gamma\varphi, \Delta\psi) = (\varphi, \Gamma\Delta\psi) = (\varphi, \psi)$. That is,

$$(\Delta\Gamma\varphi, \psi) = (\varphi, \psi)$$

for every $\psi \in \mathcal{A}^p(D)$ with compact support. Hence, as is easily verified, $\Delta\Gamma\varphi = \varphi$.

9.10. Proposition. If $\varphi \in \mathcal{A}^p(D)$, then

$$d\Gamma\varphi = \Gamma d\varphi - \int_{bD} \varphi(X) \wedge *_X\gamma(X, Y) ,$$

$$\delta\Gamma\varphi = \Gamma\delta\varphi - \int_{bD} \gamma(X, Y) \wedge *\varphi(X) .$$

These formulas are immediate consequences of Lemma 9.5 and Theorem 8.10, (i).

If φ has a compact support relative to D, then it follows from Proposition 9.10 that $d\Gamma\varphi = \Gamma d\varphi$ and $\delta\Gamma\varphi = \Gamma\delta\varphi$. That is, in the subspace of $\mathcal{A}^p(D)$ composed of forms with compact support, d and δ commute with Γ.

9.11. Theorem. Let f be a function of class C^2

which satisfies $\Delta f = 0$ at each point of D. Then f is of class C^∞ in D.

Remark. Theorem 9.11 is a special case of a more general theorem in potential theory (more generally, in the theory of elliptic differential operators) which is called "Weyl's lemma".

Proof. Let B_r be an n-ball which, together with its boundary S_r, lies in D. By Proposition 9.7, applied to B_r, we have, since $\Delta f = 0$ and $\delta f = \delta\gamma^0(X, Y) = 0$,

$$(18) \quad f(Y) = -\int_{S_r} \{f(X) \wedge *_X d_X \gamma(X, Y) - \gamma(X, Y) \wedge *df(X)\} .$$

For $X \in S$, $\gamma(X, Y)$ is of class C^∞ in its dependence on $Y \in B_r$, so we conclude that $f(Y)$ is of class C^∞ in B_r. Since D is the union of balls of the above type, it follows that f is of class C^∞ in D.

9.12. Theorem (Gauss mean-value theorem). Let f be a harmonic function in D. Then, at each point $Y \in D$, $f(Y)$ is equal to the average of its values over any $(n-1)$-sphere S_r centered at Y which lies, together with its interior, entirely in D.

Proof. We choose the $(n-1)$-sphere S_r in (18) to be centered at Y. Then

$$- t \{f(X) \wedge *_X d_X \gamma(X, Y)\} = \frac{1}{r^{n-1} v_{n-1}} f(X) r^{n-1} d\Omega$$

by Exercise 11.3. On the other hand,

$$\int_{S_r} \gamma(X, Y) \wedge *df(X) = \frac{1}{(n-2) v_{n-1} r^{n-2}} \int_{S_r} *df = 0 .$$

In fact, by Theorem 8.10, (iii), we have

$$0 = (d1, df) - (1, \Delta f) = \int_{S_r} *df \ .$$

Thus (18) gives, in this case,

$$f(Y) = \frac{1}{r^{n-1}v_{n-1}} \int_{S_r} f(X)r^{n-1}d\Omega \ .$$

Since the volume of S_r is $r^{n-1}v_{n-1}$, this means that $f(Y)$ is the average of the values of f on S_r.

9.13. Corollary (Maximum principle). A harmonic function in D which has a maximum at an interior point of D is equal to a constant.

The proof is left as an exercise.

§10. Harmonic forms and cohomology

10.1. Definition. Let D be a bounded open set of V (i.e. an open set whose closure is a compact subset of V), and suppose that each boundary point of D has a neighborhood U in V which can be mapped bidifferentiably onto the open unit coordinate ball $B = \{t \mid \Sigma_{j=1}^{n}(t^j)^2 < 1\}$ of R^n in such a way that the intersection of the boundary of D with U corresponds to the hyperplane $t^n = 0$ in B. Then we may say that D is a finite domain with smooth (i.e. differentiable) boundary.

10.2. Proposition. A finite domain D with smooth boundary is a finite domain (with boundary) in the sense of Definition 8.2, i.e. there exists a regular n-chain c of V whose support coincides with the closure of D.

A proof of this proposition is omitted.

It will be assumed throughout this section that $D \subset V$ is a finite domain with smooth boundary, where V has a fixed riemannian metric and orientation.

We shall denote by $\mathcal{A}^p_\nu$ the linear subspace of $\mathcal{A}^p$ composed of forms φ satisfying $n\varphi = nd\varphi = 0$ on bD (or, equivalently, $t*\varphi = t*d\varphi = 0$), and by $\mathcal{A}^p_\tau$ the linear subspace of $\mathcal{A}^p$ composed of forms φ satisfying $t\varphi = t\delta\varphi = 0$ on bD (or, equivalently, $n*\varphi = nd*\varphi = 0$). We denote by $\mathbb{H}^p_\nu$, $\mathbb{H}^p_\tau$ the linear subspaces of $\mathcal{A}^p_\nu$, $\mathcal{A}^p_\tau$ respectively composed of harmonic differential forms.

<u>Remarks</u>. Compact supported forms in $\mathcal{A}^p$ obviously are in $\mathcal{A}^p_\nu \cap \mathcal{A}^p_\tau$. Also, the isomorphism $*: \mathcal{A}^p \longrightarrow \mathcal{A}^{n-p}$ clearly induces an isomorphism (over R)

$$*: \mathcal{A}^p_\nu \longrightarrow \mathcal{A}^{n-p}_\tau ,$$

since $*n = t*$, $*t = n*$.

10.3. <u>Proposition</u>. The vector space $\mathbb{H}^p_\nu$ consists of the forms $\varphi \in \mathcal{A}^p$ which satisfy $d\varphi = \delta\varphi = 0$ in D and $n\varphi = 0$ on bD; the vector space $\mathbb{H}^p_\tau$ consists of the forms φ satisfying $d\varphi = \delta\varphi = 0$ in D and $t\varphi = 0$ on bD. Moreover, the isomorphism $*: \mathcal{A}^p \longrightarrow \mathcal{A}^{n-p}$ induces an isomorphism

$$*: \mathbb{H}^p_\nu \longrightarrow \mathbb{H}^{n-p}_\tau .$$

<u>Proof</u>. The first two statements follow from (iv) of Theorem 8.10 (with $\varphi = \psi$). The last statement follows from the

fact that $\Delta * \varphi = * \Delta \varphi$ (Theorem XI, 5.12).

10.4. <u>Theorem</u>. $\mathbb{H}^p_\nu$ is isomorphic to H^p.

This is a difficult theorem, whose proof is based on the theory of elliptic differential operators. However, a theorem equivalent to Theorem 10.4 in the case $p = 1$ was stated (without proof) by William Thompson (Lord Kelvin) approximately one hundred years ago. Namely, a form $\varphi \in \mathbb{H}^1_\nu$ may be interpreted as the (steady) velocity of an ideal incompressible fluid circulating in D, and Thompson asserted the existence and uniqueness of a fluid flow with prescribed circulations (periods) on the homology classes.

Theorem 10.4 shows that any cohomology class, or homology class, is represented by a unique harmonic differential form. The representation of more general types of cohomology by (more general) harmonic forms is a powerful analytic tool in investigating "structures" (such as complex analytic structures — see Chapter XIII) on compact topological spaces.

In topology there is a type of cohomology, called "relative cohomology", in particular, "cohomology relative to the boundary of D", which we shall denote by $H^p(bD)$. The vector space $\mathbb{H}^p_\tau$ is isomorphic to $H^p(bD)$ and we have the diagram

$$\begin{array}{ccc} \mathbb{H}^p_\nu & \xrightarrow{*} & \mathbb{H}^{n-p}_\tau \\ \updownarrow & & \updownarrow \\ H^p & \xrightarrow{L} & H^{n-p}(bD) \end{array}$$

where the isomorphism $L: H^p \longrightarrow H^{n-p}(bD)$, called the Lefschetz

duality theorem, corresponds to map $*: \mathbb{H}^p_\nu \longrightarrow \mathbb{H}^{n-p}_\tau$ in the representation by harmonic forms.

§11. Exercises

1. Generalized spherical coordinates $r, \theta^1, \theta^2, \ldots, \theta^{n-2}, \varphi$ may be introduced in R^n by the formulas

$$
\begin{aligned}
x^1 &= r \cos \theta^1 , \\
x^2 &= r \sin \theta^1 \cos \theta^2 , \\
&\vdots \\
x^{n-2} &= r \sin \theta^1 \sin \theta^2 \ldots \sin \theta^{n-3} \cos \theta^{n-2} , \\
x^{n-1} &= r \sin \theta^1 \sin \theta^2 \ldots \sin \theta^{n-3} \sin \theta^{n-2} \cos \varphi , \\
x^n &= r \sin \theta^1 \sin \theta^2 \ldots \sin \theta^{n-3} \sin \theta^{n-2} \sin \varphi ,
\end{aligned}
$$

where $0 < r < \infty$, $0 < \theta^i < \pi$, $i = 1, \ldots, n-2$, and $0 < \varphi < 2\pi$. Show that the Jacobian determinant of this transformation is given by

$$J_n = r^{n-1} \sin^{n-2}\theta^1 \sin^{n-3}\theta^2 \ldots \sin^2\theta^{n-3} \sin \theta^{n-2} > 0 .$$

Show that the associated basis $dr, d\theta^1, \ldots, d\theta^{n-2}, d\varphi$ is orthogonal, but not orthonormal (except at certain points of R^n). Thus

$$dx^1 \wedge \ldots \wedge dx^n = r^{n-1} dr \wedge d\Omega$$

where

$$r^{n-1} d\Omega = J_n d\theta^1 \wedge \ldots \wedge d\theta^{n-2} \wedge d\varphi ,$$

and dr is orthogonal to $d\theta^1, \ldots, d\theta^{n-2}$, $d\varphi$. Show that

$$|dr| = 1 , \qquad |r^{n-1} d\Omega| = 1 .$$

Therefore the volume element induced by the euclidean metric in the $(n-1)$-sphere

$$S_r^{n-1} = \{X \mid X \in R^n \text{ and } |X| = r\}$$

of radius r is $r^{n-1} d\Omega$. Let

$$v_{n-1} = \int_{S_1^{n-1}} d\Omega$$

denote the volume of the unit $(n-1)$-sphere. Show that we have the recurrence formula

$$v_{n-1} = v_{n-2} \int_0^{\pi} \sin^{n-2}\theta \, d\theta ;$$

where $v_1 = 2\pi$, and determine v_2, v_3, v_4 and v_5.

2. If spherical coordinates are used in computing normal and tangential components on the boundary S_r of the ball B_r, then the normal component of a differential form φ, of positive degree, at a point $X \in S_r$ is given by the terms in the expression for φ which involve the associated basis element dr.

Show that the decomposition into normal and tangential components is orthogonal, that is,

$$n\varphi \cdot t\psi = 0 ,$$

where φ and ψ are p-forms at $X \in S_r$, and therefore

$$\varphi\cdot\psi = t\varphi\cdot t\psi + n\varphi\cdot n\psi \ .$$

Show that, for any p-forms φ, ψ and $(n-p-1)$-form χ, $p = 0, 1, \ldots, n$, we have

$$n(dr\wedge\varphi) = dr\wedge t\varphi \ ,$$

$$t*(dr\wedge\varphi) = *(dr\wedge t\varphi) \ ,$$

$$*(dr\wedge\varphi)\cdot\chi = t(\varphi\wedge\chi)\cdot r^{n-1}d\Omega \ ,$$

$$t(\varphi\wedge*(dr\wedge\psi)) = (t\varphi\cdot t\psi)r^{n-1}d\Omega \ ,$$

$$t(*\varphi\wedge*(dr\wedge*\psi)) = (n\varphi\cdot n\psi)r^{n-1}d\Omega \ .$$

3. Show that, if we compute in terms of spherical coordinates centered at $Y \in R^n$, $n > 2$, we have

$$d_X\gamma(X, Y) = -\frac{n-2}{r}\, dr\wedge\gamma(X, Y) \ ,$$

$$\delta_X\gamma(X, Y) = (-1)^{np+n}\,\frac{n-2}{r}\, *_X(dr\wedge *_X\gamma(X, Y)) \ .$$

Use these formulas with those derived in Exercise 2 to verify formula (16) of §9.

4. Give proofs of Propositions 9.7 and 9.12 for the case $n = 2$.

5. Prove Corollary 9.13.

6. Show that the (open) unit ball in R^n, centered at $\tilde{0}$, is a finite domain with smooth boundary.

7. Prove Proposition 8.6.

8. Let $V = R^n$, and let D be a finite subdomain of R^n with smooth boundary. Show that formula (v) of Theorem 8.10 applied to functions f, g ($f, g \in \mathcal{A}^0(D)$) coincides with the

classical formula of calculus, by verifying that ndf is the usual "normal derivative".

9. A twice differentiable function f which satisfies $\Delta f \leq 0$ at each point of an open set $D \subset R^n$ is said to be <u>subharmonic</u> in D. Is a subharmonic function necessarily differentiable of class C^∞ in D? Give an example of a subharmonic function on R^n.

10. Let f be a subharmonic function on $D \subset R^n$. Prove that, at each point $Y \in D$, $f(Y)$ does not exceed the average of its values over a sphere centered at Y which lies, together with its interior, entirely in D.

11. Show that a subharmonic function on an open set $D \subset R^n$ which has a maximum at an interior point of D is equal to a constant.

12. Let $D \subset R^2$ be the annulus described in Exercise 7.1. Find a non-trivial harmonic form ω of degree 1 satisfying $d\omega = \delta\omega = 0$ in D with vanishing normal component.

13. Let $D \subset R^n$ be the (open) unit ball centered at $\tilde{0}$. Prove that there exists a function $g(X, Y)$, $X, Y \in D$, such that: (a) $g(X, Y) - \gamma^0(X, Y)$ is harmonic in D as a function of X for $X \neq Y$ and bounded for X near Y; (b) for each fixed $Y \in D$, $g(X, Y)$ vanishes for X on the boundary of D.

14. Show that $g(X, Y)$ is uniquely determined by the conditions (a) and (b) of Exercise 13 and that $g(X, Y) = g(Y, X)$. Write

$$(Gf)(Y) = \int_D f(X) \wedge *g(X, Y) , \qquad f \in \mathcal{A}^0(D) ,$$

and show that

$$(2) \qquad f(Y) = (G\Delta f)(Y) - \int_{bD} f(X) \wedge *_X d_X g(X, Y) .$$

15. By means of formula (2) prove that, given $f \in \mathcal{A}^0(D)$, there exists a harmonic function h whose values on the boundary of the ball D coincide with those of f (Dirichlet's problem, or first boundary-value problem, for the ball D). Is h unique?

16. Prove that a bounded harmonic function on R^n is equal to a constant (<u>theorem of Liouville</u>).

17. Suppose that f is a harmonic function on the unit ball D of R^n, centered at $\tilde{0}$, with continuous boundary values which vanish on an open set S of the unit sphere bounding D. Let $\tilde{X} = (\tilde{x}^1, \tilde{x}^2, \ldots, \tilde{x}^n)$ be the inverse point of $X = (x^1, x^2, \ldots, x^n)$ with respect to the unit sphere, namely $\tilde{x}^j = x^j/|X|^2$ where $|X|^2 = \Sigma_{j=1}^n (x^j)^2$. For $|X| > 1$ define $f(X) = -f(\tilde{X})/|X|^{n-2}$ and show that $f(X)$, so defined, provides the unique continuation of f as a harmonic function across S into the exterior of the unit sphere (<u>Kelvin reflection principle</u>).

18. Suppose that D is a finite subdomain of R^n, and let σ_{n-1} be an $(n-1)$-simplex belonging to the boundary of a regular n-chain associated with D. Suppose that f is a harmonic function in D which can be continued (locally) across $\operatorname{supp} \sigma_{n-1}$ as a continuous function which is differentiable except on

supp σ_{n-1}, satisfies $\Delta f = 0$ on either side of supp σ_{n-1}, and has a continuous normal derivative across supp σ_{n-1}. Prove that the continuation is differentiable and harmonic in a neighborhood in R^n of each interior point of supp σ_{n-1} (<u>continuation principle for harmonic functions</u>).

XIII. COMPLEX STRUCTURE

§1. Introduction

The purpose of this chapter is to show that the real tensor calculus of the preceding chapters, especially the calculus of differential forms, has a complex analogue. In particular, in terms of an hermitian metric (i.e. a riemannian metric compatible with the complex structure), the Green's formulas of Theorem XII, 8.10 have complex analogues in which the real operators d, δ and the laplacian Δ are replaced by complex ones. These integral formulas contain most of the usual ones in complex analysis, including Cauchy's formula in the case of complex dimension 1.

§2. Complex vector spaces

In Chapters I - XII, the field of scalars has always been the field R of real numbers. We now consider the case where the field of scalars is the field C of complex numbers, i.e. the numbers of the form $c = a + bi$ where $a, b \in R$ and $i^2 = -1$. So far as addition is concerned, these numbers can be identified with a real vector space $C_o = R^2$ by taking as (orthonormal) basis the elements $1 = (1, 0)$ and $i = (0, 1)$. The field C is obtained from C_o by giving an additional operation, the usual multiplication of complex numbers, relative to which C is not only an algebra with unit element (cf. Chapter II, §9) but a field, since the multiplication is commutative and every non-zero element has an inverse.

Two particular automorphisms of C_o should be noted. The first, <u>conjugation</u>, sends $c = a + bi$ into $\bar{c} = a - bi$ and is a symmetric orthogonal transformation S_o with proper values 1 and -1. The linear subspace of C_o consisting of the proper vectors corresponding to the proper value 1 is left elementwise fixed by S_o and is identified with the subfield R = the real numbers. (Actually S_o preserves multiplication as well, that is, $\overline{cc'} = \bar{c}\,\bar{c'}$, and may be considered as an automorphism of the field C which leaves the subfield R fixed.) S_o is an involution, or automorphism of period 2, since $S_o^2 = I$. The second automorphism, J_o, is the skew-symmetric orthogonal transformation of C_o determined by $J_o 1 = i$, $J_o i = -1$, that is, $J_o c = ic$, and corresponds to a rotation through 90°. The transformation J_o has no proper values (in particular, no non-zero vector remains fixed) and does not preserve multiplication in C. J_o is an automorphism of C_o of period 4, since $J_o^2 = -I$, $J_o^4 = I$.

2.1. <u>Definition</u>. A <u>complex vector space</u> is a set V satisfying the axioms of Definition I, 1.1 with the field R of real numbers replaced by the field C of complex numbers.

Most definitions and theorems of Chapters I - V and IX have complex analogues. For any notion which depends on properties common to both the real and complex numbers, it is sufficient to substitute "complex" for "real".

2.2. <u>Definition</u> (I, 3.1, 3.2). The <u>cartesian complex k-dimensional space</u>, $k > 0$, denoted by C^k, is the complex vector space whose elements are sequences $(a_1, a_2, \ldots, a_k)$ of k

complex numbers, with the operations of addition and scalar multiplication defined by

$$(a_1,a_2,\ldots,a_k) + (b_1,b_2,\ldots,b_k) = (a_1 + b_1,\ a_2 + b_2,\ldots,a_k + b_k)$$

and, for any $c \in C$,

$$c(a_1,a_2,\ldots,a_k) = (ca_1,ca_2,\ldots,ca_k) .$$

In particular, $C^1 = C$ is the field of complex numbers.

2.3. Definition (I, 7.1, 7.4). A non-empty subset U of a (complex) vector space V is called a (complex) linear subspace of V if it satisfies the conditions

(i) if $A \in U$ and $B \in U$, then $A + B \in U$,

(ii) if $A \in U$ and $c \in C$, then $cA \in U$.

Any non-empty subset D of V determines a (complex) linear subspace $L(D)$ whose elements are the vectors of V which can be expressed as a finite linear combination

$$(1) \qquad c_1A_1 + \ldots + c_kA_k , \qquad A_j \in D,\ c_j \in C .$$

2.4. Definition (I, 9.1, 10.1, 10.2, 10.3). A set D of vectors in V is called dependent (relative to C) if $\tilde{0}$ can be expressed as a non-trivial linear combination (1); otherwise D is called independent. A subset D in V is called a (complex) basis for V if D is independent and $L(D) = V$. V is called finite dimensional if it has a (complex) basis consisting of a finite number of vectors. [In this case, every (complex) basis

contains the same number of vectors (this number being called the <u>complex dimension</u> of V) and, if D is a (complex) basis for V, each vector in V can be expressed in the form (1) in one, and only one, way.]

2.5. <u>Definition</u> (II, 1.1). If V and W are (complex) vector spaces, a function $T: V \longrightarrow W$ is called a <u>linear transformation</u> if

(i) $\quad T(A + B) = TA + TB$, $\quad$ for all $A, B \in V$,

(ii) $\quad T(cA) = cTA$, $\quad$ for all $A \in V, c \in C$.

If (ii) is replaced by

($\overline{\text{ii}}$) $\quad T(cA) = \overline{c}TA$, $\quad$ for all $A \in V, c \in C$,

then T is called a <u>conjugate linear transformation</u>.

<u>Examples</u>. The identity transformation $I: V \longrightarrow V$ is linear, as is the function $J: V \longrightarrow V$ defined by $JA = iA$ for all $A \in V$. The function $S: C \longrightarrow C$, $c \longrightarrow \overline{c}$, is conjugate linear. A linear transformation T satisfies $TJ = JT$ and is therefore holomorphic (cf. Definition 3.3). A conjugate linear transformation T satisfies $TJ = -JT$.

2.6. <u>Proposition</u> (II, 2.2, 2.3, 7.1, 2.5, 7.4). If $T: V \longrightarrow W$ is linear (or conjugate linear), then:

(i) $T(U)$ is a linear subspace of W if U is a linear subspace of V,

(ii) $T^{-1}(U)$ is a linear subspace of V if U is a linear

subspace of W (in particular, $\ker T = T^{-1}(\tilde{0}_W)$ is a linear subspace of V),

(iii) T is injective if and only if $\ker T = \tilde{0}_V$,

(iv) if V is finite dimensional,

$$\dim_C V = \dim_C (\ker T) + \dim_C (\operatorname{im} T) ,$$

(v) if T is bijective, T^{-1} is linear (or conjugate linear).

2.7. Proposition (II, 8.1). If V and W are complex vector spaces, the set $L(V, W)$ of linear transformations from V to W and the set $\bar{L}(V, W)$ of conjugate linear transformations from V to W are complex vector spaces. If V and W have complex dimensions k and n, then $L(V, W)$ and $\bar{L}(V, W)$ have complex dimension kn.

2.8. Proposition (II, 2.1). The composition of two linear (or conjugate linear) transformations is linear.

2.9. Proposition (II, 9.1, 9.2). The set $E(V) = L(V, V)$ of endomorphisms of V is an algebra with unit if multiplication is defined by composition of endomorphisms. The subset $A(V) \subset E(V)$ of automorphisms of V forms a group with respect to this multiplication.

Remark. The set of conjugate linear transformations of V into itself does not have the above properties: the composition of two elements of the set is not in the set, and this set does not include I.

In defining the scalar product in the real case, the property

$a^2 \geqq 0$ for $a \in R$ is basic, and this property does not hold for complex numbers. However, a more general property: $c\bar{c} \geqq 0$ for $c \in C$, holds. (This includes the property for the subfield of the real numbers which are characterized by the condition $c = \bar{c}$.)

2.10. Definition (III, 1.1). An hermitian scalar product in a complex vector space V is a function which assigns to each pair of vectors A, B in V a complex number, denoted by $A \cdot B$, such that

1. For all $A, B \in V$, $A \cdot B = \overline{B \cdot A}$ (hermitian symmetry),
2. For all $A, B \in V$, and $c \in C$, $cA \cdot B = c(A \cdot B)$,
3. For all $A, B, C \in V$, $(A + B) \cdot C = A \cdot C + B \cdot C$,
4. For all A in V, $A \cdot A \geqq 0$,
5. $A \cdot A = 0$ if and only if $A = \tilde{0}$.

Remarks. Axioms 2 and 3 state that, for fixed $B \in V$, the function $T: V \longrightarrow C$ defined by $TA = A \cdot B$ is linear. On the other hand, the function $T: V \longrightarrow C$ defined by $TA = B \cdot A$ is conjugate linear. In fact,

$$T(cA) = B \cdot cA = \overline{cA \cdot B} = \bar{c}(\overline{A \cdot B}) = \bar{c}(B \cdot A) = \bar{c}TA \quad .$$

The standard hermitian scalar product (III, 2.3) in n-dimensional complex cartesian space C^n is defined by

$$(2) \qquad (a_1, a_2, \ldots, a_n) \cdot (b_1, b_2, \ldots, b_n) = \Sigma_{j=1}^{n} a_j \overline{b_j} \quad .$$

Since $A \cdot A \geqq 0$, by Axiom 4, it is again possible to define the length (or absolute value) of a vector A by

$$|A| = \sqrt{A \cdot A} \quad .$$

Then we have

2.11. Proposition (III, 3.2). The length function has the following properties:

(i) For each $A \in V$, $|A| \geqq 0$.

(ii) If $A \in V$, then $|A| = 0$ if and only if $A = \tilde{0}$.

(iii) For each $A \in V$ and each $c \in C$, $|cA| = |c||A|$.

(iv) For each pair A, B in V, $|A \cdot B| \leqq |A|\ |B|$ (Schwarz inequality).

(v) For each pair A, B in V, $|A + B| \leqq |A| + |B|$ (triangle inequality).

2.12. Definition (III; 5.1). Two vectors A and B in V are orthogonal if $A \cdot B = 0$. A (complex) basis in V is orthogonal if each two distinct basis vectors are orthogonal. A vector A in V is normal if $A \cdot A = 1$. A (complex) basis in V is orthonormal if it is orthogonal and if each basis vector is normal.

2.13. Proposition (III, 5.6). Each finite dimensional subspace of a complex vector space with hermitian scalar product has an orthonormal basis.

2.14. Definition (III, 6.5, 6.2). An endomorphism T of V which preserves the hermitian scalar product:

$$TA \cdot TB = A \cdot B \ , \qquad\qquad A, B \in V \ ,$$

is called <u>unitary</u> (the analogue of orthogonal in the real case). The unitary transformations form a subgroup of the group of automorphisms of V.

2.15. <u>Definition</u> (V, 4.1). If V is a complex vector space and $T \in E(V)$, then a non-zero vector $A \in V$ is called a <u>proper vector</u> if $TA = \lambda A$ for some $\lambda \in C$. The complex number λ is called a <u>proper value</u> of T.

<u>Remark</u>. We shall not, in this brief survey, review all the constructions of Chapter IX explicitly. It is to be emphasized that, whenever the symbols $\otimes$ and $\wedge$ of the tensor and exterior products are used, it must always be made clear which field of scalars, R or C, is being used. If V has complex dimension n and $T \in E(V)$, then the <u>determinant</u> of T is the complex number $\det T$ determined, as in Chapter IX, §9, by the condition

$$\wedge^n T(A_1 \wedge \ldots \wedge A_n) = TA_1 \wedge \ldots \wedge TA_n = \det T\ (A_1 \wedge \ldots \wedge A_n)$$

for arbitrary $A_j \in V$. Here the exterior product is defined with C as the field of scalars. Further,

$$\det (TS) = \det T \cdot \det S$$

and, if $T \in A(V)$, then $\det T \neq 0$ and

$$\det T^{-1} = 1/\det T \quad .$$

2.16. <u>Proposition</u> (V, 4.5, 4.7). If $T \in E(V)$, where $\dim_C V = n$, then $\det (T - zI)$ is a polynomial, with complex coefficients, of degree n in z. The set of proper values of T

coincides with the set of complex roots of the equation

$$\det (T - zI) = 0 \quad . \tag{3}$$

Remark. The field C is algebraically closed (in contrast to the field R) and the polynomial equation (3) always has n roots in C (but not necessarily distinct ones).

2.17. Proposition (V, 4.9). If (3) has distinct roots $\lambda_1, \ldots, \lambda_n$, then there is a complex basis $A_1, \ldots, A_n$ for V satisfying $TA_j = \lambda_j A_j$, $j = 1, \ldots, n$.

2.18. Definition (V, 5.1, 5.3). Let V be finite dimensional, with an hermitian scalar product, and let $T \in E(V)$. The adjoint T^* of T is the endomorphism $B \longrightarrow T^*B$ where T^*B is determined by the condition

$$TA \cdot B = A \cdot T^*B \qquad \text{for all } A \in V \quad .$$

2.19. Lemma (V, 5.4). With respect to an orthonormal basis in V, the matrix (α^*_{ij}) of T^* is the conjugate transpose of the matrix (α_{ij}) of T, that is

$$\alpha^*_{ij} = \overline{\alpha_{ji}} \ , \tag{4}$$

Remark. It is obvious from (4) that we also have $\det T^* = \overline{\det T}$.

2.20. Proposition (V, 5.5, 6.2). As a function from $E(V)$ to $E(V)$, the passage to the adjoint has the following properties:

(i) $(S + T)^* = S^* + T^*$ for all $S, T \in E(V)$,

(ii) $(cT)^* = \bar{c}T^*$ for all $T \in E(V), c \in C$,

(iii) $T^{**} = T$ for all $T \in E(V)$,

(iv) $(ST)^* = T^*S^*$ for all $S, T \in E(V)$.

Remark. Because of (ii), this function defines a conjugate linear, rather than a linear, involution of $E(V)$, considered as a vector space.

2.21. Proposition (V, 3.1). If $T \in A(V)$ is unitary, then $T^* = T^{-1}$ and $|\det T| = 1$. Further, any proper value of λ of T satisfies $|\lambda| = 1$.

Proof. If T is unitary, then

$$TA \cdot TB = A \cdot B \qquad \text{for all } A, B \in V .$$

Thus, for a fixed $B \in V$, we have

$$A \cdot T^*TB = TA \cdot TB = A \cdot B \qquad \text{for all } A \in V ,$$

which implies $T^*TB = B$ by the complex analogue of Lemma V, 5.2.

If $TA = \lambda A$, then

$$A \cdot A = TA \cdot TA = \lambda\bar{\lambda}(A \cdot A) .$$

Since $A \cdot A \neq 0$, this implies $\lambda\bar{\lambda} = 1$.

2.22. Definition (V, 5.6). $T \in E(V)$ is called hermitian (or self-adjoint) if $T^* = T$.

2.23. Proposition. If T is an hermitian transformation, then all its proper values are real, as is $\det T$.

Proof. If $TA = \lambda A$, then $T^*A = \lambda A$, and

$$\lambda(A\cdot A) = TA\cdot A = A\cdot T^*A = \overline{\lambda}(A\cdot A) \quad .$$

Since $A\cdot A \neq 0$, this implies $\lambda = \overline{\lambda}$, that is, λ is real.

2.24. Theorem (V, 7.2). Let $\dim_C V = n$. If $T \in E(V)$ is unitary or hermitian (or, more generally, satisfies $TT^* = T^*T$), then there is an orthonormal complex basis $A_1, \ldots, A_n$ consisting of proper vectors of T.

The proof of this proposition is omitted.

§3. Relations between real and complex vector spaces

On any complex vector space V, scalar multiplication by real numbers is, of course, defined. Relative to addition, and scalar multiplication by real numbers only, the elements of V clearly form a real vector space, which will be denoted by V_o and called the real vector space underlying the complex vector space V.

A subscript o will be used to distinguish structures for which the field of scalars is R rather than C.

3.1. Proposition. Let V, W be complex vector spaces and V_o, W_o the underlying real vector spaces. Every linear or conjugate linear transformation $T: V \longrightarrow W$ induces a linear transformation $T_o: V_o \longrightarrow W_o$.

Proof. Any function $T: V \longrightarrow W$ induces a function $T_o: V_o \longrightarrow W_o$. If T is linear or conjugate linear, then T satisfies (i) and (ii) (or ($\overline{\text{ii}}$)) of Definition 2.5. If c is taken to be a real number, these are precisely the conditions that T_o be

linear.

3.2. Corollary. If V_o is the underlying real vector space of a complex vector space V, then there is an automorphism J_o of V_o satisfying $J_o^2 = -I_o$, induced by the automorphism J of V given by $JA = iA$.

It is left as an exercise to verify that, if $T \in L(V, W)$, then $T_oJ_o = J_oT_o$ and that, if $S \in \overline{L}(V, W)$, then $S_oJ_o = -J_oS_o$. (On the left J_o denotes the automorphism induced on V_o and on the right, the automorphism induced on W_o.)

3.3 Definition. If V_o, W_o are real vector spaces, on each of which is given an automorphism J_o satisfying $J_o^2 = -I_o$, then $T_o \in L_o(V_o, W_o)$ is called holomorphic (relative to the given J_o's) if $T_oJ_o = J_oT_o$.

3.4. Proposition. The composition of holomorphic transformations is holomorphic. The holomorphic automorphisms of a vector space V_o (with given J_o) form a subgroup of the group of automorphisms of V_o.

The proof of this proposition is left as an exercise.

3.5. Proposition. Let V, W be complex vector spaces and V_o, W_o the underlying real vector space. Let J_o on V_o and W_o be induced by the complex structures of V and W, respectively, as in Corollary 3.2. Then every holomorphic $T_o \in L_o(V_o, W_o)$ corresponds to a $T \in L(V, W)$. Analogously, if $S_o \in L_o(V_o, W_o)$ satisfies $S_oJ_o = -J_oS_o$, then S_o induces an $S \in \overline{L}(V, W)$.

Proof. Clearly, T_o determines a function $T: V \longrightarrow W$

sending $A \in V$ into the element $TA \in W$ identified with $T_oA \in W_o$, computed for A considered as an element of V_o. Note that $cA = (a + ib)A \in V$ is to be identified with $aA + bJ_oA \in V_o$, etc. Then, since

$$T_o(aA + bJ_oA) = a\, T_oA + b\, T_oJ_oA = a\, T_oA + b\, J_oT_oA\ ,$$

it is clear that $T(cA) = cTA$.

3.6. Proposition. Let W_o be the underlying real vector space of a finite dimensional complex vector space W. Then

(i) $$\dim_R W_o = 2 \dim_C W\ .$$

(ii) If $A_1, \ldots, A_n$ is a complex basis for W, the vectors $A_1, J_oA_1, \ldots, A_n, J_oA_n$ give a basis for W_o.

(iii) All bases of the type described in (ii) correspond to the same orientation of W_o, that is, W_o has a natural orientation induced by the complex structure of W.

(iv) Any holomorphic automorphism of W_o preserves this orientation.

Proof. If $A_1, \ldots, A_n$ is a complex basis for W, then every $A \in W$ can be expressed in one and only one way in the form

$$A = \Sigma_{j=1}^n\, c^j\, A_j\ ,$$

and at least one $c^j \neq 0$ if $A \neq \tilde{0}$. If we write $c^j = a^j + ib^j$, $j = 1, \ldots, n$, then

$$A = \Sigma_{j=1}^n\, a^j\, A_j + \Sigma_{j=1}^n\, b^j(iA_j)$$

where the real coefficients a^j, b^j are all zero only if $A = \tilde{0}$. If we consider A as an element of W_o and identify iA_j with J_oA_j, this shows that the $2n$ vectors $A_1, J_oA_1, \ldots, A_n, J_oA_n$ form a (real) basis for W_o. If $B_1, J_oB_1, \ldots, B_n, J_oB_n$ is a second basis for W_o of the above type, and if

$$B_\ell = \Sigma_{j=1}^n a_\ell{}^j A_j + \Sigma_{j=1}^n b_\ell{}^j J_oA_j ,$$

then

$$J_oB_\ell = \Sigma_{j=1}^n - b_\ell{}^j A_j + \Sigma_{j=1}^n a_\ell{}^j J_oA_j \quad .$$

Now if α, β denote real $n \times n$ matrices, we have (see Proposition 3.17)

$$\det \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix} = \det \begin{pmatrix} \alpha+i\beta & 0 \\ 0 & \alpha-i\beta \end{pmatrix} = \det (\alpha+i\beta)\, \overline{\det (\alpha+i\beta)} \geq 0.$$

Thus, the determinant corresponding to the change of basis, which is not zero, must be positive. Finally, let T_o be a holomorphic automorphism of W_o. Let $A_1, J_oA_1, \ldots, A_n, J_oA_n$ be a basis for W_o, and set $B_\ell = T_oA_\ell$, $\ell = 1, \ldots, n$. Since T_o is holomorphic, $J_oB_\ell = T_oJ_oA_\ell$ and, since T_o is an automorphism, $B_1, J_oB_1, \ldots, B_n, J_oB_n$ is a basis for V_o. Then $\det T_o > 0$ as above.

3.7. <u>Definition</u>. Let W_o be a real vector space. The <u>complexification</u> of W_o, denoted by $\mathbb{C}W_o$, is the complex vector space constructed as follows:

Let $U_o = C_o \otimes W_o$, where C_o is the underlying real vector space of $C = C^1$, and the tensor product is taken over R.

Let $U = \mathbb{C}W_o$ be the complex vector space whose elements are the vectors $\Sigma\, c \otimes A$ of U_o, $c \in C$, $A \in W_o$, with scalar multiplication by complex numbers defined by $c'(c \otimes A) = c'c \otimes A$.

It is clear that U is a complex vector space with U_o as underlying real vector space. If W_o is finite dimensional, say $\dim_R W_o = k$, then $\dim_R U_o = 2k$, by Proposition IX, 2.18, and $\dim_C U = k$.

<u>Example</u>. $C^k = \mathbb{C}R^k$.

3.8. <u>Proposition</u>. Let W_o, U_o, and $U = \mathbb{C}W_o$ be as in Definition 3.7, and let J_o be the automorphism of U_o induced by the complex structure of U, that is, $J_o(c \otimes A) = ic \otimes A$. Then

$$U_o = W_o \oplus J_o W_o \quad .$$

Further, there is an automorphism S_o of U_o which is an involution: $S_o^{\,2} = I_o$, which leaves W_o elementwise fixed, and which satisfies $S_o J_o = -J_o S_o$ and therefore induces a conjugate linear involution S of $U = \mathbb{C}W_o$.

<u>Proof</u>. As a subspace of U_o, W_o is identified with the set of elements of the form $1 \otimes A$, $A \in W_o$. Then $J_o W_o$ is the set of elements of the form $J_o(1 \otimes A) = i \otimes A$, $A \in W_o$. The direct sum decomposition of U_o is then obvious. The involution S_o of U_o is defined by $S_o(c \otimes A) = \bar{c} \otimes A$, $c \in C$, $A \in W_o$, and the remaining statements are easily verified.

3.9. <u>Definition</u>. A subset W_o of a complex vector space V will be called a <u>real subspace</u> of V if W_o is a linear subspace of V_o.

If W_o is a real subspace of V and if $C_o \otimes W_o$ is the underlying real vector space of $\mathbb{C}W_o$, with J_o determined by the complex structure of $\mathbb{C}W_o$, then the inclusion $\iota_o: W_o \longrightarrow V_o$ can be extended to a holomorphic transformation $\iota_o: C_o \otimes W_o \longrightarrow V_o$ by setting $\iota_o(c \otimes A) = cA$, $c \in C$, $A \in W_o$, where $cA \in V$ is considered as an element of V_o.

3.10. <u>Definition</u>. A real subspace W_o of a complex vector space V is called a <u>real form</u> of V if $\iota_o: C_o \otimes W_o \longrightarrow V_o$ is an isomorphism.

<u>Example</u>. A real vector space is a real form of its complexification.

<u>Remarks</u>. If W is a complex linear subspace of V, then W_o cannot be a real form of V. For example, $iA \in W_o$ for $A \in W_o$, so $\iota_o(1 \otimes iA) = \iota_o(i \otimes A)$; that is, ι_o cannot be injective. If W_o is a real form of V, then the isomorphism between V_o and $C_o \otimes W_o$ can be used to define another complex structure on $C_o \otimes W_o$, induced from the complex structure on V, by defining $J_o(1 \otimes A)$ to be the element B such that $\iota_o B = J_o A$, $A \in W_o$, $J_o A \in V_o$. For the new structure, the inclusion $A \longrightarrow 1 \otimes A$ of W_o in $C_o \otimes W_o$ is holomorphic.

3.11. <u>Proposition</u>. To each conjugate linear involution of a complex vector space V there corresponds a unique real form of V (and conversely).

<u>Proof</u>. Given a conjugate linear involution S of V, let W_o be the set of fixed points of V, that is, $A \in W_o$ if and only if $SA = A$. Since S_o is linear (Proposition 3.1), W_o

is a real subspace of V. To show that $C_o \otimes W_o \longrightarrow V_o$ is an isomorphism, it is sufficient to show that V_o also has a direct sum decomposition $W_o \oplus J_o W_o$. For any $A \in V_o$, we have

$$A = \frac{1}{2}(A + S_o A) + \frac{1}{2}(A - S_o A) ,$$

where the first term is an element of W_o, and the second can be expressed as $J_o(\frac{1}{2i}(A - S_o A))$, with $\frac{1}{2i}(A - S_o A) \in W_o$. Further, $W_o \cap J_o W_o = \tilde{0}$ since, for $A, B \in W_o$, the relation $A = J_o B$ implies

$$J_o B = A = S_o A = S_o J_o B = - J_o S_o B = - J_o B ,$$

or $B = A = \tilde{0}$.

Conversely, if W_o is a real form of V, then there is a conjugate linear involution S of V, induced from the one on $\mathbb{C}W_o$ (Proposition 3.8) by way of the isomorphism between $C_o \otimes W_o$ and V_o, which leaves the elements of W_o fixed: $SA = A$ if and only if $A \in W_o$.

3.12. <u>Proposition</u>. Any finite dimensional complex vector space V has a conjugate linear involution, and therefore a real form.

<u>Proof</u>. Let $A_1, \ldots, A_n$ be a complex basis for V and define $S: V \longrightarrow V$ by

$$S(\Sigma_{j=1}^n c^j A_j) = \Sigma_{j=1}^n \overline{c^j} A_j , \qquad c^j \in C .$$

(The set W_o of fixed points of S are the vectors of the form

$\Sigma_{j=1}^{n} c^j A_j$ with $\overline{c^j} = c^j$, that is, with c^j real, $j = 1, \ldots, n$.)

3.13. Definition. Let W_o be a real vector space, $0 < \dim_R W_o < \infty$. We say that a (homogeneous) complex structure is given on W_o if there is given an endomorphism J_o of W_o satisfying $J_o^2 = -I_o$.

(The endomorphism J_o is an automorphism, since J_o^{-1} exists and is given by $-J_o$.)

3.14. Theorem. Let W_o be a real vector space with a (homogeneous) complex structure defined by J_o, with $J_o^2 = -I_o$. Then:

(i) There exists a basis for W_o of the form

$$A_1, J_oA_1, \ldots, A_n, J_oA_n ;$$

in particular, $\dim_R W_o$ is even and W_o has a natural orientation.

(ii) There exists a complex space W such that W_o is the underlying real vector space of W and J_o is induced by the complex structure of W; in particular, $\dim_R W_o = 2 \dim_C W$.

Proof. Since $\dim_R W_o > 0$, by hypothesis, there exists a vector $A_1 \neq \tilde{0}$ in W_o. Then A_1 and J_oA_1 are independent. In fact, if there exist real numbers a, b such that

$$aA_1 + bJ_oA_1 = \tilde{0} ,$$

then

$$aJ_oA_1 - bA_1 = \tilde{0}$$

and

$$\tilde{0} = a(aA_1 + bJ_oA_1) = a^2A_1 + ab\, J_oA_1 = (a^2 + b^2)A_1 .$$

Since $A_1 \neq \tilde{0}$, this implies $a = b = 0$.

We proceed by induction, and assume that an independent set

$$(1) \qquad A_1,\ J_oA_1,\ \ldots,\ A_k,\ J_oA_k$$

of vectors in W_o has been found, where $k \geqq 1$. If $\dim_R W_o = 2k$, there is nothing further to prove, since these vectors give a basis for W_o. If $\dim_R W_o > 2k$, then there is a non-zero vector $A_{k+1} \in W_o$ which is independent of the vectors in (1). The vectors

$$(2) \qquad A_1,\ J_oA_1,\ \ldots,\ A_k,\ J_oA_k,\ A_{k+1},\ J_oA_{k+1}$$

form an independent set (and $\dim_R W_o \geqq 2(k+1)$). In fact, if $a_1, a_2, \ldots, a_{k+1}, b_1, b_2, \ldots, b_{k+1}$ are real numbers such that

$$(3) \qquad \Sigma_{j=1}^{k+1} a_jA_j + \Sigma_{j=1}^{k+1} b_jJ_oA_j = \tilde{0} ,$$

then

$$(4) \qquad \Sigma_{j=1}^{k+1} a_jJ_oA_j - \Sigma_{j=1}^{k+1} b_jA_j = \tilde{0} .$$

We multiply (3) by a_{k+1} and (4) by b_{k+1} and subtract, to obtain

$$(5) \qquad \begin{aligned} &\Sigma_{j=1}^{k}(a_ja_{k+1} + b_jb_{k+1})A_k + \Sigma_{j=1}^{k+1}(b_ja_{k+1} - a_jb_{k+1})J_oA_j \\ &\qquad + (a_{k+1}^2 + b_{k+1}^2)A_{k+1} = \tilde{0} . \end{aligned}$$

Since A_{k+1} is independent of the independent set (1), all coefficients in (5) are zero; in particular, $a_{k+1}^2 + b_{k+1}^2 = 0$, or

$a_{k+1} = b_{k+1} = 0$. Then (3) implies $a_j = b_j = 0$, $j = 1, \ldots, k$, since the set (1) is independent. Thus the set (2) is independent, as was to be shown.

The complex vector space W is constructed from the elements of W_o by defining the operation of scalar multiplication by a complex number $c = a + ib$ as follows

$$cA = aA + bJ_oA \ , \qquad A \in W_o \ .$$

Again, let W_o be a real vector space with a (homogeneous) complex structure defined by J_o, where $J_o^2 = -I_o$, and consider the complexification $\mathbb{C}W_o$ (Definition 3.7). If J_o on $C_o \otimes W_o$ is induced from the complex structure on $\mathbb{C}W_o$, then the inclusion

$$\iota_o : W_o \longrightarrow C_o \otimes W_o \ ,$$

$A \longrightarrow 1 \otimes A$, is not holomorphic. In order for ι_o to be holomorphic, the construction of a complex vector space V overlying $C_o \otimes W_o$, by defining scalar multiplication by complex numbers, should be based on the formula

$$J_o(c \otimes A) = c \otimes J_oA \ , \tag{6}$$

rather than on the natural multiplication

$$J_o(c \otimes A) = ic \otimes A \tag{7}$$

as in Definition 3.7. Thus we have two complex vector spaces V and $\mathbb{C}W_o$ with the same underlying real vector space $C_o \otimes W_o$. We shall write out the natural multiplication of (7) explicitly,

for $\mathbb{C}W_o$, and reserve the symbols J_o or J for the new structure V. Then $\iota_o(W_o)$ is a real form of $\mathbb{C}W_o$, but not of V. In fact, ι_o is holomorphic relative to the new structure and therefore induces a linear transformation $\iota : W \longrightarrow V$, where W is the complex vector space constructed over W_o (Theorem 3.14, (ii)), so ιW, which lies over $\iota_o W_o$ in V, is a linear subspace of V. For simplicity, the vector space V is not mentioned, and we think of the complex structure defined by J, and induced by the (homogeneous) complex structure J_o on W_o.

Note first that multiplication by i commutes with J_o, since

$$iJ_o(c \otimes A) = i(c \otimes J_oA) = ic \otimes J_oA ,$$
$$J_o(i(c \otimes A)) = J_o(ic \otimes A) = ic \otimes J_oA$$

Consequently, natural multiplication by a fixed complex number gives a linear transformation in the new structure while, on the other hand, J_o induces a linear transformation J of the complex vector space $\mathbb{C}W_o$, with $J^2 = -I$.

Let $P(\mathbb{C}W_o)$ denote the set of elements of $\mathbb{C}W_o$ for which the new multiplication agrees with the natural multiplication; that is, an element B of $\mathbb{C}W_o$ is in $P(\mathbb{C}W_o)$ if and only if

$$JB = iB , \quad \text{or} \quad B = \tfrac{1}{2}(B - iJB) \quad .$$

It is clear (by the complex analogue of Proposition V, 4.2) that $P(\mathbb{C}W_o)$ is a complex linear subspace of $\mathbb{C}W_o$. If we define an endomorphism P of $\mathbb{C}W_o$ by

$$P = \tfrac{1}{2}(I - iJ) ,$$

then P is a projection: $P^2 = P$, and therefore defines a direct sum decomposition

$$\mathbb{C}W_o = P(\mathbb{C}W_o) \oplus Q(\mathbb{C}W_o) , \tag{8}$$

where

$$Q = I - P = \tfrac{1}{2}(I + iJ)$$

is also a projection, and $P(\mathbb{C}W_o) = \operatorname{im} P$, $Q(\mathbb{C}W_o) = \operatorname{im} Q = \ker P$ (cf. Proposition II, 11.10, but generalize to the case of complex vector spaces). Note that for elements B in $Q(\mathbb{C}W_o)$ we have $JB = -iB$. That is,

$$JP = iP , \qquad JQ = -iQ$$

To compute JB for an arbitrary $B \in \mathbb{C}W_o$, we write

$$B = PB + QB ,$$

and then (since J is linear)

$$JB = iPB - iQB ,$$

or $J = i(P - Q)$.

There is a linear involution S_o of $C_o \otimes W_o$, leaving fixed the points identified with W_o, defined by

$$S_o(c \otimes A) = \bar{c} \otimes A , \qquad A \in W_o , \ c \in C_o .$$

We have $S_o i = -iS_o$, so the involution S of $\mathbb{C}W_o$ induced by

S_o is conjugate linear relative to the natural complex structure on $\mathbb{C}W_o$. (However, $S_oJ_o = J_oS_o$; in particular, S does not coincide with the involution associated with the new structure, as in Proposition 3.11.) Let $A_1, J_oA_1, \ldots, A_n, J_oA_n$ be a real basis for W_o (Proposition 3.14); these same vectors give a complex basis for $\mathbb{C}W_o$. For any $A \in \mathbb{C}W_o$, we have

$$A = \Sigma_{\alpha=1}^{n} c^{\alpha}A_{\alpha} + \Sigma_{\alpha=1}^{n} d^{\alpha} JA_{\alpha} ,$$

and

$$SA = \Sigma_{\alpha=1}^{n} \overline{c^{\alpha}} A_{\alpha} + \Sigma_{\alpha=1}^{n} \overline{d^{\alpha}} JA_{\alpha} .$$

That is, S is computed by taking complex conjugates of the coefficients c^{α}, d^{α}.

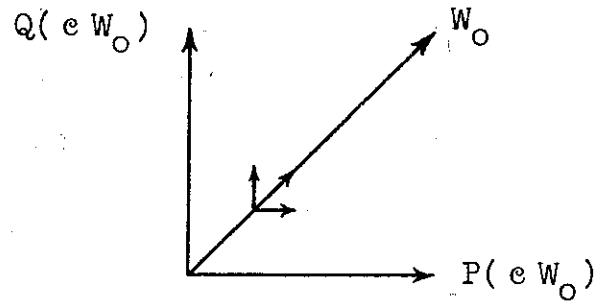
Alternatively,

$$B_{\alpha} = PA_{\alpha} = \tfrac{1}{2}(A_{\alpha} - iJA_{\alpha}) , \qquad \alpha = 1, \ldots, n ,$$

forms a (complex) basis for $P(\mathbb{C}W_o)$ and

$$B_{\overline{\alpha}} = QA_{\alpha} = \tfrac{1}{2}(A_{\alpha} + iJA_{\alpha}) , \qquad \overline{\alpha} = n + \alpha ,$$

gives a basis for $Q(\mathbb{C}W_o)$. Further, $SB_{\alpha} = B_{\overline{\alpha}}$, $SB_{\overline{\alpha}} = B_{\alpha}$, since $SA_{\alpha} = A_{\alpha}$, $SJA_{\alpha} = JA_{\alpha}$. Thus S transforms $P(\mathbb{C}W_o)$ onto $Q(\mathbb{C}W_o)$ and $Q(\mathbb{C}W_o)$ onto $P(\mathbb{C}W_o)$, while leaving fixed the subspace of $\mathbb{C}W_o$ corresponding to points of W_o. That is, $S: \mathbb{C}W_o \longrightarrow \mathbb{C}W_o$ may be regarded as reflection on the "diagonal" W_o, and the direct sum decomposition of $\mathbb{C}W_o$ may be pictured schematically as in the following diagram:



With respect to the above bases, an arbitrary element A of $\mathbb{C}W_o$ may be represented uniquely in the form

$$A = \Sigma_{\alpha=1}^{n} z^{\alpha} B_{\alpha} + \Sigma_{\bar{\alpha}=n+1}^{2n} z^{\bar{\alpha}} B_{\bar{\alpha}} . \tag{9}$$

Then

$$SA = \Sigma_{\alpha=1}^{n} \overline{z^{\alpha}} B_{\bar{\alpha}} + \Sigma_{\bar{\alpha}=n+1}^{2n} \overline{z^{\bar{\alpha}}} B_{\alpha} .$$

Thus, if a "self-conjugate" basis B_{α}, $B_{\bar{\alpha}}$ is used, S is again computed by taking conjugates. In view of these properties, the projection Q is often denoted by $\bar{P}$. However, Q has the property implied by this notation only when restricted to W_o (see, however, Definition 5.7). In fact, the elements of W_o are characterized by the condition $A = SA$, that is, $z^{\bar{\alpha}} = \overline{z^{\alpha}}$. (It is customary to write $\bar{z}^{\alpha}$ rather than $\overline{z^{\alpha}}$, etc.)

3.15. <u>Definition</u> (cf. Definition XI, 1.6). The coordinates $(z^1, \ldots, z^n, \bar{z}^1, \ldots, \bar{z}^n)$ are called a system of <u>self-conjugate complex coordinates</u> on W_o.

If $B_1', \ldots, B_n', B_{\bar{1}}', \ldots, B_{\bar{n}}'$ is another basis of the above type, then

$$(10) \qquad B_\ell' = \Sigma\, c_\ell^{\,j} B_j \,, \quad B_{\bar\ell}' = \Sigma\, c_{\bar\ell}^{\,\bar j} B_{\bar j}$$

since the new basis must span the same subspaces $P(cW_o)$ or $Q(cW_o)$, respectively. Further, $SB_\ell' = B_{\bar\ell}'$ implies $c_{\bar\ell}^{\,\bar j} = \bar c_\ell^{\,j}$. Note that

$$B_1' \wedge \ldots \wedge B_n' = \det c\ (B_1 \wedge \ldots \wedge B_n) \,,$$

where $\det c$ is the determinant of the change of basis on $P(cW_o)$. (Note: the exterior product is defined with respect to the complex numbers.) From (10) we obtain

3.16. Proposition. Under a change of self-conjugate coordinates on W_o, the new coordinates z'^β are given by a linear function of the z^α and do not involve the $\bar z^\alpha$. Similarly, the coordinates $\bar z'^\beta$ do not depend on the z^α.

If $A_1', J_oA_1', \ldots, A_n', J_oA_n'$ is the real basis of W_o corresponding to the basis $B_1', \ldots, B_n', B_{\bar 1}', \ldots, B_{\bar n}'$, then

$$(11) \qquad \begin{aligned} A_\ell' &= \Sigma\, a_\ell^{\,j} A_j + \Sigma\, b_\ell^{\,j} JA_j \,, \\ JA_\ell' &= \Sigma - b_\ell^{\,j} A_j + \Sigma\, a_\ell^{\,j} JA_j \,, \end{aligned}$$

where $c_\ell^{\,j} = a_\ell^{\,j} + ib_\ell^{\,j}$.

3.17. Proposition. The (real) determinant of the change of basis (11) is equal to the determinant of the change of basis (10), and is therefore positive.

Proof. We have

$$B_1 \wedge \ldots \wedge B_n \wedge B_{\bar{1}} \wedge \ldots \wedge B_{\bar{n}}$$
$$= \frac{1}{2^{2n}} (A_1 - iJA_1) \wedge \ldots \wedge (A_n - iJA_n) \wedge (A_1 + iJA_1) \wedge \ldots \wedge (A_n + iJA_n)$$
$$= \frac{i^n}{2^n} A_1 \wedge \ldots \wedge A_n \wedge JA_1 \wedge \ldots \wedge JA_n ,$$

where the final expression is obtained from the one preceding it by discarding all terms in the exterior product which vanish because of a repeated factor, and noting that the 2^n terms which do not vanish are all equal. The same formula holds with primes added. Thus the determinants for (10) and (11) are equal, and the determinant corresponding to (10) has the value $\det c \, \overline{\det c} \geq 0$.

Finally, for $\alpha = 1, \ldots, n$, write

$$z^\alpha = x^\alpha + iy^\alpha , \qquad \bar{z}^\alpha = x^\alpha - iy^\alpha ;$$

then the real values x^α, y^α are given by

$$x^\alpha = \frac{1}{2}(z^\alpha + \bar{z}^\alpha) , \qquad y^\alpha = \frac{1}{2i}(z^\alpha - \bar{z}^\alpha) .$$

For a vector $A \in W_o$, we have $z^{\bar{\alpha}} = \bar{z}^\alpha$, and (9) gives

$$A = \Sigma_{\alpha=1}^n (x^\alpha A_\alpha + y^\alpha JA_\alpha) , \tag{12}$$

where $(x^1, y^1, \ldots, x^n, y^n)$ is a real coordinate system on W_o.

A complex basis for the complex vector space W constructed over W_o is given by the vectors $A_1, \ldots, A_n$ and any vector $A \in W$ can be written as

$$A = \Sigma_{\alpha=1}^n z^\alpha A_\alpha . \tag{13}$$

With $z^\alpha = x^\alpha + iy^\alpha$, the inclusion

$$\iota : W \longrightarrow \mathbb{C}W_o ,$$

holomorphic relative to J, sends A into ιA where ιA has the expression (12) or, equivalently,

$$\iota A = \Sigma_{\alpha=1}^{n} (z^\alpha B_\alpha + \bar{z}^\alpha B_{\bar{\alpha}}) , \qquad \bar{\alpha} = n + \alpha .$$

Now $(z^1, \ldots, z^n)$ gives a complex coordinate system on W and, in terms of the corresponding complex coordinate system $(z^1, \ldots, z^n, z^{\bar{1}}, \ldots, z^{\bar{n}})$ on $\mathbb{C}W_o$, the component functions of ι are $z^\alpha = z^\alpha$, $z^{\bar{\alpha}} = \bar{z}^\alpha$, $\alpha = 1, \ldots, n$. Caution: ιA does <u>not</u> have the expression (13) — this also gives an inclusion of W in $\mathbb{C}W_o$ but not the one induced by ι_o.

§4. Exercises

Let V be a complex vector space with an hermitian scalar product.

1. Show that for any $A, B \in V$, we have

$$|A|^2 + |B|^2 \geq 2 \operatorname{Re} A\cdot B$$

where $\operatorname{Re} A\cdot B$ denotes the real part of the complex number $A\cdot B$.

2. Under what conditions is it true that $|A + B| = |A| + |B|$? (Cf. Exercise III, 4.9).

3. Let V be finite dimensional and let $T: V \longrightarrow C$ be linear (conjugate linear). Show that there is a unique vector $A \in V$ (depending on T) such that $TX = X\cdot A$ ($TX = A\cdot X$) for each $X \in V$.

4. Show that the automorphism J of V defined by $JA = iA$ is unitary.

5. Let V be a finite dimensional complex vector space, and let $T \in E(V) = L(V, V)$. Let

$$\wedge V = \wedge^0 V \oplus \wedge^1 V \oplus \ldots \oplus \wedge^n V , \qquad n = \dim_C V ,$$

denote the complex exterior algebra over V. Show that a unique $\wedge T \in E(\wedge V)$ is determined by the axioms

(1) $\wedge T(X + Y) = \wedge TX + \wedge TY , \qquad X, Y \in \wedge V ,$

(2) $\wedge T(X \wedge Y) = \wedge TX \wedge \wedge TY , \qquad X, Y \in \wedge V ,$

(3) $\wedge Tc = c , \qquad c \in C = \wedge^0 V ,$

(4) $\wedge TX = TX , \qquad X \in V = \wedge^1 V .$

6. Let V be as in Exercise 5, but take $T \in \overline{L}(V, V)$. Show that a unique $\wedge T \in \overline{L}(\wedge V, \wedge V)$ is determined by the Axioms (1), (2), and (4) above, and

($\overline{3}$) $\wedge Tc = \overline{c} , \qquad c \in C = \wedge^0 V .$

7. Let $S, T: V \longrightarrow V$ be linear (or conjugate linear, or one of each) and let $R = ST$. Show that $\wedge R = \wedge S \wedge T$, where the extensions are defined by the appropriate exercises above. Derive the following corollaries: (i) if S, T satisfy $ST = TS$, then the corresponding extensions $\wedge S, \wedge T$ satisfy $\wedge S \wedge T = \wedge T \wedge S$; (ii) if T satisfies $T^2 = I$, then $(\wedge T)^2 = I$; (iii) if $T \in E(V)$ satisfies $T^2 = -I$, then $(\wedge T)^2 X = (-1)^p X$ for

$X \in \wedge^p V$.

8. Let V be a complex vector space. For any $T \in E(V)$ and any $\omega \in V^* = L(V, C)$, show that $T^*\omega \in V^*$, where $T^*\omega: V \longrightarrow C$ is defined by

$$< X, T^*\omega > = < TX, \omega > \qquad \text{for all } X \in V \ .$$

Show that the function $T^*: V^* \longrightarrow V^*$ is linear. For $S \in \overline{L}(V, V)$ and any $\omega \in V^*$, show that $S^*\omega \in V^*$, where $S^*\omega: V \longrightarrow C$ is defined by

$$< X, S^*\omega > = \overline{< SX, \omega >} \qquad \text{for all } X \in V \ .$$

Show that the function $S^*: V^* \longrightarrow V^*$ is conjugate linear. If $S, T: V \longrightarrow V$ are linear (or conjugate linear, or one of each) and satisfy $ST = TS$, show that $T^*S^* = S^*T^*$.

Let W_o be a real vector space on which a scalar product is given.

9. Show that the scalar product on W_o induces an hermitian scalar product on cW_o, by defining, on $C_o \otimes W_o$,

$$c \otimes A \cdot c' \otimes A' = c\overline{c'}(A \cdot A') \qquad A, A' \in W_o, \ c, c' \in C \ .$$

(Note: this intermediate function is not a scalar product on the real space $C_o \otimes W_o$.) For $W_o = R^n$, with the standard scalar product, show that the standard hermitian scalar product is induced on $cW_o = C^n$.

10. Let W_o have a homogeneous complex structure defined

by an automorphism J_o, $J_o^2 = -I_o$, and suppose that J_o is orthogonal. Show that the induced automorphism J of $\mathbb{C}W_o$ is unitary relative to the hermitian scalar product of Exercise 9. (Then the given scalar product on W_o is called <u>hermitian</u> relative to J_o, cf. Definition 9.1.) Show also that $A \cdot JB = -JA \cdot B$ for $A, B \in \mathbb{C}W_o$.

11. Under the assumptions of Exercise 10, show that the decomposition (8) of §3 induced by J is orthogonal, that is,

$$A \cdot B = 0, \qquad A \in P(\mathbb{C}W_o),\ B \in Q(\mathbb{C}W_o) .$$

12. Show that J_o in Exercise 10 is orthogonal if it is only assumed that J_o preserves lengths, that is $|A| = |J_oA|$ for any $A \in W_o$. [Hint: start from $|A + B| = |J_oA + J_oB|$.]

§5. Complex differential calculus of forms

5.1. <u>Definition</u>. A subset D of a complex vector space W is called <u>open</u> if and only if the corresponding subset D_o of the underlying real vector space W_o is open. Analogous definitions are implied in the use of terms corresponding to other topological properties, such as compact, closed, etc.

5.2. <u>Definition</u>. Let W and $\tilde{W}$ be complex vector spaces, and let $D \subset W$ be open. A map or function $F: D \longrightarrow \tilde{W}$ is called <u>differentiable of class</u> C^k, $k = 0, 1, \ldots,$ if and only if the induced function $F_o: D_o \longrightarrow \tilde{W}_o$ is differentiable of class C^k.

All maps and functions will be assumed to be differentiable, i.e. differentiable of class C^∞, unless it is explicitly stated

otherwise.

The basic definitions for the operation of differentiation apply in the complex case also. Let W and $\tilde{W}$ be finite dimensional complex vector spaces and let $D \subset W$ be open. Let $F: D \longrightarrow \tilde{W}$ be differentiable. For $X \in D$ and $Y \in W$, define $F'(X, Y) \in \tilde{W}$ by

$$F'(X, Y) = \lim_{h \to 0} \frac{F(X + hY) - F(X)}{h}$$

This limit surely exists, since F is supposed differentiable, and we have

(1) $F'(X, Y + Z) = F'(X, Y) + F'(X, Z)$, $\qquad Y, Z \in W$,

(2) $F'(X, aY) = aF'(X, Y)$, $\qquad Y \in W, a \in R$,

(as in Propositions VII, 1.5 and 1.3). The method of proof for (2) fails if we try to replace $a \in R$ by $c \in C$, and the conclusion fails also. In fact, using only (1) and (2), we can verify that

$$(3) \quad \begin{aligned} F'(X, cY) = {} & c\{\tfrac{1}{2}(F'(X, Y) - iF'(X, iY))\} \\ & + \bar{c}\{\tfrac{1}{2}(F'(X, Y) + iF'(X, iY))\} \quad . \end{aligned}$$

Thus, in general, $F'(X, cY) \neq cF'(X, Y)$, and we cannot identity the complex vector space of tangent vectors at X with a complex vector space of differentiation operators on functions.

<u>Remark</u>. It is clear from (3) that the identification could be carried out if the set of functions F is restricted to those F for which $F'(X, Y) = -iF'(X, iY)$ for every $Y \in W$.

This is an important class of functions, which will be considered in more detail in §6. As a simple example, consider the case that $F: W \longrightarrow \tilde{W}$ is induced by a linear transformation $F_o: W_o \longrightarrow \tilde{W}_o$. For this choice of F, it is easily verified that the condition $F'(X, Y) = -iF'(X, iY)$ for every $Y \in W$ is equivalent to $F(iX) = iF(X)$, that is, to the condition that F be holomorphic.

Equation (3) involves multiplication by complex scalars in both W and $\tilde{W}$. If we write $JY = iY$ to express the multiplication in W, then (3) yields the decomposition

$$F'(X, Y) = \frac{1}{2}(F'(X, Y) - iF'(X, JY)) + \frac{1}{2}(F'(X, Y) + iF'(X, JY)) \quad .$$

This suggests that, in the complex case, the theory of differentiation operators should be based on the choice of $\mathbb{C}W_o$, rather than W, as "tangent space" at a point $X \in D$.

5.3. <u>Definition</u>. The elements of $\mathbb{C}W_o$ are called <u>complex tangent vectors</u> at $X \in D$, and $\mathbb{C}W_o$ can be identified with a complex vector space of differentiation operators on functions by defining

$$(4) \qquad u \cdot f = c(v \cdot f) \qquad \text{if } u_o = c_o \otimes v_o,\ v_o \in W_o ,$$

for (differentiable) functions $f: D \longrightarrow \mathbb{C}$.

We now have a precise complex analogue of the real case considered in Chapter XI, §2 and we define associated complex structures as follows (all tensor and exterior products, etc. being taken with respect to the field $\mathbb{C}$ as scalars).

5.4. <u>Definition</u> (XI, 2.3, 2.5). Let D be an open

subset of a finite dimensional complex vector space W. The <u>complex tangent space</u> $CT = CT(D)$ is the space $D \times \mathbb{C}W_o$, where the fibre $(CT)_X = \pi^{-1}(X)$ is the space of complex tangent vectors at $X \in D$. The (differentiable) sections of CT are the <u>complex vector fields</u> on D, and the set τ of complex vector fields on D is a complex vector space (of infinite dimension) or a free A^o-module of complex dimension $2n$, if $n = \dim_{\mathbb{C}} W$, where $A^o = A^o(D)$ is the ring of differentiable sections of $D \times \mathbb{C}$. Analogously, a section of $D \times \overset{s}{\otimes} \mathbb{C}W_o$ is a <u>complex tensor field</u>, contravariant of order s, etc. A section of $CT^* = D \times (\mathbb{C}W_o)^*$, is a <u>complex differential form of degree</u> 1, and the set of these sections is denoted by A^1. A section of $\Lambda^p CT^* = D \times \Lambda^p(\mathbb{C}W_o)^*$ is called a <u>complex differential form of degree</u> p, and the set of these sections is denoted by A^p. The <u>complex exterior algebra</u> of differential forms associated with D is

$$A = A^o \oplus A^1 \oplus \dots \oplus A^{2n} \quad .$$

<u>Remark</u>. The above definitions do not involve the complex structure of W, and can equally well be given for any real vector space W_o and any open $D_o \subset W_o$.

5.5. <u>Definition</u>. The (complex) <u>exterior derivative</u> is the operator

$$d: A \longrightarrow A$$

defined by the axioms of Definition XI, 4.1, except that in Axiom (3), the phrase "$u \in T_X$" is to be replaced by "$u \in (CT)_X$".

Remarks. The existence and uniqueness of d follow as in the real case, as do the properties (i) $d^2 = 0$, (ii) d is an operator of degree 1, (iii) $df = 0$ if $f \in A^0$ is constant, and (iv) d is an endomorphism of A if A is considered as a complex vector space.

So far, the complex exterior algebra and the complex exterior derivative appear to be a more or less routine copy of the analogous real case. However, the basic fibre $\mathbb{C}W_0$ has additional structure:

(i) a conjugate linear involution S of $\mathbb{C}W_0$ leaving fixed the points identified with points of W_0 (S does not depend on the choice of (homogeneous) complex structure on W_0);

(ii) an endomorphism J of $\mathbb{C}W_0$ such that $J^2 = -I$ (J represents the complex structure of W in $\mathbb{C}W_0$);

(iii) a direct sum decomposition

$$\mathbb{C}W_0 = P(\mathbb{C}W_0) \oplus Q(\mathbb{C}W_0) \tag{5}$$

where the projections P, Q are given by

$$P = \tfrac{1}{2}(I - iJ) \ , \quad Q = \tfrac{1}{2}(I + iJ) \ .$$

Both S and J induce (conjugate linear or linear) transformations, which (for simplicity) will again be denoted by S and J, in $(\mathbb{C}W_0)^*$ and in all tensor or exterior products of $\mathbb{C}W_0$ and of $(\mathbb{C}W_0)^*$ (cf. Exercises 4.5 - 4.8 and Definition 6.1). These transformations, defined first in each fibre, do not involve differentiation, so they extend immediately to the set of sections

of the corresponding product space. The direct sum decomposition (5) induces direct sum decompositions in the (complex) exterior algebra (as in Corollary IX, 7.7).

5.6. Definition. Any transformation induced by the conjugate linear involution S of $\mathbb{C}W_0$ will be called conjugation. If φ is a complex vector field, or a complex tensor field, or a complex differential form, etc., the transform $S\varphi$ of φ is called the conjugate of φ and is denoted by $\overline{\varphi}$. φ is called real if $\varphi = S\varphi$, i.e. if $\varphi = \overline{\varphi}$.

Examples. For $C = \Lambda^0(\mathbb{C}W_0)^*$, S is ordinary conjugation of complex numbers; thus, for $f \in A^0$, the function $\overline{f} \in A^0$ is defined by $\overline{f}(X) = \overline{f(X)}$, $X \in D$. If $f = \overline{f}$, then $f(X)$ must lie in the subset of C left fixed by S, that is, in R. Thus a real function on D is the same as a real-valued function on D. A real vector field on D must assign to each $X \in D$ a tangent vector in $\mathbb{C}W_0$ which is left fixed by S, that is, an element of W_0 (which therefore corresponds to an actual direction at X).

Any complex differential form φ can be expressed in terms of real differential forms by writing

$$\varphi = \frac{\varphi + \overline{\varphi}}{2} + i\,\frac{\varphi - \overline{\varphi}}{2i} ,$$

where $(\varphi + \overline{\varphi})/2$ and $(\varphi - \overline{\varphi})/2i$ are real. Note that this decomposition, although uniquely defined, does not give a direct sum decomposition of A over the complex numbers, but represents a relationship which might be written symbolically as $A = \mathbb{C}A_0$.

5.7. Definition. Let $E: A \longrightarrow A$ be any operator

(or transformation) on the complex exterior algebra A. The <u>conjugate operator</u> to E, denoted by $\overline{E}$, is defined by $\overline{E} = SES$. E is called a <u>real operator</u> if $\overline{E} = E$.

<u>Remark</u>. The notion of conjugate operators is not restricted to the case of operators on A, and requires only the properties of S. With this concept in mind, one may safely write $Q = \overline{P}$ in §3.

5.8. <u>Proposition</u>. If $E: A \longrightarrow A$ and $\varphi \in A$, then

$$\overline{E\varphi} = \overline{E}\,\overline{\varphi}\ , \quad \overline{E}\varphi = \overline{E\,\overline{\varphi}}\ .$$

The conjugate of $\overline{E}$ is E. If E is linear (or conjugate linear), then $\overline{E}$ is linear (or conjugate linear). E is a real operator if and only if E commutes with S, and then $E\varphi$ is real if φ is real. If $F: A \longrightarrow A$, then $\overline{F + E} = \overline{F} + \overline{E}$, $\overline{FE} = \overline{F}\,\overline{E}$. $E + \overline{E}$ is real.

<u>Proof</u>. The first statement is equivalent to the identity $SE = \overline{E}S$, and the remaining statements are proved similarly, using the properties of S.

<u>Examples</u>. Conjugation S is a real operator, as is J. This last follows from the fact that $JS = SJ$ on $\mathcal{C}W_o$, so that the same is true for the extensions of S and J. The exterior derivative d is real, since it can be expressed as the sum of conjugate operators, as will be seen in Theorem 5.13. Any real operator induces an operator in the (real) exterior algebra of differential forms associated with D_o; the transformation induced by S is the identity.

The direct sum decomposition (5) of $\mathbb{C}W_o$ gives a direct sum decomposition of the complex tangent space at X, for each $X \in D$, which we write as

$$(CT)_X = P((CT)_X) \oplus Q((CT)_X) \quad .$$

The direct sum decomposition (5) of $\mathbb{C}W_o$ also induces (e.g. by Exercise 4.8) a direct sum decomposition

$$(6) \qquad (\mathbb{C}W_o)^* = P((\mathbb{C}W_o)^*) \oplus Q((\mathbb{C}W_o)^*)$$

where, for $\omega \in (\mathbb{C}W_o)^*$, $P\omega \in (\mathbb{C}W_o)^*$ is determined by

$$< u, P\omega > = < Pu, \omega > \qquad \text{for all } u \in \mathbb{C}W_o \quad .$$

Note that $< u, P\omega > = 0$ if $u \in Q(\mathbb{C}W_o)$ and $< u, Q\omega > = 0$ if $u \in P(\mathbb{C}W_o)$.

The decomposition (6) induces a direct sum decomposition

$$A^1 = A^{1,o} \oplus A^{o,1}$$

where $\varphi \in A^{1,o}$ is a complex differential form which assigns to each $X \in D$ a 1-form in $P((\mathbb{C}W_o)^*)$.

As in Corollary IX, 7.7, the decomposition (6) also induces a decomposition of $\Lambda^p(\mathbb{C}W_o)^*$, $p > 1$, from which we obtain

$$A^p = \Sigma_{r+s=p} A^{r,s} = A^{p,o} \oplus A^{p-1,1} \oplus \ldots \oplus A^{o,p} \quad .$$

Note that for $p > n = \dim_{\mathbb{C}} W$, $A^{r,s} = \tilde{0}$ if $r > n$ or $s > n$. If we set $A^{o,o} = A^o$, then

$$A = \Sigma_p A^p = \Sigma_{0 \leq r,\, s \leq n} A^{r,s} \quad .$$

We denote by $\Pi^{r,s}$ the projection $A \longrightarrow A^{r,s}$ (cf. Proposition II, 11.13). Then $A^{r,s} = \Pi^{r,s}A$, and

$$\Pi^{r,s}\,\Pi^{r,s} = \Pi^{r,s} \ ,$$

$$\Pi^{r,s}\,\Pi^{r',s'} = \tilde{0} \qquad \text{unless} \quad r = r',\ s = s' \ ,$$

$$\Sigma_{0 \leq r,\, s \leq n}\, \Pi^{r,s} = I \quad .$$

5.9. Definition. The elements of $A^{r,s}$, $0 \leq r,\ s \leq n$, are called differential forms of type (r, s). For any $\varphi \in A$, $\Pi^{r,s}\varphi$ is called the component of φ of type (r, s).

The following statements are easily verified: (a) $S: A^{r,s} \longrightarrow A^{s,r}$ and $\overline{\Pi}^{r,s} = \Pi^{s,r}$; (b) $\varphi \in A$ is of type (r, s) if and only if $\Pi^{r,s}\varphi = \varphi$; (c) if $\varphi \in A$ is of type (r, s), then

$$< u_1 \wedge \ldots \wedge u_{r+s}, \varphi > = 0 \ , \qquad u_j \in (CT)_X \ ,$$

unless exactly r of the u_j's are in $P((CT)_X)$ and s of the u_j's are in $Q((CT)_X)$; (d) $\varphi \in A$ is zero if and only if $\Pi^{r,s}\varphi = 0$ for every pair (r, s); (e) $\varphi \wedge \psi \in A^{p+r,q+s}$ if $\varphi \in A^{p,q}$, $\psi \in A^{r,s}$.

5.10. Associated bases. Two particular kinds of basis for $\mathbb{C}W_0$ have been considered in §3. A "real" basis is one derived from a real basis $A_1, J_0A_1, \ldots, A_n, J_0A_n$ for W_0. The images $A_1, JA_1, \ldots, A_n, JA_n$ of these vectors in $\mathbb{C}W_0$ form a complex basis for $\mathbb{C}W_0$. A basis of this kind serves to introduce

a real linear coordinate system $(x^1, y^1, \ldots, x^n, y^n)$ in W_o and in any open $D_o \subset W_o$. As complex tangent vectors A_α, JA_α satisfy

$$A_\alpha \cdot f = \frac{\partial f}{\partial x^\alpha}, \qquad JA_\alpha \cdot f = \frac{\partial f}{\partial y^\alpha},$$

so that we may denote by $\frac{\partial}{\partial x^\alpha}$, $\frac{\partial}{\partial y^\alpha}$ the sections $X \longrightarrow A_\alpha$, $X \longrightarrow JA_\alpha$ of CT. These form a basis (over A^o) for τ. Similarly, the dual basis for A^1 may be denoted by dx^α, dy^α (cf. XI, 2.4, 2.5, 4.1). For these basis elements we have

$$S(\frac{\partial}{\partial x^\alpha}) = \frac{\partial}{\partial x^\alpha}, \quad S(\frac{\partial}{\partial y^\alpha}) = \frac{\partial}{\partial y^\alpha}, \quad Sdx^\alpha = dx^\alpha, \quad Sdy^\alpha = dy^\alpha .$$

Since S is a conjugate linear transformation, it follows that the real vector fields, tensor fields, differential forms (Definition 5.6) are precisely those which have real coefficients when expressed in terms of a "real" associated basis. To compute J in terms of this basis, we use

$$J(\frac{\partial}{\partial x^\alpha}) = \frac{\partial}{\partial y^\alpha}, \quad J(\frac{\partial}{\partial y^\alpha}) = -\frac{\partial}{\partial x^\alpha}, \quad Jdx^\alpha = -dy^\alpha, \quad Jdy^\alpha = dx^\alpha,$$

and the fact that J is linear.

From any "real" basis $A_1, JA_1, \ldots, A_n, JA_n$ for cW_o we can construct a complex basis

$$B_\alpha = \frac{1}{2}(A_\alpha - iJ_\alpha), \qquad B_{\bar\alpha} = \frac{1}{2}(A_\alpha + iJ_\alpha)$$

for cW_o, where the B_α form a complex basis for $P(cW_o)$ and the $B_{\bar\alpha}$ a complex basis for $Q(cW_o)$. A basis of this kind serves

to introduce a (self-conjugate) complex linear coordinate system $(z^1, \ldots, z^n, \bar{z}^1, \ldots, \bar{z}^n)$ in W_o, and in any open $D_o \subset W_o$, where

$$z^\alpha = x^\alpha + iy^\alpha , \qquad \bar{z}^\alpha = x^\alpha - iy^\alpha ,$$

$$x^\alpha = \frac{1}{2}(z^\alpha + \bar{z}^\alpha) , \qquad y^\alpha = \frac{1}{2i}(z^\alpha - \bar{z}^\alpha) .$$

As a complex tangent vector at $X \in D$, B_α satisfies

$$B_\alpha \cdot f = \frac{1}{2}(A_\alpha - iJA_\alpha)\cdot f = \frac{1}{2}(\frac{\partial f}{\partial x^\alpha} - i\frac{\partial f}{\partial y^\alpha}) = \frac{\partial f}{\partial z^\alpha} ,$$

and we denote by $\frac{\partial}{\partial z^\alpha}$ the section $X \longrightarrow B_\alpha$ of CT. Similarly,

$$B_{\bar{\alpha}} \cdot f = \frac{1}{2}(A_\alpha + iJA_\alpha)\cdot f = \frac{1}{2}(\frac{\partial f}{\partial x^\alpha} + i\frac{\partial f}{\partial y^\alpha}) = \frac{\partial f}{\partial \bar{z}^\alpha} .$$

However, the section $X \longrightarrow B_{\bar{\alpha}}$ will be denoted by $\frac{\partial}{\partial z^{\bar{\alpha}}}$, rather than by $\frac{\partial}{\partial \bar{z}^\alpha}$, even though its action as a differentiation operator is that of $\frac{\partial}{\partial \bar{z}^\alpha}$. This is done in order to have reliable rules for writing formulas. The rules will then be the same as those outlined in Chapter XI, 3.13, with the added requirement that formulas be homogeneous with respect to barred and unbarred indices. For example, a complex vector field u is expressed by

$$(8) \qquad u = \Sigma_{\alpha=1}^n u^\alpha \frac{\partial}{\partial z^\alpha} + \Sigma_{\alpha=1}^n u^{\bar{\alpha}} \frac{\partial}{\partial z^{\bar{\alpha}}} = Pu + Qu .$$

The dual basis for A^1 is easily verified to be

$$dz^\alpha = dx^\alpha + i\, dy^\alpha , \quad dz^{\bar{\alpha}} = dx^\alpha - i\, dy^\alpha ,$$

where the notation $dz^{\bar{\alpha}}$ is used for the reasons stated above.

(As a differential form, $dz^{\overline{\alpha}}$ coincides with the form df for $f = \overline{z}^{\alpha}$.) For these basis elements, we have

$$S(\frac{\partial}{\partial z^{\alpha}}) = \frac{\partial}{\partial z^{\overline{\alpha}}}, \quad S(\frac{\partial}{\partial z^{\overline{\alpha}}}) = \frac{\partial}{\partial z^{\alpha}}, \quad Sdz^{\alpha} = dz^{\overline{\alpha}}, \quad Sdz^{\overline{\alpha}} = dz^{\alpha},$$

and

$$J(\frac{\partial}{\partial z^{\alpha}}) = i\frac{\partial}{\partial z^{\alpha}}, \quad J(\frac{\partial}{\partial z^{\overline{\alpha}}}) = -i\frac{\partial}{\partial z^{\overline{\alpha}}}, \quad Jdz^{\alpha} = idz^{\alpha}, \quad Jdz^{\overline{\alpha}} = -idz^{\overline{\alpha}}.$$

The basis elements dz^{α} are all of type $(1, 0)$ and span $A^{1,0}$; the elements $dz^{\overline{\alpha}}$ are of type $(0, 1)$ and span $A^{0,1}$. For a form φ of type (r, s), $0 \leq r, s \leq n$, we have

$$(9) \qquad \varphi = \sum_{\substack{\alpha_1<\ldots<\alpha_r \\ \beta_1<\ldots<\beta_s}} \varphi_{\alpha_1\ldots\alpha_r\overline{\beta}_1\ldots\overline{\beta}_s} dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge dz^{\overline{\beta}_1}\wedge\ldots\wedge dz^{\overline{\beta}_s}$$

and for arbitrary φ, the component $\Pi^{r,s}\varphi$ is found by taking only those terms in the expression for φ which involve exactly r factors dz^{α} of type $(1, 0)$ and s factors $dz^{\overline{\beta}}$ of type $(0, 1)$.

As examples of computation we have, for u as in (8) and φ as in (9),

$$\overline{u} = \Sigma\, \overline{u^{\alpha}} \frac{\partial}{\partial z^{\overline{\alpha}}} + \Sigma\, \overline{u^{\overline{\alpha}}} \frac{\partial}{\partial z^{\alpha}},$$

$$Ju = i(\Sigma\, u^{\alpha} \frac{\partial}{\partial z^{\alpha}} - \Sigma\, u^{\overline{\alpha}} \frac{\partial}{\partial z^{\overline{\alpha}}}),$$

$$\overline{\varphi} = \sum_{\substack{\alpha_1<\ldots<\alpha_r \\ \beta_1<\ldots<\beta_s}} \overline{\varphi_{\alpha_1\ldots\alpha_r\overline{\beta}_1\ldots\overline{\beta}_s}}\, dz^{\overline{\alpha}_1}\wedge\ldots\wedge dz^{\overline{\alpha}_r}\wedge dz^{\beta_1}\wedge\ldots\wedge dz^{\beta_s}.$$

Also,

(10) $$J\varphi = i^r(-1)^s\varphi = i^{r+3s}\varphi = i^{r-s}\varphi \ , \qquad \varphi \in A^{r,s} \ .$$

5.11. Definition. An operator $E: A \longrightarrow A$ is called of type (μ, ν) if $E(A^{r,s}) \subset A^{r+\mu,s+\nu}$ for each pair (r, s), $0 \leq r, s \leq n$.

Remarks. If E is of type (μ, ν), then $\overline{E}$ is of type (ν, μ) so, if E is real, we must have $\mu = \nu$. The real operator J is of type $(0, 0)$. If $E: A \longrightarrow A$ is additive, that is, if $E(\varphi + \psi) + E\varphi + E\psi$, then E can be decomposed (uniquely) into a sum of operators $E^{\mu,\nu}$ of type (μ, ν), $0 \leq \mu, \nu \leq n$, by setting

$$E^{\mu,\nu} = \Sigma_{0 \leq r,s \leq n} \Pi^{r+\mu,s+\nu} E \Pi^{r,s} \ .$$

It can be also shown that E is zero if and only if $E^{\mu,\nu}$ is zero for each pair (μ, ν). If E is of type (μ, ν) and $F: A \longrightarrow A$ is of type (ρ, σ), then EF and FE are of type $(\mu + \rho, \nu + \sigma)$.

5.12. Definition. The operator

$$\partial: A \longrightarrow A$$

is defined by the axioms

[1] $\partial(\varphi + \psi) = \partial\varphi + \partial\psi \ , \qquad \varphi, \psi \in A \ ,$

[2] $\partial(\varphi \wedge \psi) = \partial\varphi \wedge \psi + (-1)^p\varphi \wedge \partial\psi, \qquad \varphi \in A^p, \psi \in A \ ,$

[3] for $f \in A^0$, ∂f is the differential form of degree 1 determined at each $X \in D$ by

$$< u, \partial f > = Pu \cdot f \qquad \text{for all } u \in (CT)_X \ ,$$

[4] $(\partial d + d\partial)f = 0$, $f \in A^0$.

The conjugate operator

$$\bar{\partial}: A \longrightarrow A$$

is (obviously) determined by the Axioms [1], [2], and [4], with $\bar{\partial}$ replacing ∂, and

[$\bar{3}$] for $f \in A^0$, $\bar{\partial}f$ is the differential form of degree 1 determined at each $X \in D$ by

$$< u, \bar{\partial}f > = Qu \cdot f \qquad \text{for all } u \in (CT)_X .$$

5.13. <u>Theorem</u>. Operators ∂ and $\bar{\partial}$ satisfying the above axioms exist and are uniquely determined by them. Moreover, ∂ and $\bar{\partial}$ are operators of types (1, 0) and (0, 1) respectively, and they satisfy

(11) $$\partial^2 = 0 , \qquad \partial\bar{\partial} + \bar{\partial}\partial = 0 , \qquad \bar{\partial}^2 = 0 ,$$

(12) $$d = \partial + \bar{\partial} .$$

<u>Remarks</u>. Thus, ∂ is the component of d of type (1, 0) and $\bar{\partial}$ the component of type (0, 1):

(13) $$\partial = \Sigma_{r,s}\, \Pi^{r+1,s}\, d\, \Pi^{r,s} , \qquad \bar{\partial} = \Sigma_{r,s}\, \Pi^{r,s+1}\, d\, \Pi^{r,s} .$$

<u>Proof</u>. For any $f \in A^0 = A^{0,0}$, the form ∂f determined by [3] lies in $A^{1,0}$ and the form $\bar{\partial}f$ determined by [$\bar{3}$] lies in $A^{0,1}$ (see (6)), and

$$df = \partial f + \bar{\partial} f \ , \qquad f \in A^0 \ , \tag{14}$$

since $P + Q = I$ and $< u, df > = u \cdot f$. In terms of a system $(z^1, \ldots, z^n, \bar{z}^1, \ldots, \bar{z}^n)$ of (self-conjugate) complex coordinates on D, these forms may be given explicitly as follows. For any $\omega \in A^1$, we have

$$\omega = \Sigma_{\alpha=1}^n \omega_\alpha dz^\alpha + \Sigma_{\alpha=1}^n \omega_{\bar{\alpha}} dz^{\bar{\alpha}} = \Pi^{1,0} \omega + \Pi^{0,1} \omega \ ,$$

where

$$\omega_\alpha = < \frac{\partial}{\partial z^\alpha}, \omega > \ , \qquad \omega_{\bar{\alpha}} = < \frac{\partial}{\partial z^{\bar{\alpha}}}, \omega > \ .$$

By [3] and [$\bar{3}$], we then have

$$\begin{aligned} \partial f &= \Sigma_{\alpha=1}^n < \frac{\partial}{\partial z^\alpha}, \partial f > dz^\alpha + \Sigma_{\alpha=1}^n < \frac{\partial}{\partial z^{\bar{\alpha}}}, \partial f > dz^{\bar{\alpha}} \\ &= \Sigma_{\alpha=1}^n \frac{\partial f}{\partial z^\alpha} dz^\alpha \end{aligned} \tag{15}$$

and

$$\bar{\partial} f = \Sigma_{\alpha=1}^n \frac{\partial f}{\partial \bar{z}^\alpha} dz^{\bar{\alpha}} \tag{16}$$

since $P(\frac{\partial}{\partial z^\alpha}) = \frac{\partial}{\partial z^\alpha}$, $P(\frac{\partial}{\partial z^{\bar{\alpha}}}) = 0$, $Q(\frac{\partial}{\partial z^\alpha}) = 0$, $Q(\frac{\partial}{\partial z^{\bar{\alpha}}}) = \frac{\partial}{\partial z^{\bar{\alpha}}}$.

In particular, for $f = z^\alpha$ or $f = \bar{z}^\alpha$, we have

$$\partial z^\alpha = dz^\alpha, \qquad \bar{\partial} z^\alpha = 0, \qquad \partial \bar{z}^\alpha = 0, \qquad \bar{\partial}\, \bar{z}^\alpha = dz^{\bar{\alpha}} \ .$$

Then by [4] and the fact that $d^2 = 0$ we obtain

$$\partial dz^\alpha = -d\partial z^\alpha = 0, \qquad \bar{\partial} dz^\alpha = 0, \qquad \partial dz^{\bar{\alpha}} = 0, \qquad \bar{\partial} dz^{\bar{\alpha}} = 0 \ .$$

Next, using [2], we have

$$(17)\qquad \begin{aligned} \partial(dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge dz^{\overline{\beta}_1}\wedge\ldots\wedge dz^{\overline{\beta}_s}) &= 0\ , \\ \overline{\partial}(dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge dz^{\overline{\beta}_1}\wedge\ldots\wedge dz^{\overline{\beta}_s}) &= 0\ , \end{aligned}$$

for the basis elements of $A^{r,s}$. Any $\varphi \in A^{r,s}$ can be expressed in the form (9), and then by [1], [2], and (17),

$$\partial\varphi = \Sigma_{\substack{\alpha_1<\ldots<\alpha_r \\ \beta_1<\ldots<\beta_s}} \partial\varphi_{\alpha_1\ldots\alpha_r\,\overline{\beta}_1\ldots\overline{\beta}_s}\, dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge dz^{\overline{\beta}_1}\wedge\ldots\wedge dz^{\overline{\beta}_s}$$

and

$$\overline{\partial}\varphi = \Sigma_{\substack{\alpha_1<\ldots<\alpha_r \\ \beta_1<\ldots<\beta_s}} \overline{\partial}\varphi_{\alpha_1\ldots\alpha_r\,\overline{\beta}_1\ldots\overline{\beta}_s}\, dz^{\alpha_1}\wedge\ldots\wedge dz^{\alpha_r}\wedge dz^{\overline{\beta}_1}\wedge\ldots\wedge dz^{\overline{\beta}_s}$$

with $\partial\varphi_{\alpha_1\ldots\alpha_r\,\overline{\beta}_1\ldots\overline{\beta}_s} \in A^{1,0}$, $\overline{\partial}\varphi_{\alpha_1\ldots\alpha_r\,\overline{\beta}_1\ldots\overline{\beta}_s} \in A^{0,1}$. Then $\partial\varphi \in A^{r+1,s}$, $\overline{\partial}\varphi \in A^{r,s+1}$ or

$$\begin{aligned} &\partial\colon A^{r,s} \longrightarrow A^{r+1,s}\ , && 0 \leqq r < n\ , \quad 0 \leqq s \leqq n\ , \\ &\partial\colon A^{n,s} \longrightarrow 0\ , && 0 \leqq s \leqq n\ , \end{aligned}$$

and

$$\begin{aligned} &\overline{\partial}\colon A^{r,s} \longrightarrow A^{r,s+1}\ , && 0 \leqq r \leqq n\ , \quad 0 \leqq s < n\ , \\ &\overline{\partial}\colon A^{r,n} \longrightarrow 0\ , && 0 \leqq r \leqq n\ . \end{aligned}$$

An arbitrary $\varphi \in A$ is a sum of forms of type (r, s), $0 \leqq r, s \leqq n$, so $\partial\varphi$ and $\overline{\partial}\varphi$ can be computed by [1]. Consequently operators ∂ and $\overline{\partial}$ satisfying the axioms in Definition 5.12 exist and are uniquely determined by them, and are operators of types $(1, 0)$

and (0, 1) respectively.

Now let $D = d - (\partial + \bar{\partial})$. By [1], [2], and (14), [4], and $d^2 = 0$, we have

(1) $\quad D(\varphi + \psi) = D\varphi + D\psi$,

(2) $\quad D(\varphi \wedge \psi) = D\varphi \wedge \psi + (-1)^p \varphi \wedge D\psi$, $\quad \varphi \in A^p,\ \psi \in A$,

(3) $\quad Df = 0$, $\quad f \in A^0$,

(4) $\quad Ddf = -dDf = 0$, $\quad f \in A^0$.

It is clear that $D\varphi$, $\varphi \in A$, is uniquely determined by the above properties and that $D\varphi = 0$, i.e.

$$d\varphi = \partial\varphi + \bar{\partial}\varphi \ , \qquad \varphi \in A \ .$$

and this is (12).

Finally,

$$0 = d^2 = (\partial + \bar{\partial})^2 = \partial^2 + (\partial\bar{\partial} + \bar{\partial}\partial) + \bar{\partial}^2 \ .$$

On the right, the first operator is of type (2, 0), the second is of type (1, 1), and the last is of type (0, 2). If the sum is zero, each component must be zero, and this is (11).

§6. Holomorphic maps and holomorphic functions

Let W, $\tilde{W}$ be finite dimensional complex vector spaces. Any differentiable map $F: D \longrightarrow \tilde{D}$ induces correspondences F_*, F^* of the structures associated with D and $\tilde{D}$. For example, F_* is defined for complex tangent vectors by $F_* u = \tilde{u}$

where, if u is the complex tangent vector at $X \in D$ corresponding to $c_o \otimes v_o$, then $\tilde{u}$ is the complex tangent vector at $\tilde{X} = F(X) \in \tilde{D}$ which corresponds to $c_o \otimes F_{o*}v_o$. These induced maps are linear (with respect to complex coefficients) and commute with conjugation.

In particular, if $A = A(D)$ is the complex exterior algebra of differential forms associated with D and $\tilde{A} = A(\tilde{D})$ is the complex exterior algebra associated with $\tilde{D}$, then

$$F^*: \tilde{A} \longrightarrow A$$

is linear and satisfies

$$F^*S = SF^* \quad .$$

The above statements do not depend on the complex structures of W and $\tilde{W}$, expressed by the operators J for W and for $\tilde{W}$. The extension of J to A is obtained as follows: from cW_o to $(cW_o)^*$, by Exercise 4.8, to $\Lambda(cW_o)^*$ by Exercise 4.5, and then fibrewise to A. The resulting construction may be summarized as follows:

6.1. <u>Definition</u>. Let W be a finite dimensional complex vector space, and let $D \subset W$ be open. Let $A = A(D)$ be the complex exterior algebra of differential forms associated with D. The operator

$$J: A \longrightarrow A$$

is characterized by the properties:

(1) $J(\varphi + \psi) = J\varphi + J\psi$, $\varphi, \psi \in A$,

(2) $J(\varphi \wedge \psi) = J\varphi \wedge J\psi$, $\varphi, \psi \in A$,

(3) $Jf = f$, $f \in A^0$,

(4) for $\omega \in A^1$, $J\omega \in A^1$ is determined at each $X \in D$ by

$$< u, J\omega > = < Ju, \omega > , \qquad \text{for all } u \in (CT)_X .$$

6.2. Definition. A differentiable map $F: D \longrightarrow \tilde{D}$ is called holomorphic if and only if the induced correspondence $F^*: \tilde{A} \longrightarrow A$ commutes with J, i.e.

$$F^*J = JF^* .$$

A bidifferentiable holomorphic map is called biholomorphic.

Remarks. It is easily checked that, if F is biholomorphic, then F^{-1} is holomorphic. The qualification "bidifferentiable" in the above definition can be weakened to "bijective" since it can be shown that a bijective holomorphic map is bidifferentiable.

6.3. Proposition. For any differentiable map $F: D \longrightarrow \tilde{D}$, the following conditions are equivalent:

(5) $F^*J = JF^*$

(6) $F^*\Pi^{r,s} = \Pi^{r,s}F^*$ for all r, s ,

(7) $F^*\partial = \partial F^*$,

(8) $F^*\bar{\partial} = \bar{\partial}F^*$.

Proof. Formula (10) of §5 implies

$$J = \Sigma_{r,s} i^{r-s} \Pi^{r,s} ; \tag{9}$$

since F^* is linear, (6) and (9) imply (5). The equivalence of (7) and (8) follows from $F^*d = dF^*$ and $d = \partial + \bar{\partial}$. The fact that (5) implies (7) and (8) is easily seen from the formulas for ∂ and $\bar{\partial}$ of Exercise 8.4. To complete the proof, we show that (7) and (8) imply (6).

Let $(w^1,\ldots,w^k, \bar{w}^1,\ldots,\bar{w}^k)$ be a linear system of self-conjugate complex coordinates on $\tilde{D}$. Then the basis elements dw^α are of type (1, 0) and the basis elements $dw^{\bar{\alpha}}$ are of type (0, 1). Further, $\bar{\partial}w^\alpha = \partial w^{\bar{\alpha}} = 0$. Then (7) gives

$$F^*dw^\alpha = F^*\partial w^\alpha = \partial F^* w^\alpha , \qquad \alpha = 1,\ldots,k , \tag{10}$$

that is, F^*dw^α is of type (1, 0). Similarly, $F^*dw^{\bar{\alpha}}$ is of type (0, 1). Thus (7) and (8) imply that F^* preserves type for the basis elements dw^α, $dw^{\bar{\alpha}}$, from which (6) follows since F^* is linear and satisfies the complex analogue of XI, 3.7.

If a differentiable map $F: D \longrightarrow \tilde{D}$ is expressed by its component functions

$$w^\alpha = f^\alpha(X), \ \bar{w}^\alpha = \overline{f^\alpha(X)} = \bar{f}^\alpha(X) , \qquad X \in D , \tag{11}$$

then $f^\alpha = F^*w^\alpha$ and

$$\partial f^\alpha + \bar{\partial} f^\alpha = dF^*w^\alpha = F^*dw^\alpha = F^*\partial w^\alpha , \qquad \alpha = 1,\ldots,k .$$

Thus (10) is true, i.e. F is holomorphic, if and only if the

component functions satisfy

$$(12) \qquad \bar{\partial} f^{\alpha} = 0 , \qquad \alpha = 1,\ldots,k .$$

If $(z^1,\ldots,z^n, \bar{z}^1,\ldots,\bar{z}^n)$ is a linear system of self-conjugate complex coordinates on D, then (12) is equivalent to

$$(13) \qquad \frac{\partial f^{\alpha}}{\partial \bar{z}^{\beta}} = 0 , \qquad \alpha = 1,\ldots,k;\ \beta = 1,\ldots,n ,$$

by (16) of §5. For a holomorphic map F we write

$$w^{\alpha} = f^{\alpha}(X) = f^{\alpha}(z^1,\ldots,z^n) , \qquad \alpha = 1,\ldots,k ,$$

but, in the case of an arbitrary differentiable map F, we write

$$w^{\alpha} = f^{\alpha}(X) = f^{\alpha}(z^1,\ldots,z^n, \bar{z}^1,\ldots,\bar{z}^n) , \qquad \alpha = 1,\ldots,k ,$$

to express the fact that (13) does not hold. By Exercise 8.1, we have

$$\overline{\frac{\partial f^{\alpha}}{\partial z^{\beta}}} = \frac{\partial f^{\alpha}}{\partial \bar{z}^{\beta}} .$$

If F is holomorphic, these values are zero, by (13), and we write

$$\bar{w}^{\alpha} = \bar{f}^{\alpha}(\bar{z}^1,\ldots,\bar{z}^n) .$$

<u>Remark</u>. The above considerations can be used to show that for the case $D = W$, $\tilde{D} = \tilde{W}$, F_o linear, Definition 3.3 is equivalent to Definition 6.2. In particular, a linear transformation $F: W \longrightarrow \tilde{W}$ is holomorphic.

6.4. <u>Change of coordinates</u>. From Proposition 6.3, it follows that a bidifferentiable map $F: D \longrightarrow \tilde{D}$ preserves all

aspects of the complex structure if and only if it is biholomorphic. Consequently, only biholomorphic maps will be used to introduce new complex coordinates on D, as in Chapter XI, 3.13.

Remarks. An arbitrary bidifferentiable map $F: D \longrightarrow \tilde{D}$ will transport complex coordinates $(w^1,\ldots,w^n, \bar{w}^1,\ldots,\bar{w}^n)$ from $\tilde{D}$ to D but, if F is not holomorphic, the forms $dw^1,\ldots,dw^n$ would not give a basis for the forms of type $(1, 0)$ on D, etc.

Again, if $F_o: D_o \longrightarrow \tilde{D}_o$ is bidifferentiable and $\tilde{D}$ has a complex structure, there is a unique complex structure on D_o which makes $F: D \longrightarrow \tilde{D}$ a holomorphic map, viz. the structure on D_o determined by the automorphism $F_o^* J_o F_o^{*-1}$ of the complex exterior algebra associàted with D_o, where J_o expresses the complex structure on $\tilde{D}$.

More generally, an arbitrary homeomorphism $F_o: D_o \longrightarrow \tilde{D}_o$, where D_o is a topological space, serves to transport both the differentiable and complex structures of $\tilde{D}$ to D_o. For example, a function f on D_o is called differentiable, relative to the induced structure, if and only if fF^{-1} is differentiable. The complex tangent space $CT(D)$ is constructed as follows. If $\tilde{u}$ is a complex tangent vector at $F(X) \in \tilde{D}$, $X \in D$, then $u \in (CT)_X$ is the differentiation operator, on functions f on D, defined by

$$u \cdot f = \tilde{u} \cdot fF^{-1} ,$$

and $F_* u$ is $\tilde{u}$ by construction. A section $u: D_o \longrightarrow CT(D_o)$

is a complex vector field on D if and only if F_*u induces a complex vector field on $\tilde{D}$. A coordinate system on $\tilde{D}$ gives a coordinate system on D by assigning to a point $X \in D$ the coordinates of $F(X) \in \tilde{D}$, etc.

In the particular case $\tilde{D} = C$, $F: D \longrightarrow C$ is a complex-valued function on D. F may be expressed by the complex-valued component functions f, $\bar{f}$, or by a pair of real-valued functions

$$\operatorname{Re} f = \frac{1}{2}(f + \bar{f}) , \qquad \operatorname{Im} f = \frac{1}{2i}(f - \bar{f}) ,$$

called the real and imaginary parts of f. Then

$$f = \operatorname{Re} f + i\operatorname{Im} f , \qquad \bar{f} = \operatorname{Re} f - i\operatorname{Im} f .$$

In view of (12), an equivalent definition of "holomorphic" for the special case of functions is

6.5. <u>Definition</u>. A function $f \in A^0 = A^0(D)$ is called <u>holomorphic</u> if and only if $\bar{\partial} f = 0$ at each $X \in D$.

6.6. <u>Definition</u>. The equation

$$\bar{\partial} f = 0 , \qquad f \in A^0 , \tag{14}$$

is called the <u>Cauchy-Riemann equation</u>.

The equation (14) is equivalent to the system of partial differential equations

$$\frac{\partial f}{\partial \bar{z}^\alpha} = \frac{1}{2}\left(\frac{\partial f}{\partial x^\alpha} + i\frac{\partial f}{\partial y^\alpha}\right) = 0 , \qquad \alpha = 1,\ldots,n , \tag{15}$$

where $z^\alpha = x^\alpha + iy^\alpha$. The real and imaginary parts of (15) must vanish independently, so (15) is equivalent to the system

$$(16) \qquad \frac{\partial \mathrm{Re}\, f}{\partial x^\alpha} = \frac{\partial \mathrm{Im}\, f}{\partial y^\alpha}\,, \quad \frac{\partial \mathrm{Re}\, f}{\partial y^\alpha} = -\frac{\partial \mathrm{Im}\, f}{\partial x^\alpha}\,, \qquad \alpha = 1,\ldots,n\,,$$

of real equations to be satisfied by the real functions Re f, Im f if $f = \mathrm{Re}\, f + i\mathrm{Im}\, f$ is to be holomorphic. These equations assume a more complicated form if expressed in terms of arbitrary real coordinates on D.

6.7. <u>Proposition</u>. If D_0 is such that $H_1 = \tilde{\mathcal{O}}$ (Chapter XII), then a real function g on D is the real part of a holomorphic function on D if and only if $i\partial\bar{\partial}g = 0$ or, equivalently, $\partial\bar{\partial}g = 0$.

<u>Remark</u>. It is easily verified that $i\partial\bar{\partial}$ is a real operator of type (1, 1).

<u>Proof</u>. If there exists a holomorphic function f on D such that

$$g = \mathrm{Re}\, f = \tfrac{1}{2}(f + \bar{f})\,,$$

then $\bar{\partial}f = 0$, $\partial\bar{f} = 0$, and

$$\partial\bar{\partial}g = \partial(\tfrac{1}{2}\,\bar{\partial}\bar{f}) = -\tfrac{1}{2}\,\bar{\partial}\partial\bar{f} = 0$$

using (11) of §5. Conversely, suppose that g satisfies $\partial\bar{\partial}g = 0$. Then ∂g is closed, since $d(\partial g) = \bar{\partial}\partial g = -\partial\bar{\partial}g = 0$. Moreover,

$$2\partial g = (\partial g + \bar{\partial}g) + i\left(\frac{\partial g - \bar{\partial}g}{i}\right)$$

gives a decomposition in which the 1-forms $\partial g + \bar{\partial} g$ and $(\partial g - \bar{\partial} g)/i$ are real and closed. Then if D is connected and X_0 is a fixed point of D, the integral

$$f(X) = 2\int_{X_0}^{X} \partial g + g(X_0)$$

will be single-valued in D, by Stokes' formula (Theorem XII, 4.4), using the hypothesis $H_1 = \tilde{0}$. That is, $f \in A^0 = A^0(D)$. Moreover, $df = 2\partial g$ so $\partial f = 2\partial g$, $\bar{\partial} f = 0$. In particular, f is holomorphic. Also, $d\bar{f} = 2\bar{\partial} g$ (since g is real). Then

$$d\mathrm{Re}\, f = \frac{1}{2}(df + d\bar{f}) = \partial g + \bar{\partial} g = dg \quad .$$

Since $(\mathrm{Re}\, f)(X_0) = g(X_0)$, we conclude that $g = \mathrm{Re}\, f$.

If D is not connected, the construction can be carried out in each connected component of D.

6.8. <u>Proposition</u>. Let D_0 be such that $H_1 = \tilde{0}$ and let $f \in A^0$ satisfy $\partial\bar{\partial} f = 0$ in D. Then

$$f = F + \bar{G} \ ,$$

where F and G are holomorphic functions in D.

<u>Proof</u>. Since $i\partial\bar{\partial}$ is a real operator, $\partial\bar{\partial} f = 0$ implies

$$\partial\bar{\partial}\mathrm{Re}\, f = 0 \ , \qquad \partial\bar{\partial}\mathrm{Im}\, f = 0 \quad .$$

By Proposition 6.7, $\mathrm{Re}\, f$ is the real part of a holomorphic function F_1 and $\mathrm{Im}\, f$ is the real part of a holomorphic function G_1. Then for

$$F = \frac{1}{2}(F_1 + iG_1) , \qquad G = \frac{1}{2}(F_1 - iG_1) ,$$

we have

$$F + \overline{G} = \text{Re } F_1 + i\text{Re } G_1 = \text{Re } f + i\text{Im } f = f \quad .$$

6.9. Definition. A differential form $\varphi \in A^p = A^p(D)$ is called holomorphic in D if and only if φ is of type (p, o) and $\overline{\partial}\varphi = 0$.

6.10. Proposition. If $\varphi \in A^{p,o}$ is holomorphic, then each coefficient in a representation of φ in terms of complex coordinates is a holomorphic function.

The proof is left as an exercise.

6.11. Definition. A differentiable 2p-chain c_{2p} on D will be said to be of type $(n-p, n-p)$ if

$$\int_{c_{2p}} \varphi = 0$$

for every $\varphi \in A^{2p}$ satisfying $\Pi^{p,p} \varphi = 0$.

Remarks. If c_{2p} is of type $(n-p, n-p)$, then

$$\int_{c_{2p}} \Pi^{r,s} \varphi = 0 \qquad \text{unless } r = p, s = p ,$$

and, in general,

$$\int_{c_{2p}} \varphi = \int_{c_{2p}} \Pi^{p,p} \varphi \quad .$$

In the case $p = n$, every finite 2n-chain is of type (o, o).

6.12. Lemma. Let $\sigma_{2p} = (s_{2p}, F)$ be a (differentiable) 2p-simplex on D. If s_{2p} is the standard 2p-simplex of $C_o{}^p$ and if F is a holomorphic map, then σ_{2p} is of type $(n-p, n-p)$.

The proof is left as an exercise.

6.13. <u>Proposition</u>. Let $\varphi \in A^{p,p-1} = A^{p,p-1}(D)$. Then $\bar{\partial}\varphi = 0$ if and only if

$$\int_{c_{2p-1}} \varphi = 0 \tag{17}$$

for every $(2p-1)$-cycle c_{2p-1} in D such that $c_{2p-1} = \partial c_{2p}$ with c_{2p} of type $(n-p, n-p)$.

<u>Proof</u>. If $c_{2p-1} = \partial c_{2p}$ where c_{2p} is of type $(n-p, n-p)$ then, by Stokes' formula,

$$\int_{c_{2p-1}} \varphi = \int_{c_{2p}} d\varphi = \int_{c_{2p}} \bar{\partial}\varphi ,$$

since $\Pi^{p,p} d\varphi = \bar{\partial}\varphi$ if $\varphi \in A^{p,p-1}$, and is zero if $\bar{\partial}\varphi = 0$.

Conversely, assume that the integral of φ over c_{2p-1} vanishes for every $c_{2p-1} = \partial\sigma_{2p}$ where $\sigma_{2p} = (s_{2p}, F)$, s_{2p} is the standard simplex of $C_o^{\,p}$, and F is a holomorphic affine map; that is, the component functions of F are of the form

$$z^\alpha = c_1^{\,\alpha} t^1 + \ldots + c_p^{\,\alpha} t^p + c^\alpha , \qquad \alpha = 1,\ldots,n , \tag{18}$$

where the $c_j^{\,\alpha}$, c^α are complex constants. Then σ_{2p} is of type $(n-p, n-p)$ by Lemma 6.12, and

$$0 = \int_{c_{2p-1}} \varphi = \int_{s_{2p}} F^* \bar{\partial}\varphi \quad .$$

Given $X_o \in D$ and an arbitrary open set U in D containing X_o, we can choose a holomorphic affine map F such that the support of σ_{2p} is contained in U. Since U can be taken arbitrarily small, we conclude that $F^*\bar{\partial}\varphi = 0$ at X_o, i.e.

$$(19)\quad \sum_{\substack{\alpha_1<\ldots<\alpha_p\\ \beta_1<\ldots<\beta_p}} (\bar\partial\varphi)_{\alpha_1\ldots\alpha_p\,\bar\beta_1\ldots\bar\beta_p}\,\frac{\partial(z^{\alpha_1},\ldots,z^{\alpha_p})}{\partial(t^1,\ldots,t^p)}\,\overline{\frac{\partial(z^{\beta_1},\ldots,z^{\beta_p})}{\partial(t^1,\ldots,t^p)}} = 0$$

at X_o. The Jacobian determinant in the formula analogous to (12) of XI, 3.6 here splits into two factors because F is holomorphic. The entries in each determinant are constants if F is affine. Then (19) holds for all holomorphic affine F such that the c_j^{α} in (18) are sufficiently small. That this implies $(\bar\partial\varphi)_{\alpha_1\ldots\alpha_p\,\bar\beta_1\ldots\bar\beta_p} = 0$ at X_o, for each choice of indices, will follow from Lemma 6.14 below. Since X_o is an arbitrary point of D, we conclude that $\bar\partial\varphi$ vanishes in D.

6.14. <u>Lemma</u>. If the complex numbers $a_{\alpha_1\ldots\alpha_p\,\bar\beta_1\ldots\bar\beta_p}$, $\alpha_1<\ldots<\alpha_p \leqq n$, $\beta_1<\ldots<\beta_p \leqq n$, satisfy

$$(20)\quad \sum_{\substack{\alpha_1<\ldots<\alpha_p\\ \beta_1<\ldots<\beta_p}} a_{\alpha_1\ldots\alpha_p\,\bar\beta_1\ldots\bar\beta_p}\,\begin{vmatrix} c_1^{\alpha_1} & \ldots & c_p^{\alpha_1}\\ & \ldots & \\ c_1^{\alpha_p} & \ldots & c_p^{\alpha_p}\end{vmatrix}\cdot\begin{vmatrix} \bar c_1^{\beta_1} & \ldots & \bar c_p^{\beta_1}\\ & \ldots & \\ \bar c_1^{\beta_p} & \ldots & \bar c_p^{\beta_p}\end{vmatrix} = 0$$

for all choices of the complex numbers c_j^{α}, $\alpha = 1,\ldots,n$, $j = 1,\ldots,p$, such that $|c_j^{\alpha}| < \varepsilon$, where $\varepsilon > 0$, then each $a_{\alpha_1\ldots\alpha_p\,\bar\beta_1\ldots\bar\beta_p}$ is zero.

<u>Proof</u>. Let $\alpha_1<\ldots<\alpha_p$, $\beta_1<\ldots<\beta_p$ be a given selection of indices, and suppose that q of the β's are different from the α's. (If $p = n$, only the case $q = 0$ can occur; if $p = 1 < n$, then q may be 0 or 1, etc). Consider the special choices of the c_j^{α} such that

$$c_k^{\alpha_j} = 0 \qquad \text{for } k \neq j,\ j = 1,\ldots,p\ ,$$
$$c_k^{\beta_\ell} = 0 \qquad \text{for } k \neq \ell,\ \beta_\ell \neq \alpha_1,\ldots,\alpha_p\ ,$$
$$c_k^{\gamma} = 0\ , \qquad \gamma \neq \alpha_1,\ldots,\alpha_p,\ \beta_1,\ldots,\beta_p\ ,$$

and the remaining c_j^{α} are different from zero. Then many of the terms in (20) vanish. For example, suppose $q = 2$ and $\beta_1 < \beta_2 < \alpha_1 < \alpha_2 < \ldots < \alpha_p$, with $\beta_j = \alpha_j$, $j = 3,\ldots,p$. If we divide by $c_3^{\alpha_3}\ldots c_p^{\alpha_p}\,\bar{c}_3^{\alpha_3}\ldots\bar{c}_p^{\alpha_p} \neq 0$, then (20) gives

$$\begin{aligned}
0 \equiv\ & a_{\alpha_1\alpha_2\alpha_3\ldots\alpha_p\,\bar{\beta}_1\bar{\beta}_2\bar{\alpha}_3\ldots\bar{\alpha}_p}\, c_1^{\alpha_1}\, c_2^{\alpha_2}\, \bar{c}_1^{\beta_1}\, \bar{c}_2^{\beta_2} \\
& - a_{\beta_1\alpha_2\alpha_3\ldots\alpha_p\,\bar{\beta}_2\bar{\alpha}_1\bar{\alpha}_3\ldots\bar{\alpha}_p}\, c_1^{\beta_1}\, c_2^{\alpha_2}\, \bar{c}_2^{\beta_2}\, \bar{c}_1^{\alpha_1} \\
& - a_{\beta_2\alpha_1\alpha_3\ldots\alpha_p\,\bar{\beta}_1\bar{\alpha}_2\bar{\alpha}_3\ldots\bar{\alpha}_p}\, c_2^{\beta_2}\, c_1^{\alpha_1}\, \bar{c}_1^{\beta_1}\, \bar{c}_2^{\alpha_2} \\
& + a_{\beta_1\beta_2\alpha_3\ldots\alpha_p\,\bar{\alpha}_1\bar{\alpha}_2\bar{\alpha}_3\ldots\bar{\alpha}_p}\, c_1^{\beta_1}\, c_2^{\beta_2}\, \bar{c}_1^{\alpha_1}\, \bar{c}_2^{\alpha_2}\ ,
\end{aligned}$$

or

$$Axy\overline{zw} + Bzy\overline{wx} + Cwx\overline{zy} + Dzw\overline{xy} \equiv 0\ ,$$

where $x \neq 0$, $y \neq 0$, $z \neq 0$, $w \neq 0$. Now,

$$\frac{\partial\bar{x}}{\partial x} = \frac{1}{2}\left(\frac{\partial(\mathrm{Re}\ x - i\,\mathrm{Im}\ x)}{\partial \mathrm{Re}\ x} - i\,\frac{\partial(\mathrm{Re}\ x - i\,\mathrm{Im}\ x)}{\partial \mathrm{Im}\ x}\right) = 0\ ,$$

etc., so by q differentiations we obtain

$$\ldots,\ Ay\overline{zw} + Cw\overline{zy} \equiv 0\ , \qquad A\overline{zw} \equiv 0\ .$$

Since $zw = c_1^{\beta_1} c_2^{\beta_2} \neq 0$, this implies that

$$A = a_{\alpha_1 \ldots \alpha_p \overline{\beta}_1 \overline{\beta}_2 \overline{\alpha}_3 \ldots \overline{\alpha}_p} = 0\ .$$

6.15. <u>Proposition</u> (Cauchy's Theorem). Let $\varphi \in A^{1,0} = A^{1,0}(D)$. Then

(i) φ is holomorphic if and only if

(21) $$\int_{c_1} \varphi = 0$$

for every 1-cycle c_1 in D such that $c_1 = \partial c_2$ with c_2 of type $(n-1, n-1)$,

(ii) φ is a closed holomorphic 1-form (i.e. $\partial\varphi = 0$, $\overline{\partial}\varphi = 0$) if and only if (21) is satisfied for every 1-cycle c_1 in D which is homologous to zero.

Part (i) is the special case $p = 1$ of Proposition 6.13. The proof of (ii) is similar to that of Proposition 6.13, but easier. It is left as an exercise.

<u>Remarks</u>. In the case of complex dimension $n = 1$, every c_2 is of type $(0, 0)$; then the statements (i) and (ii) of Proposition 6.15 coincide (since $\partial\varphi$ is automatically zero if $n = 1$) and constitute the classical theorem of Cauchy: a complex-valued function f on an open set $D \subset C$ is holomorphic in D if and only if the integral of $\varphi = f dz$ around any closed curve c_1 in D vanishes [where c_1 is homologous to zero]. The

qualification in brackets is often omitted in textbooks, but the theorem is incorrect without it. For example, take D to be the complex plane C with the origin removed, and take $f = 1/z$. Then f is holomorphic in D, i.e. $\bar{\partial} f = 0$, but the integral of f around the unit circumference does not vanish!

Proposition 6.15 remains valid under the weaker assumption that f is continuous (class C^0) in D. In the case of complex dimension $n = 1$, this stronger theorem is often called the Cauchy-Goursat Theorem.

§7. Poincaré Lemma

We now consider the complex analogues of Propositions XI, 4.12 and XI, 4.14.

7.1. Definition. The unit polycylinder in n-dimensional complex cartesian space C^n is the set of points $(a^1,\ldots,a^n)$ such that $a^\alpha \bar{a}^\alpha < 1$, $\alpha = 1,\ldots,n$. If $D \subset W$ is open, with complex coordinates $(z^1,\ldots,z^n, \bar{z}^1,\ldots,\bar{z}^n)$ on D, then the subset of D of points for which $z^\alpha \bar{z}^\alpha < 1$, $\alpha = 1,\ldots,n$, is called a unit (coordinate) polycylinder in D.

7.2. Definition. Let $U \subset D$ be the unit coordinate polycylinder corresponding to a particular choice of complex coordinates $(z^1,\ldots,z^n, \bar{z}^1,\ldots,\bar{z}^n)$ in D. We denote by $A_b^0 = A_b^0(U)$ the subspace of $A^0(U)$ consisting of functions f such that f and all partial derivatives of f of the form $\partial^q f/\partial \bar{z}^{\gamma_1}\ldots\partial \bar{z}^{\gamma_q}$, $q = 1, 2,\ldots,$ are bounded in U. We denote by $A_b = A_b(U)$ the subalgebra of the complex exterior algebra $A(U)$

associated with U, consisting of those differential forms which have coefficients in A_b^o when expressed in terms of the basis for $A(U)$ associated with the given complex coordinates.

It is clear that $\bar{\partial}: A_b \longrightarrow A_b$.

7.3. Proposition. Let U be a unit coordinate polycylinder and let $A_b(U)$ be the algebra defined in Definition 7.2. There exists a linear operator

$$\kappa: A_b \longrightarrow A_b ,$$

of type $(0, -1)$ such that

$$(1) \qquad \kappa\bar{\partial}\varphi + \bar{\partial}\kappa\varphi = \varphi , \qquad \varphi \in A_b^{r,s},\ s > 0 ,$$

$$(2) \qquad \kappa\bar{\partial}\varphi = \varphi - h\varphi , \qquad \varphi \in A_b^{r,o} ,$$

where $h: A_b^{r,o} \longrightarrow A_b^{r,o}$ is a projection of $A_b^{r,o}$ onto the linear subspace of the holomorphic forms in $A_b^{r,o}$.

The complete proof of this proposition will not be given. An operator κ of type $(0, -1)$ having the properties stated in Proposition 7.3 can be defined, for forms in $A^{r,s}$ with $s > 0$, in terms of the operators

$$T^{\bar{\alpha}}: A_b^o \longrightarrow A_b^o , \qquad \alpha = 1, \ldots, n ,$$

defined by

$$T^{\bar{\alpha}}f(z, \bar{z})$$

$$= \frac{1}{2\pi i} \int_{|\tau|<1} f(z^1, \bar{z}^1, \ldots, z^{\alpha-1}, \bar{z}^{\alpha-1}, \tau, \bar{\tau}, z^{\alpha+1}, \bar{z}^{\alpha+1}, \ldots, z^n, \bar{z}^n) \frac{d\tau d\bar{\tau}}{\tau - z^\alpha} .$$

These operators satisfy

$$\frac{\partial}{\partial \bar{z}^\beta}(T^{\bar{\alpha}} f) = \begin{cases} T^{\bar{\alpha}} \dfrac{\partial f}{\partial \bar{z}^\beta} & \text{for } \alpha \neq \beta , \\ f & \text{for } \alpha = \beta , \end{cases}$$

(Exercise 11.4). For example, if $\varphi \in A_b^{0,1}$, say

$$\varphi = \Sigma_{\alpha=1}^n \varphi_{\bar{\alpha}} \, dz^{\bar{\alpha}} ,$$

then $\kappa\varphi \in A_b^0$ is defined by

$$(3) \quad \kappa\varphi = \Sigma_{k=0}^{n-1} \frac{(-1)^k}{(k+1)!} \Sigma_{\alpha=1}^n \Sigma_{\substack{\beta_1,\ldots,\beta_k \\ \beta_\ell \neq \alpha}} T^{\bar{\beta}_1} \frac{\partial}{\partial \bar{z}^{\beta_1}} \cdots T^{\bar{\beta}_k} \frac{\partial}{\partial \bar{z}^{\beta_k}} T^{\bar{\alpha}} \varphi_{\bar{\alpha}} .$$

To prove (2) in the case $r = 0$, for example, we define the operator h by the equation

$$hf = f - \kappa\bar{\partial} f , \qquad f \in A_b^{0,0} ,$$

where $\kappa\bar{\partial} f$ is given by (3) with $\varphi_{\bar{\alpha}} = \partial f/\partial \bar{z}^\alpha$. We then verify that $\partial(\kappa\bar{\partial} f)/\partial \bar{z}^\beta = \partial f/\partial \bar{z}^\beta$. Thus $\partial(hf)/\partial \bar{z}^\beta = 0$, $\beta = 1,\ldots,n$; that is, hf is a holomorphic function.

The analogue of Proposition XI, 4.14 is

7.4. <u>Proposition</u> (Poincaré Lemma for $\bar{\partial}$). If $D \subset W$ is an open set and $X_1 \in D$, then there is an open set D_1 containing X_1 and contained in D such that, if $\varphi \in A^{r,s}(D)$, $s > 0$, satisfies $\bar{\partial}\varphi = 0$, then there is a differential form $\eta \in A^{r,s-1}(D_1)$ satisfying $\bar{\partial}\eta = \varphi$ in D_1.

<u>Proof</u>. A suitable change of coordinates by means of a holomorphic affine map will give a unit coordinate polycylinder U

such that $X_1 \in U$ and $\overline{U} \subset D$. We take $D_1 = U$. The restriction of φ to $D_1 = U$ gives a form in $A_b^{r,s}(D_1)$ since φ is differentiable on D and $\overline{U}$ is a compact subset of D. We then take $\eta = \kappa\varphi \in A_b^{r,s-1}(D_1)$, where κ is an operator having the properties described in Proposition 7.3. Since $\overline{\partial}\varphi = 0$, (1) gives $\overline{\partial}\eta = \varphi$ on D_1.

7.5. <u>Definition</u>. Let D be an open set in a complex vector space. A differential form $\varphi \in A = A(D)$ is said to be <u>$\overline{\partial}$-closed</u> if $\varphi \in \ker \overline{\partial}$, i.e. if $\overline{\partial}\varphi = 0$; φ is said to be <u>$\overline{\partial}$-exact</u> if $\varphi \in \operatorname{im} \overline{\partial}$, i.e. if there exists a differential form $\eta \in A$ such that $\varphi = \overline{\partial}\eta$. The (complex) vector space of $\overline{\partial}$-closed differential forms of type (r, s) will be denoted by $Z^{r,s} = Z^{r,s}(D)$; the (complex) vector space of $\overline{\partial}$-exact differential forms of type (r, s) will be denoted by $B^{r,s} = B^{r,s}(D)$.

In particular, $Z^{r,o}$ consists of the holomorphic r-forms, $B^{r,o} = \tilde{0}$, and $Z^{r,n} = A^{r,n}$.

Since $\overline{\partial}^2 = 0$, $B^{r,s} \subset Z^{r,s}$ and, as in XI, §4 and XII, §6, we have

$$\tilde{0} \longrightarrow B^{r,s} \xrightarrow{\iota} Z^{r,s} \longrightarrow H_{\overline{\partial}}^{r,s} \longrightarrow \tilde{0}$$

where

$$H_{\overline{\partial}}^{r,s} = H_{\overline{\partial}}^{r,s}(D) = Z^{r,s}/B^{r,s} \quad .$$

7.6. <u>Definition</u>. The complex vector space $H_{\overline{\partial}}^{r,s} = H_{\overline{\partial}}^{r,s}(D)$ is called the <u>$\overline{\partial}$-cohomology</u> of D of type (r, s) (or, sometimes, the $\overline{\partial}$-cohomology of $A = A(D)$ of type (r, s)).

In Chapter XII we saw that the d-cohomology reflects the "geometric structure" of $D \subset V$ with V a real vector space. The $\bar{\partial}$-cohomology, on the other hand, reflects the "complex structure" of $D \subset W$.

7.7. Definition. If $D \subset W$, $D \neq W$, and $H^{0,s}_{\bar{\partial}}(D) = \tilde{0}$ for $s > 0$, then D is called a domain of holomorphy.

The condition of Definition 7.7 can be shown to be equivalent to the statement that there exists at least one holomorphic function in D which is not the restriction to D of any holomorphic function in D', for any $D' \subset W$ with $D \subset D'$, $D \neq D'$.

If $\dim_{\mathbb{C}} W = 1$, every D is a domain of holomorphy but, if $\dim_{\mathbb{C}} W > 1$, only domains D of very special type have this property. For example, a unit coordinate polycylinder or a unit coordinate ball $(\Sigma_{\alpha=1}^{n} z^{\alpha}\bar{z}^{\alpha} < 1)$ are domains of holomorphy. However, the image of either of these domains under a homeomorphism of W onto itself is, in general, not a domain of holomorphy. This shows that the $\bar{\partial}$-coholomogy is not determined by the d-cohomology, i.e. by the ordinary geometrical structure.

Remark. Proposition 7.3 does not immediately imply that $H^{r,s}_{\bar{\partial}}(U) = \tilde{0}$ for $s > 0$, where U is a unit coordinate polycylinder, but only $H^{r,s}_{b,\bar{\partial}} = \tilde{0}$, where $H^{r,s}_{b,\bar{\partial}} = Z^{r,s}_{b}/B^{r,s}_{b}$ and $Z^{r,s}_{b}$ is the space $\bar{\partial}$-closed forms in $A^{r,s}_{b}$, $B^{r,s}_{b}$ the space of $\bar{\partial}$-exact forms in $A^{r,s}_{b}$. In order to conclude, for example, that $H^{0,s}_{\bar{\partial}} = \tilde{0}$, $s > 0$, given that $H^{0,s}_{b,\bar{\partial}} = \tilde{0}$, $s > 0$, it is necessary to show that for any $\varphi \in Z^{0,s}$, $s > 0$, there exists a $\zeta \in A^{0,s-1}$

such that $\varphi - \bar{\partial}\zeta \in Z_b^{o,s}$.

§8. Exercises

1. Let u be a complex tangent vector or complex vector field, and let $\bar{u}$ be its conjugate. Show that, as differentiation operators on functions, u and $\bar{u}$ are conjugate operators, i.e. show that

$$\overline{u \cdot f} = \bar{u} \cdot \bar{f} , \qquad u \cdot \bar{f} = \overline{\bar{u} \cdot f} ,$$

for all complex-valued functions f.

2. Give the analogues, for ∂ and $\bar{\partial}$, of formula (11) of XI, §4.

3. Characterize $S: A \longrightarrow A$ by its properties as an operator on A, in analogy with those given for $J: A \longrightarrow A$ in Definition 6.1. Do the same for the function $F^*: \tilde{A} \longrightarrow A$ defined at the beginning of §6.

4. Show that

$$J\partial = i\partial J , \qquad J\bar{\partial} = - i\bar{\partial} J ,$$

and that

$$\partial = \tfrac{1}{2}(d + iJ^{-1}dJ) , \qquad \bar{\partial} = \tfrac{1}{2}(d - iJ^{-1}dJ) .$$

5. Using (13) of §5, show that (6) of Proposition 6.3 implies (7) and (8) of the same proposition.

6. Show that a holomorphic transformation of coordinates preserves orientation.

7. Prove Lemma 6.12.

8. Prove part (ii) of Proposition 6.15.

9. Prove Cauchy's Theorem (Proposition 6.15) under the weaker assumption that φ is continuous.

§9. Hermitian and Kähler metrics

Let W_o be a finite dimensional vector space with (homogeneous) complex structure defined by an automorphism J_o satisfying $J_o^2 = -I_o$, and let W be the complex vector space lying over W_o (Theorem 3.14). Let $D \subset W$ be open.

9.1. Definition. A riemannian metric (Definition XI, 5.1) on D_o is said to be hermitian if, for any $X \in D_o$, and any (real) tangent vector $u \in T_X$, u and $J_o u$ have the same length, i.e. $|u| = |J_o u|$. The metric induced on D is called an hermitian metric on D.

In fact (cf. Exercises 4.9 - 4.12), the real scalar product in T_X induces an hermitian scalar product in $(CT)_X$, relative to which $J: (CT)_X \longrightarrow (CT)_X$ is unitary and the decomposition $(CT)_X = P((CT)_X) \oplus Q((CT)_X)$ is orthogonal. Further,

$$(1) \qquad u \cdot Jv = - Ju \cdot v , \qquad u, v \in (CT)_X.$$

The extension of the metric to forms is defined, at any $X \in D$, by

$$\varphi \cdot \psi = c\, \overline{c'}(\eta \cdot \zeta)$$

if $\varphi = c\eta$, $\psi = c'\zeta$ and η, ζ are real. Note that an arbitrary form φ can be expressed as

$$\varphi = \frac{\varphi + \overline{\varphi}}{2} + i\left(\frac{\varphi - \overline{\varphi}}{2i}\right) = \mathrm{Re}\ \varphi + i\mathrm{Im}\ \varphi \ ,$$

and we have

$$\varphi\cdot\psi = \mathrm{Re}\ \varphi\cdot\mathrm{Re}\ \psi + \mathrm{Im}\ \varphi\cdot\mathrm{Im}\ \psi + i(\mathrm{Im}\ \varphi\cdot\mathrm{Re}\ \psi - \mathrm{Re}\ \varphi\cdot\mathrm{Im}\ \psi) \quad .$$

Since the fundamental decomposition of $(CT)_X$ is orthogonal, it follows that the induced decomposition into types (Definition 5.9) is also orthogonal, that is,

$$(2) \qquad \Pi^{r,s}\varphi\cdot\Pi^{r',s'}\psi = 0 \ , \qquad (r,\ s) \neq (r',\ s') \quad .$$

Similarly, the extension of J to forms is unitary:

$$(3) \qquad J\varphi\cdot J\psi = \varphi\cdot\psi \quad .$$

Since an hermitian scalar product is conjugate linear, rather than linear, in the second argument, the complex analogue of formula (1) of XI, §5 does not give a (complex) covariant tensor of order 2. Instead:

9.2. <u>Definition</u>. The <u>fundamental form</u> θ of the hermitian metric on D is the differential form of degree 2 defined by

$$(4) \qquad \langle u\wedge v,\theta\rangle = Ju\cdot\overline{v} \ , \qquad u,\ v \in (CT)_X \quad .$$

It is left as an exercise to verify that (4) does indeed define a differential form (use (1)), and that $u\cdot v = \langle u\wedge J\overline{v},\theta\rangle$. Further, θ is real, that is, $\overline{\theta} = \theta$, since

$$< u \wedge v, \overline{\theta} > = \overline{< \overline{u} \wedge \overline{v}, \theta >} = \overline{Ju \cdot v} = Ju \cdot \overline{v} = < u \wedge v, \theta > ,$$

and θ is of type $(1, 1)$. In fact, if $u, v \in P((CT)_X)$, then $Ju \in P((CT)_X)$ and $\overline{v} \in Q((CT)_X)$, so $Ju \cdot \overline{v} = 0$; the same conclusion holds if $u, v \in Q((CT)_X)$. Thus $\Pi^{2,0}\theta = \Pi^{0,2}\theta = 0$. The expression of θ in terms of complex coordinates $(z^1, \ldots, z^n, \overline{z}^1, \ldots, \overline{z}^n)$ on D is

$$(4') \qquad \theta = i \, \Sigma_{1 \leq \alpha, \beta \leq n} \, \lambda_{\alpha\overline{\beta}} \, dz^\alpha \wedge dz^{\overline{\beta}} ,$$

where

$$\lambda_{\alpha\overline{\beta}}(z, \overline{z}) = \frac{\partial}{\partial z^\alpha} \cdot \frac{\partial}{\partial z^\beta} = \overline{\frac{\partial}{\partial z^\beta} \cdot \frac{\partial}{\partial z^\alpha}} = \overline{\lambda_{\beta\overline{\alpha}}} .$$

If θ^n denotes the exterior product of θ with itself, n times, then

$$(5) \qquad \theta^n = i^n n! \det(\lambda) \, dz^1 \wedge dz^{\overline{1}} \wedge \ldots \wedge dz^n \wedge dz^{\overline{n}} ,$$

where

$$\det(\lambda) = \left(\frac{\partial}{\partial z^1} \wedge \ldots \wedge \frac{\partial}{\partial z^n}\right) \cdot \left(\frac{\partial}{\partial z^{\overline{1}}} \wedge \ldots \wedge \frac{\partial}{\partial z^{\overline{n}}}\right)$$

is real and positive: $\det(\lambda) = \overline{\det(\lambda)} > 0$.

Let

$$x^\alpha = \frac{1}{2}(z^\alpha + \overline{z}^\alpha) , \qquad y^\alpha = \frac{1}{2i}(z^\alpha - \overline{z}^\alpha)$$

be the corresponding real coordinates on D_o, so that $\frac{\partial}{\partial x^\alpha}$, $\frac{\partial}{\partial y^\alpha} = J_o \frac{\partial}{\partial x^\alpha}$ form a real basis for T_X and a complex basis for $(CT)_X$, $X \in D$. The fundamental tensor γ of any

riemannian metric on D_o can be expressed in the form (tensor product over R)

$$\gamma = \Sigma_{1\leq\alpha,\beta\leq n}(g_{\alpha\beta}dx^\alpha \otimes dx^\beta + g_{\alpha\hat{\beta}}dx^\alpha \otimes dy^\beta + g_{\hat{\alpha}\beta}dy^\alpha \otimes dx^\beta + g_{\hat{\alpha}\hat{\beta}}dy^\alpha \otimes dy^\beta)$$

where the coefficients are real and symmetric:

$$g_{\alpha\beta} = g_{\beta\alpha}\ , \quad g_{\alpha\hat{\beta}} = g_{\hat{\beta}\alpha}\ , \quad g_{\hat{\alpha}\hat{\beta}} = g_{\hat{\beta}\hat{\alpha}}\ .$$

If the riemannian metric is hermitian, then

$$g_{\alpha\beta} = g_{\hat{\alpha}\hat{\beta}}, \quad g_{\alpha\hat{\beta}} = -\,g_{\hat{\alpha}\beta}\ .$$

If the hermitian metric θ on D is induced by γ, then (Exercise 4.9)

$$\lambda_{\alpha\bar{\beta}} = \frac{\partial}{\partial z^\alpha}\cdot\frac{\partial}{\partial z^\beta} = \frac{1}{2}\left(\frac{\partial}{\partial x^\alpha} - i\,\frac{\partial}{\partial y^\alpha}\right)\cdot\frac{1}{2}\left(\frac{\partial}{\partial x^\beta} - i\,\frac{\partial}{\partial y^\beta}\right)$$

$$= \frac{1}{4}(g_{\alpha\beta} + ig_{\alpha\hat{\beta}} - ig_{\hat{\alpha}\beta} + g_{\hat{\alpha}\hat{\beta}}) \tag{6}$$

$$= \frac{1}{2}(g_{\alpha\beta} + ig_{\alpha\hat{\beta}})\ .$$

Therefore

$$g = \det\begin{pmatrix} (g_{\alpha\beta}) & (g_{\alpha\hat{\beta}}) \\ (g_{\alpha\hat{\beta}}) & (g_{\hat{\alpha}\hat{\beta}}) \end{pmatrix} = \det\begin{pmatrix} (g_{\alpha\beta}) & -(g_{\alpha\hat{\beta}}) \\ (g_{\alpha\hat{\beta}}) & (g_{\alpha\beta}) \end{pmatrix}\ .$$

Then (as in the proof of Proposition 3.6)

$$g = \det \begin{pmatrix} (g_{\alpha\beta} + ig_{\alpha\hat{\beta}}) & 0 \\ 0 & (g_{\alpha\beta} - ig_{\alpha\hat{\beta}}) \end{pmatrix} = \det \begin{pmatrix} (2\lambda_{\alpha\bar{\beta}}) & 0 \\ 0 & (2\overline{\lambda_{\alpha\bar{\beta}}}) \end{pmatrix}$$

or

$$g = 2^{2n}\det(\lambda)\,\overline{\det(\lambda)} = 2^{2n}(\det(\lambda))^2 \quad . \tag{7}$$

Now the complex structure determines a natural orientation in the underlying real vector space and therefore determines the choice of the real 2n-form of unit length in $\Lambda^{2n}T_X^*$ for each $X \in D$, and therefore a real form Θ in $A^{2n} = A^{2n}(D)$. Since $x^1, y^1, \ldots, x^n, y^n$ is a positively oriented coordinate system and since $dx^\alpha \wedge dy^\alpha = \frac{i}{2}(dz^\alpha \wedge d\bar{z}^\alpha)$, we have

$$\begin{aligned} \Theta &= \sqrt{g}\, dx^1 \wedge dy^1 \wedge \ldots \wedge dx^n \wedge dy^n \\ &= (\tfrac{i}{2})^n \sqrt{g}\, dz^1 \wedge d\bar{z}^1 \wedge \ldots \wedge dz^n \wedge d\bar{z}^n \\ &= \frac{\theta^n}{n!} , \end{aligned} \tag{8}$$

using (5) and (7).

<u>Example</u>. Let $W = C^n$, $W_o = R^{2n}$, so that $\mathbb{C}W_o = C^{2n}$. The euclidean metric on R^{2n} is obviously hermitian in the sense of Definition 9.1, and the hermitian scalar product induced in $(CT)_X$ for each $X \in W_o$ is the same as would be induced by the standard hermitian scalar product in $C^{2n} = \mathbb{C}W_o$. If

$(x^1, y^1, \ldots, x^n, y^n)$ denote the usual euclidean coordinates in R^{2n}, then vectors $\frac{\partial}{\partial x^1}, \frac{\partial}{\partial y^1} = J_o \frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial y^n} = J_o \frac{\partial}{\partial x^n}$ give an orthonormal basis for T_X and the corresponding vectors in $(CT)_X$ given an orthonormal basis for $(CT)_X$ consisting of real tangent vectors. Note that the basis for $(CT)_X$ consisting of the vectors $\frac{\partial}{\partial z^\alpha}, \frac{\partial}{\partial z^{\bar\beta}}, \frac{\partial}{\partial z^\alpha} = \frac{1}{2}(\frac{\partial}{\partial x^\alpha} - i\frac{\partial}{\partial y^\alpha})$ etc., is orthogonal, but not orthonormal since $\left| \frac{\partial}{\partial z^\alpha} \right| = \frac{1}{2}$. For this hermitian metric, we have

$$(9) \qquad \theta = \frac{i}{2} \Sigma_{\alpha=1}^n dz^\alpha \wedge dz^{\bar\alpha} ,$$

and for the "volume element" Θ,

$$(10) \quad \Theta = dx^1 \wedge dy^1 \wedge \ldots \wedge dx^n \wedge dy^n = (\tfrac{i}{2})^n dz^1 \wedge dz^{\bar 1} \wedge \ldots \wedge dz^n \wedge dz^{\bar n} .$$

9.3. <u>Definition</u>. The $*$-operator in the exterior algebra associated with D_o, determined by the riemannian metric on D_o and the natural orientation, is extended to the complex exterior algebra of differential forms associated with D by taking

$$(11) \qquad *\varphi = f(*\zeta)$$

if $\varphi = f\zeta \in A = A(D)$ and ζ is real, $f \in A^o$.

<u>Remark</u>. Since $*$ can be defined in the real case, for given ζ, by the condition (see (15) of Chapter IX, §9)

$$*\zeta \cdot \eta = (\zeta \wedge \eta) \cdot \Theta \qquad \text{for all real } \eta ,$$

an equivalent definition, for given φ, is

(11') $$*\varphi\cdot\eta = (\varphi\wedge\overline{\eta})\cdot\frac{\theta^n}{n!} \qquad \text{for all } \eta .$$

9.4. Theorem. Let an hermitian metric be given on $D \subset W$. Then the corresponding $*$-operator is a real automorphism

$$*: A \longrightarrow A , \qquad A = A(D) ,$$

with

(12) $$*: A^{r,s} \longrightarrow A^{n-s,n-r}$$

where $n = \dim_{\mathbb{C}} W$, satisfying

(13) $$**\varphi = (-1)^p\varphi , \qquad \varphi \in A^p ,$$

(14) $$\varphi\wedge*\overline{\psi} = \overline{\psi}\wedge*\varphi , \qquad \varphi, \psi \in A^p ,$$

(15) $$*\Pi^{r,s} = \Pi^{n-s,n-r}* .$$

Proof. (12) follows from (11') and (2), which also gives (15). The remaining statements follow as in the real case (Theorem XI, 5.8), using the fact that $\dim_{\mathbb{R}} W_o$ is even.

9.5. Definition. The real operator $\delta = - *d*$ (cf. Definition XI, 5.9) is decomposed into a sum of conjugate operators

(16) $$\delta = \vartheta + \overline{\vartheta}$$

by defining

(17) $$\vartheta = - *\partial*, \quad \overline{\vartheta} = - *\overline{\partial}* .$$

9.6. Theorem. The operator ϑ is of type $(0, -1)$, that is,

$$\vartheta : A^{r,s} \longrightarrow A^{r,s-1} , \qquad s > 0 ,$$

and

(18) $$\vartheta = \Sigma_{r,s} \, \Pi^{r,s-1} \, \delta \, \Pi^{r,s} .$$

Similarly, $\bar{\vartheta}$ is an operator of type $(-1, 0)$. Moreover,

(19) $$\vartheta^2 = 0, \quad \vartheta\bar{\vartheta} + \bar{\vartheta}\vartheta = 0 , \qquad \bar{\vartheta}^2 = 0 ,$$

(20) $$*\vartheta\partial = \partial\vartheta* , \qquad *\bar{\partial}\,\bar{\vartheta} = \bar{\vartheta}\,\bar{\partial}* ,$$

(21) $$\vartheta\partial* = *\partial\vartheta , \qquad \bar{\partial}\,\bar{\vartheta}* = *\,\bar{\vartheta}\,\bar{\partial} .$$

The proof is similar to that for the analogous real case (Theorem XI, 5.10).

9.7. <u>Definition</u>. Corresponding to a given hermitian metric on $D \subset W$, the <u>complex Laplace-Beltrami</u> operators are defined by

(22) $$\Box = \vartheta\,\bar{\partial} + \bar{\partial}\,\vartheta , \quad \bar{\Box} = \bar{\vartheta}\,\partial + \partial\,\bar{\vartheta} .$$

9.8. <u>Theorem</u>. The operators $\Box$, $\bar{\Box}$ of Definition 9.7 are of type $(0, 0)$, that is,

(23) $$\Box : A^{r,s} \longrightarrow A^{r,s} , \quad \bar{\Box} : A^{r,s} \longrightarrow A^{r,s}$$

so that

(23') $$\Box\,\Pi^{r,s} = \Pi^{r,s}\Box , \quad \bar{\Box}\,\Pi^{r,s} = \Pi^{r,s}\,\bar{\Box} .$$

Further,

(24) $$*\,\Box = \Box\,* , \qquad *\,\bar{\Box} = \bar{\Box}\,* ,$$

(25) $$\bar{\partial}\Box = \Box\bar{\partial} = \bar{\partial}\vartheta\bar{\partial}\ , \qquad \partial\bar{\Box} = \bar{\Box}\partial = \partial\bar{\vartheta}\partial\ ,$$

(26) $$\vartheta\Box = \Box\vartheta = \vartheta\bar{\partial}\vartheta\ , \qquad \bar{\vartheta}\bar{\Box} = \bar{\Box}\bar{\vartheta} = \bar{\vartheta}\partial\bar{\vartheta}\ .$$

The above properties are the analogues of those in the real case (Theorem XI, 5.12). However, the operators $\Box$, $\bar{\Box}$ do not, in general, give a decomposition of the real laplacian Δ into a sum of conjugate operators. In fact,

(27) $$\begin{aligned}\Delta = d\delta + \delta d &= (\partial + \bar{\partial})(\vartheta + \bar{\vartheta}) + (\vartheta + \bar{\vartheta})(\partial + \bar{\partial}) \\ &= \Box + \bar{\Box} + (\vartheta\partial + \partial\vartheta) + (\bar{\vartheta}\bar{\partial} + \bar{\partial}\bar{\vartheta})\ ,\end{aligned}$$

where $\vartheta\partial + \partial\vartheta$ is an operator of type $(1, -1)$ and its conjugate $\bar{\vartheta}\bar{\partial} + \bar{\partial}\bar{\vartheta}$ of type $(-1, 1)$.

9.9. <u>Definition</u>. An hermitian metric on D is said to be <u>kählerian</u>, or to be a Kähler metric, if and only if its fundamental form θ is closed, i.e.

(28) $$d\theta = 0\ .$$

<u>Remark</u>. Since θ is of type $(1, 1)$, the condition (28) is equivalent to

(28') $$\partial\theta = 0\ , \qquad \bar{\partial}\theta = 0\ .$$

<u>Example</u>. The hermitian metric for C^n is kählerian (see (9)).

9.10. <u>Theorem</u>. If the hermitian metric on D is a Kähler metric, then

(29) $$\square = \bar{\square} = \frac{1}{2}\Delta \quad .$$

9.11. Corollary. Then $\square$ and $\bar{\square}$ are real operators, and

(30) $$\vartheta\partial + \partial\vartheta = 0 \ , \qquad \bar{\vartheta}\bar{\partial} + \bar{\partial}\bar{\vartheta} = 0 \ ,$$

(31) $$\Delta\Pi^{r,s} = \Pi^{r,s}\Delta \quad .$$

A proof of Theorem 9.10 will be omitted (since it involves a rather tedious calculation). In the case where $D \subset C^n$ and the hermitian metric is given by (9), the proof is left as an exercise (Exercise 11.1). The identities (30) follow from (29) and (27). The property (31) follows from (29) and (23').

The property (31) of a Kähler metric is extremely important since, in certain applications, it provides a relationship between d-coholomogy and $\bar{\partial}$-cohomology.

Any $D \subset W$, where W is a finite dimensional complex vector space, carries at least one Kähler metric, viz. the metric induced by the euclidean Kähler metric of C^n, $n = \dim_C W$, under a linear isomorphism. Further, any two metrics are equivalent in the sense that a volume which is positive for one is also positive for the other.

9.12. Proposition. No finite 2p-cycle of type (n-p, n-p) whose volume is positive is homologous to zero.

Proof. Let c_{2p} be a 2p-cycle of type (n-p, n-p) with positive volume, that is,

$$\int_{c_{2p}} \theta^p > 0 ,$$

where θ is the fundamental form of the Kähler metric, and θ^p is of type (p, p). If $c_{2p} = \partial c_{2p-1}$ where c_{2p-1} is a finite chain, then by Stokes' formula

$$\int_{c_{2p}} \theta^p = \int_{c_{2p-1}} d(\theta^p) = 0$$

(since θ^p is closed if $d\theta = 0$) and this gives a contradiction.

For the case $D = C^n$ with θ given by (9), it follows from Proposition 9.12 and Lemma 6.12 that no finite singular differentiable 2p-cycle of positive volume can be constructed of simplexes $\sigma_{2p} = (s_{2p}, F)$ where s_{2p} is the standard 2p-simplex in C^p and F is holomorphic. This result is often summarized as follows: a compact <u>complex-analytic</u> "variety" (of positive dimension) cannot be imbedded <u>complex-analytically</u>, i.e. holomorphically, into a complex euclidean space of arbitrarily high dimension. If "complex-analytic" is changed to differentiable, such an imbedding is always possible. Complex-analytic "structures" are more rigid than differentiable ones!

§10. Complex Green's formulas

Let $D \subset W$ be a finite domain (Definition.XII, 8.2) and let $\alpha^{r,s}(D)$ denote the subspace of $A^{r,s}(D)$ consisting of forms which are differentiable of class C^∞ on D in the sense of Definition XII, 8.5.

10.1. <u>Definition</u>. Let an hermitian metric be given on

D. Given $\varphi, \psi \in \mathcal{A}(D)$, we define

(1) $$(\varphi, \psi) = \int_D \varphi \wedge * \overline{\psi}$$

(cf. Definition XII, 8.8).

10.2. Proposition. Formula (1) defines an hermitian scalar product in $\mathcal{A} = \mathcal{A}(D)$, that is,

$$\begin{aligned}
&(\varphi, \psi) = \overline{(\psi, \varphi)}, && \varphi, \psi \in \mathcal{A}, \\
&(c\varphi, \psi) = c(\varphi, \psi), && c \in C, \varphi, \psi \in \mathcal{A}, \\
&(\varphi + \eta, \psi) = (\varphi, \psi) + (\eta, \psi) && \varphi, \eta, \psi \in \mathcal{A}, \\
&(\varphi, \varphi) \geqq 0, && \varphi \in \mathcal{A},
\end{aligned}$$

(so we may take $\|\varphi\| = \sqrt{(\varphi, \varphi)}$,

$$\|\varphi\| = 0 \text{ if and only if } \varphi = 0$$

Further, for $\varphi, \psi \in \mathcal{A}$, we have

(2) $$(*\varphi, *\psi) = (\varphi, \psi),$$

(3) $$(J\varphi, J\psi) = (\varphi, \psi),$$

(4). $$(\Pi^{r,s}\varphi, \Pi^{r',s'}\psi) = 0 \qquad \text{unless } (r, s) = (r', s'),$$

(5) $$(\Pi^{r,s}\varphi, \psi) = (\varphi, \Pi^{r,s}\psi)$$

The proof is left as an exercise. Note that (5) states that the operator $\Pi^{r,s}$ is self-adjoint in the sense of Definition 2.18.

10.3. Theorem (complex Green's formulas). Let $D \subset W$ be a finite domain on which is given an hermitian metric. Then,

for $\varphi \in \mathcal{A}^{r,s-1}(D)$, $\psi \in \mathcal{A}^{r,s}(D)$,

(i) $$(\bar\partial\varphi, \psi) - (\varphi, \vartheta\psi) = \int_{bD} \varphi \wedge *\bar\psi \ ;$$

and, for $\varphi, \psi \in \mathcal{A}^{r,s}(D)$,

(ii) $$(\bar\partial\vartheta\varphi, \psi) - (\vartheta\varphi, \vartheta\psi) = \int_{bD} \vartheta\varphi \wedge *\bar\psi \ ,$$

(iii) $$(\bar\partial\varphi, \bar\partial\psi) - (\varphi, \vartheta\bar\partial\psi) = \int_{bD} \varphi \wedge *\overline{\bar\partial\psi} \ ,$$

(iv) $$(\Box\varphi, \psi) - (\bar\partial\varphi, \bar\partial\psi) - (\vartheta\varphi, \vartheta\psi) = \int_{bD} (\vartheta\varphi \wedge *\bar\psi - \bar\psi \wedge *\bar\partial\varphi) \ ,$$

(v) $$(\Box\varphi, \psi) - (\varphi, \Box\psi) = \int_{bD} (\vartheta\varphi \wedge *\bar\psi - \overline{\vartheta\psi} \wedge *\varphi + \varphi \wedge *\overline{\bar\partial\psi} - \bar\psi \wedge *\bar\partial\varphi) \ .$$

<u>Proof</u>. Take $\varphi = \Pi^{r,s-1}\varphi$, $\psi = \Pi^{r,s}\psi$ in (i) of Theorem XII, 8.10; then (i) follows by (13) of §5, (18) of §9, and (5). The remaining formulas are derived from (i) as in the proof of the corresponding real theorem.

10.4. <u>Corollary</u>. If either φ or ψ has a compact support relative to D, then

$$(\bar\partial\varphi, \psi) = (\varphi, \vartheta\psi) \ ,$$

that is, ϑ is the adjoint of $\bar\partial$ in the sense of Definition 2.18; also,

$$(\Box\varphi, \psi) = (\varphi, \Box\psi) \ ,$$

that is, $\Box$ is self-adjoint, and

$$(\Box\varphi, \psi) = (\bar\partial\varphi, \bar\partial\psi) + (\vartheta\varphi, \vartheta\psi)$$

We now specialize to the case that $D \subset C^n$ is a finite

domain and that the metric is the euclidean Kähler metric defined by (9) of §9. Let

$$\gamma(X, Y) = \gamma^{o}(X, Y) = \begin{cases} \dfrac{1}{2(n-1)v_{2n-1}} \dfrac{1}{|X - Y|^{2(n-1)}}, & n > 1, \\ \dfrac{1}{2\pi} \log \dfrac{1}{|X - Y|}, & n = 1, \end{cases}$$

where v_{2n-1} is the volume of the unit $(2n-1)$-sphere of C^n and $X, Y \in C^n$ (compare Definition XII, 9.4).

10.5. <u>Proposition</u>. For $f \in \mathcal{A}^{o}(D)$, $Y \in D$, we have

$$(6) \qquad \tfrac{1}{2}f(Y) = (\bar{\partial}f(X), \bar{\partial}_X\gamma(X, Y)) - \int_{bD} f(X) \wedge * \overline{\bar{\partial}_X\gamma(X, Y)},$$

where, in the right-hand member of (6), X denotes the "variable of integration".

The proof is left as an exercise (compare Proposition XII, 9.7).

10.6. <u>Corollary</u>. If $f \in \mathcal{A}^{o}(D)$ is holomorphic, then

$$(7) \qquad f(Y) = -2\int_{bD} f(X) \wedge * \overline{\bar{\partial}_X\gamma(X, Y)}, \qquad Y \in D.$$

For the special case $n = 1$, let $(z, \bar{z})$ be the complex coordinates of X, and $(w, \bar{w})$ the complex coordinates of Y. Then

$$|X - Y|^2 = (z - w)(\bar{z} - \bar{w}),$$

and

$$\gamma(X, Y) = -\frac{1}{2\pi} \log |X - Y| = -\frac{1}{4\pi}\{\log (z - w) + \log (\bar{z} - \bar{w})\}$$

Then, since $\bar{\partial}z = 0$ and $\bar{\partial}\bar{z} = d\bar{z}$,

$$\bar{\partial}_X \gamma(X, Y) = -\frac{1}{4\pi}\frac{d\bar{z}}{\bar{z} - \bar{w}}, \qquad X \neq Y .$$

Thus,

$$*\partial_X \gamma(X, Y) = -\frac{1}{4\pi}\frac{*dz}{z - w} = -\frac{1}{2\pi i}\frac{dz}{z - w} .$$

[By (11') of §9, $*dz \cdot d\bar{z} = 0$ and $*dz \cdot dz = (dz \wedge d\bar{z}) \cdot \frac{i}{2}(dz \wedge d\bar{z}) = \frac{2}{i}$ since $|\theta| = 1$; that is, $*dz = dz/i$, since $dz \cdot dz = 2$.]

10.7. Proposition. If $D \subset C$ is a finite domain with the euclidean metric, and $f \in \mathcal{A}^0(D)$, then

$$(8)\quad f(w, \bar{w}) = -\frac{1}{\pi}(\bar{\partial} f(z, \bar{z}), \bar{\partial}_z \log r) + \frac{1}{2\pi i}\int_{bD} f(z, \bar{z})\frac{dz}{z - w}, \quad w \in D .$$

If f is holomorphic, then

$$(9)\qquad f(w) = \frac{1}{2\pi i}\int_{bD} f(z)\frac{dz}{z - w}, \qquad w \in D .$$

Remarks. Formula (9) is Cauchy's integral formula. If the coordinates w and $\bar{w}$ were independent, the Cauchy-Riemann equation $\partial f/\partial \bar{w} = 0$ would be equivalent to the statement that "f is independent of $\bar{w}$". If $f \in \mathcal{A}^0(D)$ satisfies the Cauchy-Riemann equation at each point of $D \subset C$, then (9) holds. The Cauchy integral formula shows, in a more precise sense, that f depends only on w. In fact, if we take $D = \{(z, \bar{z}) \mid z \in C$ and $z\bar{z} < 1\}$, and suppose $f \in \mathcal{A}^0(D)$ is holomorphic, then (9) holds for $w\bar{w} < 1$. However, if we write $w = a + ib$, where $a, b \in R$, then formula (9) remains valid for complex values of a and b (with $\operatorname{Im} a$ and $\operatorname{Im} b$ sufficiently small), in which case w and $\bar{w}$ are in fact independent. Thus the statement "a holomorphic function f is independent of $\bar{w}$" is quite true

(in this sense).

In the case where the complex dimension n exceeds 1, the integral formula (7) involves not only $w^1,\ldots,w^n$, but also $\overline{w}^1,\ldots,\overline{w}^n$. However, there is another kind of integral formula for a holomorphic function f (Exercise 11.5) which shows, in the same way, that f depends only on $w^1,\ldots,w^n$ (and not on $\overline{w}^1,\ldots,\overline{w}^n$). In any case, the same conclusion follows from the fact that a holomorphic function f is a holomorphic function of w^α for fixed $w^1,\ldots,w^{\alpha-1}, w^{\alpha+1},\ldots,w^n$, $\alpha = 1, 2,\ldots,n$.

10.8. <u>Proposition</u>. Let $D \subset C$ be the unit coordinate disk and, for $f \in \mathcal{A}^0(D)$, let

$$(10) \qquad g(w, \overline{w}) = \frac{1}{2\pi i} \int_{|z|<1} f(z, \overline{z}) \frac{dz \wedge d\overline{z}}{z - w}, \qquad |w| < 1 .$$

Then

$$(11) \qquad \frac{\partial g}{\partial \overline{w}} = f , \qquad |w| < 1 .$$

To prove this result, one starts from the complex analogue of Theorem XII, 9.9; that is, for $D \subset C^n$ and $f \in \mathcal{A}^0(D)$, define

$$\Gamma f(Y) = (f, \gamma)$$

and show that

$$\Box \Gamma f = \frac{1}{2} f .$$

For $n = 1$, we have (for functions)

$$\frac{\partial^2}{\partial \overline{w} \partial w} = \frac{\partial^2}{\partial w \partial \overline{w}} = -\frac{1}{2} \Box ,$$

so it is sufficient to show that (10) is equivalent to

$$g = -4 \frac{\partial}{\partial w} (\Gamma f) .$$

Then

$$\frac{\partial g}{\partial \bar{w}} = \frac{\partial}{\partial \bar{w}} \left(-4 \frac{\partial}{\partial w} (\Gamma f)\right) = 2i \square \Gamma f = f .$$

§11. Exercises

1. For the euclidean metric of C^n, prove that $\square = \bar{\square} = \frac{1}{2} \Delta$.

2. Given a Kähler metric on $D \subset W$, construct a family of Kähler metrics, depending differentiably on a real parameter, which converge to the given metric as the parameter tends to zero.

3. Prove Proposition 10.5.

4. Show that Proposition 10.8 remains true if it is assumed that $f(z, \bar{z})$ is bounded in the unit coordinate disk D, rather than that $f \in \mathcal{A}^0(D)$. Show that g is also bounded in D.

5. Let $D \subset C^n$ be the unit coordinate polycylinder. If $f \in \mathcal{A}^0(D)$ is holomorphic in D, show that

$$f(w) = \frac{1}{(2\pi i)^n} \int f(z) \frac{dz^1 \wedge dz^2 \wedge \ldots \wedge dz^n}{(z^1 - w^1)(z^2 - w^2)\ldots(z^n - w^n)} , \quad w \in D ,$$

where the integral is extended over $|z^1| = |z^2| = \ldots = |z^n| = 1$.

6. Show, from the integral formula of Exercise 5, that if $f \in \mathcal{A}^0(D)$ is holomorphic in the unit polycylinder $D \subset C^n$, then f can be represented as a power series in $w^1, \ldots, w^n$ which is uniformly convergent for $w \in D$, and conversely.

ANNALS OF MATHEMATICS
Vol. 67, No. 3, May, 1958
Printed in Japan

POTENTIAL THEORY IN BOUNDED SYMMETRIC HOMOGENEOUS COMPLEX DOMAINS

By DAVID B. LOWDENSLAGER

(Received May 27, 1957)

Introduction[1]

The purpose of this paper is to treat a generalization of the elementary potential theory of the unit circle. The main tools are those of the theory of transformation groups. For this reason, the domains treated are required to have a large amount of symmetry; they form a class of homogeneous spaces first discussed in full generality by E. Cartan [6]. A certain class of partial differential equations is shown to arise naturally from the requirements of symmetry. These equations are elliptic on the interior of the domain but degenerate on the boundary; very few existence theorems or explicit solutions for such equations are known. Fortunately, the requirements of symmetry also give simple and explicit solutions by a generalization of the Poisson integral formula.

My interest in this problem stems from a problem studied by S. Bergman [1]. The solution of the differential equations mentioned forms an extended class of functions in the sense of Bergman, and the results presented here will make it possible to extend his results to some new situations; the results also throw some new light, I believe, on his own work as well.

The present paper is devoted to the elementary details of this theory. I have tried to restrict the proofs to symmetry arguments, using mainly integration over a compact group, and, in particular, the integral representation formula proved in the next section. It turns out to be unnecessary to compute any of the integrals explicitly; nevertheless, one formula is computed in passing for a class of domains of some independent interest. Integral formulas for analytic functions or for the real parts of analytic functions on some of these domains have been given by S. Bochner [2] and J. Mitchell [10]. L. K. Hua [8a] has determined kernels including those of Bochner and generalizing the Szego kernels, for the domains considered here. A subsequent paper will examine the rather complicated geometrical details of the boundary value problem considered here, as well as some applications to function theory.

[1] Presented to the Summer Institute on Differential Geometry, June, 1956, and to the American Mathematical Society, December, 1956.

Recently, J. Mitchell remarked that she already knew her Poisson formula [10, eq. (3.6)] is invariant in the sense of this paper. Thus, the results here imply that her formula holds not only for the real parts of analytic functions, but also for solutions of the invariant Laplace equation. She also showed me some mimeographed notes written about the summer of 1956 by L. K. Hua, and containing a very explicit calculation of a complete family of solutions of this equation for the domains of Type I. Hua also indicated the computation of a kernel function which turned out to give the same formula, and how this can be used to get some of the most important results of the present paper for this case. Hua also stated that it is possible to get similar results for the other types of domains.

1. A class of integral representations

Suppose there is given a homomorphism of a topological group G into the group of homeomorphisms of the Hausdorff space X; then I shall say that G acts on X and shall denote the value of the image of g in G at the place x in X by gx. It is required also that gx be continuous in g and x jointly. If f is a function on X, f_g will denote the function such that $f_g(x) = f(gx)$. A family $\mathscr{F}$ of functions will be called *invariant* under G if, for any g in G, $f \in \mathscr{F}$ implies $f_g \in \mathscr{F}$. A point x in X will be called *invariant* under G if $gx = x$ for all g in G. *G acts transitively* on a subset B of X if for every b in B, the set $\{gb \mid g \in G\}$ is exactly B. If the compact group G acts transitively on B, there is a unique integral μ on B satisfying

$$\mu(f) = \mu(f_g) \quad \text{and} \quad \mu(1) = 1 ,$$

for every continuous function f, and g in G. This integral will be called the Haar integral; it is given by

$$\mu(f) = \int_B f(b)\, db = \int_G f(gb)\, dg , \tag{1}$$

for any point b in B, where the right side is the familiar Haar integral on G.

THEOREM 1. *Let the compact group G act on the Hausdorff space X; suppose x in X is invariant and G acts transitively on the subset B. Let $\mathscr{F}$ be a linear set of continuous real functions on x which is invariant under G and contains the constant functions. If*

$$|f(x)| \leq \max_{b \in B} |f(b)| \tag{2}$$

for every f in $\mathscr{F}$, then for every f in $\mathscr{F}$,

$$f(x) = \int_B f(b)\, db , \tag{3}$$

where the integral is given by (1).

PROOF. For any function f on X, $'f$ is the restriction of f to B. Inequality (2) is equivalent to the fact that $I('f) = f(x)$ is a linear function on the set $'\mathscr{F}$ of restrictions of functions of $\mathscr{F}$ to B, and that this function has norm 1 in the norm of the Banach space $C(B)$ of continuous real functions on B. Extend the function I to $C(B)$, and denote the extension again by I. Then I is given by an integral with respect to a measure m. Thus, if $f \in \mathscr{F}$,

$$\begin{aligned} I('f_g) &= f(gx) \\ &= f(x) \\ &= I('f) \\ &= \int_B f(b)\, dm(b) \ , \end{aligned}$$

and

$$\begin{aligned} I('f_g) &= \int_G I('f_g)\, dg \\ &= \int_G \int_B f(gb)\, dm(b)\, dg \\ &= \int_B \left[\int_G f(gb)\, dg \right] dm(b) \ . \end{aligned}$$

Since G is transitive on B, the function in the brackets is constant on B; the value of I on a constant function is the constant, since the constant functions are in $\mathscr{F}$; therefore the right side of the last equation is the right side of (1). Since the left side is equal to $f(x)$, Equation (3) follows.

Theorem 1 may be applied to the family of real parts of analytic functions on the unit disc; the familiar mean value theorem for harmonic functions results. But the unit disc admits a transitive group of complex analytic transformations which can be extended so as to be continuous on the closed disc. By homogeneity, a formula established for any one point can be extended to every point of the domain, and so the Poisson formula results. The Poisson measure corresponding to a point z is the measure on the boundary invariant under the *stability group* of z, i.e., the subgroup of those transformations leaving z fixed. It may easily be shown that the Poisson kernel is harmonic, (see Theorem 3 for a geometrical proof of the general case.) Then a direct examination shows that the Poisson kernel yields a solution of the Dirichlet problem for the disc. It is this chain of reasoning which I wish to generalize.

I shall consider bounded regions in finite dimensional complex space

which admit a transitive group of complex analytic (i.e., pseudoconformal) transformations (called *automorphisms*). In addition an automorphism of the domain with only one fixed point is assumed to be in th NE group; it has been long conjectured but never established that this assumption is a consequence of homogeneity [5], [6]. Cartan has shown that every such domain is the product of certain domains of the same type which are irreducible, in the sense that they are not products of still other domains of the same type. These symmetric complex homogeneous domains (or *Cartan domains*) were classified by Cartan; the irreducible ones fall into four infinite classes with classical simple Lie groups as automorphisms, and two special domains with exceptional simple Lie groups as their groups of automorphisms. Cartan exhibited explicitly only the domains corresponding to classical Lie groups. Recently Harish-Chandra has shown the existence of all of them, by a general method not involving the classification of the real simple Lie groups [7]. However, since the application of the theorems of this paper to these domains depends on explicit verification of the hypotheses for each domain, the final results are proved only for those domains with classical semi-simple groups of automorphisms.

For any point z in a Cartan domain, local co-ordinates with z at the origin may be chosen in such a way that the stability group of z acts by linear transformations; these co-ordinates may be extended to be single valued and 1–1 on the whole domain. Moreover, the domain is again bounded in these co-ordinates [3], [6], [7]. These co-ordinates will be mainly those in use in the rest of this paper; it is not hard to prove that they are essentially unique.

The boundary B in Theorem 1, over which the integration takes place, is generally a proper subset of the topological boundary, and is called the *Bergman-Šilov*[2] *boundary*. For the bicylinder in two variables, the product of two unit discs, the B.-S. boundary is the product of the two circles bounding the discs, and has two real dimensions; the topological boundary is three dimensional.

I show in Theorem 4 that the kernel arising from the integral of Theorem 1 may be applied to any continuous function F on the B.-S. boundary, and yields a function f with F as radial boundary values. However, not every continuous real function on the B.-S. boundary is the boundary value of the real part of an analytic function. In fact, for the bi-

[2] Hereafter abbreviated. S. Bergman called this boundary the Maximumfläche [1a]; later, Šilov introduced the concept in an abstract setting [12].

cylinder just mentioned, if the Fourier series of F is

$$F(e^{ix_1}, e^{ix_2}) \sim \sum a_{mn} e^{imx_1 + inx_2} ,$$

then F is the boundary value of the real part of an analytic function if and only if $a_{mn} = 0$ whenever $mn < 0$. Thus Theorem 1 gives an integral representation of a class of functions more extended than the real parts of analytic functions. This extended class of functions is the set of solutions of the second order partial differential equation

$$\Delta f = 0 , \tag{4}$$

where Δ is the Laplace-Beltrami operator corresponding to Bergman's invariant metric in the domain [1]. For an irreducible Cartan domain, Schur's lemma implies that there is only one linearly independent Riemann metric, since the stability group is irreducible on the tangent space [6]. For the general case (at least for the " classical " Cartan domains), the corollary to Theorem 4 shows that the solutions of all the different invariant Laplace equations are the same. In Section 2, I show that the solutions of (4) satisfy the hypotheses of Theorem 1, by explicit calculation for each class of domain. In order to see more closely the relation of Equation (4) to the real parts of analytic functions, recall that the equations determining these last are

$$\frac{\partial^2 f}{\partial z^i \partial \bar{z}^j} = 0, \qquad i, j = 1, \cdots, n . \tag{5}$$

But since the Bergman metric is a Kaehler metric as well [1], its Laplacian is

$$\Delta = g^{i\bar{j}} \frac{\partial^2}{\partial z^i \partial \bar{z}^j} , \tag{6}$$

where $g^{i\bar{j}}$ is the tensor inverse to the metric tensor [4, p. 132]. From (5) and (6), it is obvious that the real parts of analytic functions are among the solutions of (4); in fact, it is possible to regard the purpose of this paper as the investigation of the solutions of (5) by relaxing the overdetermined system (5) to a single equation, (4).

The reason why the solutions of (4) are determined by their values merely on the B.-S. boundary, is that the operator (6) is singular on the boundary. The coefficients are continuous there, but their matrix has rank less than the dimension, and vanishes, in fact, on the B.-S. boundary. This reflects the fact that these domains are complete in the Bergman metric, which must therefore be infinite at the boundary. The equation for the bicylinder is

$$(1 - |z_1^2|)^2 \frac{\partial^2 f}{\partial z^1 \partial \bar{z}^1} + (1 - |z_2^2|)^2 \frac{\partial^2 f}{\partial z^2 \partial \bar{z}^2} = 0 . \tag{7}$$

The solutions of this differential equation are the functions harmonic in each variable separately, as follows from the corollary to Theorem 3. Although the present use of this differential equation is new, this class NEXT functions (defined by their property of being harmonic in each variable separately) has been used before, in a series of investigations by Bergman. Bergman's classes of extended functions are defined most often for domains similar to the bicylinder, where they are used to extend the powerful potential theoretic methods used in the theory of functions of one variable to the several variable case. The present paper shows that the solutions of (4) form a class of functions with the properties Bergman desires of an extended class of functions, and so offers the possibility of applying Bergman's results to a new and interesting class of domains.

2. The classical Cartan domains and their Laplacians

In this section are described the classical Cartan domains; it is shown that the functions satisfying the invariant Laplace-Beltrami equation satisfy the conditions of Theorem 1. In each domain, the B.-S. boundary B is the set of points where the matrix of coefficients of the operator Δ vanishes; this is also the distinguished boundary in the function theoretic sense. The following lemma is useful in proving inequality (2).

LEMMA 1. *Let*

$$\Delta = a^{ij}(x)\frac{\partial^2}{\partial x^i \partial x^j} + b^k(x)\frac{\partial}{\partial x^k}$$

be a differential operator with coefficients continuous on the closure $\overline{R}$ of a bounded region R in euclidean n-space. Let the matrix of coefficients be positive semidefinite on $\overline{R}$. Suppose that for every point p on the boundary, a local co-ordinate system exists with p as origin such that for some k, the set of points $y^1 = \cdots = y^k = 0$, $|y^{k+1}| < \delta, \cdots, |y^n| < \delta$, is on the boundary of R, and the differential operator Δ involves only derivatives with respect to $y^{k+1}, \cdots, y^n$.

Then, if f is any real non-constant function which is twice continuously differentiable, and satisfies

$$\Delta f \geqq 0 \ ,$$

the set of points where f achieves its maximum intersects the set B where the coefficients of the operator Δ vanish.

PROOF. Suppose the conclusion is false; then $(\sup f) - f(x) > 0$, if $x \in B$. Let w be a twice continuously differentiable function on $\overline{R}$, which is strictly convex, i.e., the matrix of second partials of w is positive defi-

nite everywhere. For some $t > 0$, the function $f + tw$ has a maximum in $\overline{R}$ at a point p not in B. Let $y^1, \cdots, y^n$, be the co-ordinate system guaranteed by the hypothesis, (or the original system translated to p, if p is in the interior). Since $f + tw$ has a maximum at p, its first partials with respect to the y's vanish at p, and the matrix of its second partials is negative semidefinite. Hence $\Delta(f + tw) \leqq 0$ at p. But $\Delta f \geqq 0$, and, since w is strictly convex and the matrix of coefficients of Δ is positive semidefinite and not zero, $\Delta w > 0$. This contradiction establishes the lemma.

There is an irreducible Cartan domain of the first series for every pair of integers m, n satisfying $n \geqq m \geqq 0$. If the elements of an m by n matrix Z are considered as co-ordinate in complex mn-space, the set of matrices with

$$(8) \qquad I - Z^*Z \quad \text{positive definite}$$

form an irreducible Cartan domain contained in the unit sphere [6], [8]. Here I is the n by n unit matrix and Z^* is the conjugate transpose of Z. Condition (8) is equivalent to

$$(8') \qquad J - ZZ^* \quad \text{positive definite},$$

where J is the m by m unit matrix. The group of automorphisms leaving the zero matrix fixed includes the set of transformations.

$$(9) \qquad Z \to UZV, \qquad U,\ m \text{ by } m \text{ unitary};\ V,\ n \text{ by } n \text{ unitary}.$$

The B.-S. boundary is the set of matrices with the matrix of (8′) equal to zero (but not that of (8), unless $m = n$.) This condition is equivalent to the requirement that the m rows of the matrix Z be m orthonormal vectors in complex n-space, and the transitivity of the stability group of zero follows from this geometrical consideration. An invariant metric is given by

$$(10) \qquad \operatorname{tr}\left[(J - ZZ^*)^{-1}\, dZ (I - Z^*Z)^{-1}\, dZ^*\right],$$

where dZ is the matrix of differentials [8].

An invariant Laplacian for the domain may be formed from (10) by computing the inverse of the metric tensor and using (6). The computation is easy; the interested reader may verify that the operator is

$$(11) \qquad \Delta = (J - ZZ^*)^{jp}(I - Z^*Z)^{qk} \frac{\partial^2}{\partial Z^{pq} \partial \overline{Z}^{jk}},$$

where the indices denote the matrix elements. (see, for the case $n = m$, [10].) The continuity condition of Lemma 1 is evidently satisfied, so that

only the condition on tangential derivatives must be verified on the boundary. The operator clearly vanishes on the B.-S. boundary. If Z is any other boundary point, $J - ZZ^*$ is merely positive semidefinite, that 0 is a proper value of multiplicity $m - k$ of it. The transformations (9) correspond to changes of basis in m-space and n-space. Choose a basis in m-space with $ZZ^*e_j = e_j$ for the first $m - k$ basis vectors; then the first $m - k$ vectors Z^*e_j are orthonormal in n-space and may be chosen as the first vectors of a basis there. Then condition (8) implies that Z is transformed into a matrix of the form

$$\begin{pmatrix} I & 0 \\ 0 & W \end{pmatrix}, \tag{12}$$

where I is a unit matrix and W is an $m - k$ by $n - k$ matrix satisfying the analog of (8). All such matrices are on the boundary, and the Laplacian (11) involves only derivatives with respect to the variables in W. These co-ordinates satisfy the conditions of Lemma 1.

The domains of Types II and IV have been extensively studied by Siegel [11], and Klingen [9] (whose numbering is not that of Cartan.) There is one domain of each series for each positive integer $n > 1$. This domain is given by the n by n matrices satisfying (8) and

$$\begin{aligned} \text{II:} \quad & Z + {}^tZ = 0 , \\ \text{IV:} \quad & Z - {}^tZ = 0 \end{aligned} \tag{13}$$

respectively, where ${}^tZ$ is the transpose of Z. The stability group of 0 contains the transformations of the form (9) with $V = {}^tU$. The B.-S. boundary in each case is given by the matrices such that ZZ^* has the maximum multiplicity for the proper value 1. By the theorems of Section 1 of [9], the stability groups are transitive on these B.-S. boundaries.

An invariant metric for either domain is given by (10). The invariant Laplacian is not given by (11) and it must be calculated by calculating the inverse of the restriction of the metric matrix to the subspaces of skew and symmetric matrices respectively. However, since $\operatorname{tr} AB = 0$ if A is symmetric and B is skew, under the metric (10) a symmetric and a skew matrix of differentials are orthogonal along either the submanifold of skew matrices or that of symmetric ones. Thus, if one changes co-ordinates to

$$X = \frac{1}{2}(Z + {}^tZ) , \qquad Y = \frac{1}{2}(Z - {}^tZ) , \tag{14}$$

dX is orthogonal to dY along either manifold, and the inverse of the restriction of the metric matrix is the restriction of the inverse. The

change of variables (14) is particularly simple so that it is not necessary to display the explicit form of the Laplace operator. The terms below the diagonals of the matrices X and Y are not used as co-ordinates, since they are superfluous.

By the results in Section 1 of [9], every matrix in the boundary can be moved by the stability group of 0 to a matrix of the form

$$\begin{pmatrix} 0 & I & 0 \\ -I & 0 & 0 \\ 0 & 0 & W \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} I & 0 \\ 0 & W \end{pmatrix}, \qquad I - W^* W \text{ positive definite},$$

in Case II or IV respectively. It is easy to see that the variables in the matrices W are the only ones occurring in the operator (11). Hence, because of the special form of the change of co-ordinates (14), these variables are the only ones occurring in the invariant Laplacian for the domains considered. Also, since the change of co-ordinates is linear, the coefficients of this Laplacian have the same continuity properties as those of (11). Thus Lemma 1 holds for these domains.

The third class of Cartan domains has not been studied so much as the other three, although, in some respects, it is the simplest and is of some independent interest. There is an irreducible Cartan domain of Type III for every positive integer n. Its points are the vectors in n-space (or n by 1 matrices) satisfying

$$\begin{aligned} &|{}^t ww|^2 - 2w^*w + 1 > 0\,, \\ &|{}^t ww|^2 - 1 < 0\,. \end{aligned} \tag{15}$$

The B.-S. boundary is the set of points satisfying

$$\begin{aligned} &|{}^t ww|^2 - 2w^*w + 1 = 0\,, \\ &|{}^t ww|^2 - 1 = 0\,, \end{aligned} \tag{16}$$

which is equivalent to

$$\begin{aligned} &w_k = e^{i\phi} u_k\,, \qquad u_k = \bar{u}_k\,, \\ &u_1^2 + u_2^2 + \cdots + u_n^2 = 1\,. \end{aligned} \tag{16'}$$

The subgroup of automorphisms leaving the origin 0 fixed includes the transformations

$$\begin{aligned} &w_k \to e^{i\theta} w_k\,, \\ &w \to Qw\,, \qquad Q \text{ a real orthogonal matrix}\,. \end{aligned} \tag{17}$$

These are transitive on the B.-S. boundary.

It is useful to change co-ordinates using the generalized Cayley transformation [2], [6]

$$z_k = -2iw_k W^{-1}, \qquad k = 1, \cdots, n-1, \tag{18}$$
$$z_n + i = 2(w_n - i)W^{-1},$$

where

NEXT

$$W = \sum_2^{n-1} w_k^2 + (w_n - i)^2;$$

the inverse is

$$w_k = -2iz_k Z^{-1}, \qquad k = 1, \cdots, n-1, \tag{18'}$$
$$w_n - i = -2(z_n + i)Z^{-1},$$

where

$$Z = \sum_1^{n-1} z_k^2 - (z_n + i)^2 = -4W^{-1}.$$

The inequalities (15) are transformed into

$$y_n > (y_1^2 + \cdots + y_{n-1}^2)^{1/2} - \infty < x_k < +\infty, \tag{19}$$

where

$$z_k = x_k + iy_k, \qquad x_k, y_k \text{ real}, 1 \leqq k \leqq n.$$

The B.-S. boundary (16) is given by

$$y_1 = \cdots = y_n = 0, \tag{20}$$
$$-\infty < x_k < +\infty,$$

except for a subset of lower dimension. This domain arises naturally in trying to solve the wave equation using the Laplace transformation in several variables.

The Jacobian matrix of the transformation (18′) is given by the equation

$$\frac{i}{2}\frac{\partial w_k}{\partial z_j} = \begin{cases} \delta_{kj}Z^{-1} - z_k \dfrac{\partial Z}{\partial z_j} Z^{-2}, & k < n, \\ -i\delta_{nj}Z^{-1} + i(z_n + i)\dfrac{\partial Z}{\partial z_j} Z^{-2}, & k = n. \end{cases} \tag{21}$$

Multiply the last row by $-i(z_k)(z_n + i)^{-1}$ and add this to the kth row, for $k < n$; then multiply the jth column $(j < n)$ by $z_j(z_n + i)^{-1}$ and add this to the nth column. The result is a diagonal matrix, with Z^{-1} along the first $n - 1$ diagonal positions, and

$$-iZ^{-1} + iZ^{-2}\left(\sum_1^{n-1} z_k \frac{\partial Z}{\partial z_k} + (z_n + 1)\frac{\partial Z}{\partial z_n}\right) = iZ^{-1}$$

in the nth place.

Therefore the Jacobian of the transformation (18′) is

$$J = -2^n(-i)^{n+1}Z^{-n}. \tag{22}$$

The measure on the boundary (16) invariant under (17) is ordinary euclidean surface measure, The total measure is $\pi\sigma_{n-1}$, where $\sigma_{n-1} =$

$(2\pi)^{n/2}(\Gamma(n/2))^{-1}$ is the measure of the surface of the unit sphere in real n-space.

The Jacobian of the transformation from (16) to (20) is the absolute value of the complex Jacobian (22). Therefore, the Poisson integral for the point $(0, \cdots, 0, i)$ is

$$f(0, \cdots, 0, i) = \frac{2^n}{\pi\sigma_{n-1}} \int_{R^n} \frac{f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n}{|\sum_1^{n-1} x_j^2 - x_n^2 + 1 - 2ix_n|^n} \tag{23}$$

Now, consider the following sequence of transformations:

$$\begin{aligned}
&Z_k \to tz_k\,, && t > 0\,;\\
&Z_1 \to z_1 \cosh s + z_n \sinh s\,, &&\\
&Z_k \to z_k\,, && 2 \leqq k \leqq n\,,\\
&Z_n \to z_1 \sinh s + z_n \cosh s\,; &&
\end{aligned} \tag{24}$$

$$Z_k \to Q_{kj} z_j,\quad k, j < n\,,\quad Q \text{ an } n-1 \text{ by } n-1 \text{ real orthogonal matrix}\,,$$

$$\begin{aligned}
&Z_n \to z_n\,;\\
&Z_k \to z_k + a_k\,, && a_k \text{ real}\,.
\end{aligned}$$

The original point $(0, \cdots, 0, i)$, may be moved to an arbitrary point of the domain (19) by a properly chosen sequence of transformations (24), and so the integral (23) may be computed for an arbitrary point of this domain. The result is:

A Poisson integral for the domain (19) *is given by*

$$f(u + iv) = \frac{2^n}{\pi\sigma_{n-1}} \int_{R^n} \frac{f(x)\,(v, v)^{n/2}\, dx}{[((x-u, x-u) - (v, v))^2 + 4(x-u, v)^2]^{n/2}} \tag{25}$$

where $dx = dx_1 \cdots dx_n$, *and* $(u, v) = u_n v_n - \sum_1^{n-1} u_k v_k$ *is the Lorentz inner product of the two real vectors* u *and* v.

Next, the invariant Laplacian must be computed for these domains, and must be proved to satisfy the hypotheses of Lemma 1. The invariant metric for the Cartan domains of Type III seems not to be available, so it will be computed here. It is clear that a positive multiple of

$$\sum_1^n |dw_k|^2$$

is the only metric tensor on the domain invariant under the group of automorphisms (17). It follows from the form of the Jacobian matrix (21), that an invariant metric at the point $(0, \cdots 0, i)$ in the co-ordinates (19) is given by

$$\sum_1^n |dz_k|^2\,.$$

Transforming this metric by a sequence of transformations like (24), it may be seen that an invariant metric for the domain is given by

$$ds^2 = \frac{1}{Y^2}\{(y_1^2 + \cdots + y_n^2)(dz_n d\bar{z}_n) - \textstyle\sum_1^{n-1} 2y_j y_n (dz_j d\bar{z}_n + d\bar{z}_j dz_n)$$

$$+ Y(\textstyle\sum_1^{n-1} dz_j d\bar{z}_j) + 2\,|\textstyle\sum_1^{n-1} y_j dz_j|^2\} \tag{26}$$

$$Y = y_n^2 - y_1^2 - \cdots - y_{n-1}^2 .$$

It may be checked that the matrix of the following differential operator is the inverse of the matrix of (26):

$$= \textstyle\sum_1^{n-1} (Y + 2y_k^2) \dfrac{\partial^2}{\partial z_k \partial \bar{z}_k} + (2y_n^2 - Y)\dfrac{\partial^2}{\partial z_n \partial \bar{z}_n}$$

$$+ 2\textstyle\sum_{j<k} y_j y_k \left(\dfrac{\partial^2}{\partial \bar{z}_j \partial z_k} + \dfrac{\partial^2}{\partial z_j \partial \bar{z}_k}\right). \tag{27}$$

This operator has coefficients which vanish on the B.-S. boundary, and are continuous on the closure of the domain. By a transformation leaving the point $(0, \cdots, 0, i)$ fixed, any other boundary point may be moved to the point $(i, 0, \cdots, 0, i)$. The points $(u, 0, \cdots, 0, u)$ are all in the boundary for u near i, and at the point $u = i$ the Laplacian (27) is $2(\partial^2/\partial u \partial \bar{u})$. This verifies the hypotheses of Lemma 1.

The most general classical Cartan domain (one with classical semi-simple group of automorphisms) is a product of the irreducible domains just enumerated. There is, in general, no longer a Laplace operator invariant under the automorphism group and unique up to a positive multiple. Suppose $D = D_1 \times D_2$, where D_1 and D_2 are irreducible; then, if Δ_1 and Δ_2 are *any* invariant Laplacians on D_1 and D_2 respectively, the formal sum $\Delta = \Delta_1 + \Delta_2$ is invariant under the connected component of the identity of the group of automorphisms of D. The converse is true. If D_1 is not isomorphic to D_2, then Δ is invariant under every automorphism of D, but if D_1 is isomorphic to D_2, Δ is invariant under switching co-ordinates only when $\Delta_1 = \Delta_2$. In order to simplify the statements, only connected groups of automorphisms will be considered here.

The B.-S. boundary of the product of two domains in the sense of function theory is the product of the two B.-S. boundaries. From the remarks above, it follows that this is also the set of points where the coefficients of any invariant Laplacian vanish. The stability group of a point (x_1, x_2) is the product of the stability groups of x_1 and x_2. Hence the property of transitivity on the B.-S. boundary carries over. The preceding results may be summarized by

THEOREM 2. *Let D be a classical Cartan domain. Any invariant second order elliptic operator Δ has coefficients which are continuous on the closure, $\overline{D}$. If B denotes the set of points where Δ vanishes, the stability group G of any point $x \in D$ is transitive on B. The solutions of*

$$\Delta f = 0$$

which are twice continuously differentiable on $\overline{D}$ satisfy the hypotheses of Theorem 1.

3. The Dirichlet problem

The transformations of the automorphism group of a classical Cartan domain D can be extended continuously to the closure $\overline{D}$ where they are differentiable, and even real analytic (using the special co-ordinates in which some stability group consists of linear transformations [11], [7]). These transformations carry the B.-S. boundary B onto itself; and so there is a fixed measure m on B such that the Poisson measure for each point $x \in D$ is absolutely continuous with respect to m, and even has a continuous positive Radon-Nikodym derivitive $K(x, b)$ with respect to m. K will be called a *Poisson kernel* on B. Two different Poisson kernels K and K' for different measures m and m' are related by the property that $K^{-1}K'$ is a continuous function of b only.

THEOREM 3. *Let D be a classical Cartan domain, Δ an invariant Laplacian, and K a Poisson kernel for D. Then K as a function on D satisfies*

$$\Delta K = 0, \textit{ for all } b \in B\,. \tag{28}$$

PROOF. If $D = D_1 \times D_2$, then $\Delta = \Delta_1 + \Delta_2$, $B = B_1 \times B_2$, and m can be chosen on B such that $K = K_1K_2$, where K_j is a Poisson kernel for B_j. The theorem for any one Poisson kernel implies that it holds for all Poisson kernels for the domain. Therefore, it suffices to prove the theorem for irreducible domains only. By homogeneity, it suffices to prove the theorem for just one point x in the irreducible domain D, and the basic measure m may even be chosen so as to be the invariant measure for the stability group G of x.

Let $x(t)$, $-1 \leqq t \leqq +1$, be a curve in D with non-vanishing tangent and $x(0) = x$. There is a positive constant c such that for every differentiable function f on D, the following equation holds at the point x:

$$c\Delta f = \int_G \left[\frac{d^2}{dt^2} f_g(x(t))\right](0)\, dg\,. \tag{29}$$

In fact, the right side is a second order differential operator at x which

is invariant under G. Since D is irreducible, G acts irreducibly on the tangent space [6]. By Schur's lemma, an invariant second order differer tial operator is unique up to scalar multiples. By choosing f to have nondegenerate minimum at x, it is seen that the constant in question must be positive.

Next, I claim that, for the special choice of K used,

$$K(gz, b) = K(z, g^{-1}b) , \qquad \text{for all } g \in G, z \in D, b \in B . \tag{30}$$

Let the stability subgroup of z be H; that of g_z is gHg^{-1}, which is isomorphic to H. Denoting the Poisson integral for z by $I_z(f)$, and using the notation and results of Theorem 1,

$$\begin{aligned} I_{gz}(f) &= \int_H f(ghg^{-1}b_1)\, dh = \int_H f_g(hg^{-1}b_1)\, dh \\ &= \int_H f_g(hb_2)\, dh = \int_B f_g(b)K(z, b)\, db , \end{aligned}$$

since the point b in (5) is arbitrary. Since $I_z(f_g) = I_z(f)$, the last integral is

$$\int_B f(b)K(z, g^{-1}b)\, db ,$$

But, also,

$$I_{gz}(f) = \int_B f(b)K(gz, b)\, db ,$$

and comparing these two integrals gived equation (30). Let $x(t)$, $-1 \leqq t \leqq +1$, be a curve through x as above. Then, using (30) and the fact that Haar measure on the compact group G is invariant under the transformation $g \to g^{-1}$,

$$\begin{aligned} \int_G K(gx(t), b)\, dg &= \int_G K(x(t), g^{-1}b)\, dg \\ &= \int_G K(x(t), gb)\, dg = \int_B K(x(t), b)\, db = 1 , \end{aligned}$$

the last integral being the value at $x(t)$ of the function constantly 1.

To complete the proof, invert the order of integration and differentiation on the right of (29), use the last equation, and (28) is the result.

COROLLARY. *Let D be the product of two classical Cartan domains D_1 and D_2. If f is a twice continuously differentiable function on D which satisfies*

$$\Delta f = 0 , \quad \text{and} \quad \Delta = \Delta_1 + \Delta_2 ,$$

where Δ_1 and Δ_2 are invariant Laplacians for D_1 and D_2, then

$$\Delta_1 f = \Delta_2 f = 0 \,. \tag{31}$$

The proof uses the fact that f is given by an iterated Poisson integral, the factor of which statisfies (31). NEXT

The corollary does not hold if f is assumed to be harmonic only in a subdomain of D. Consider, in the bicylinder, which is the product of two discs of radius 1, the differential equation (4). In the product of two discs of radius 1/2 whose closure is contained in the interior of the domain, the equation (4) is elliptic, not degenerating on the three dimensional boundary. Classical existence theorems give harmonic functions with preassigned values on the boundary, and so these can be constructed so that their maximum is not attained on the B.-S. boundary of the subdomain. These solutions cannot be harmonic in each variable separately. It is also possible, using the iterated Poisson integral, to find functions harmonic in each variable separately, and thus satisfying (4), which have arbitrary boundary values on the two dimensional B.-S. boundary of the subdomain.

The last part of the solution of the Dirichlet problem is the construction of a function u satisfying $\Delta u = 0$, with preassigned boundary values on the B.-S. boundary. What will be proved here is that u can be found with arbitrary continuous function f as radial limit. Suppose the co-ordinates are chosen so that the stability group of the origin is linear. By a *radius* is meant a straight line segment with one endpoint at the origin and the other on the B.-S. boundary. All radii are conjugate under the stability group of the origin, since this group is transitive on the B.-S. boundary.

THEOREM 4. *Let D be a classical Cartan domain with* B.-S. *boundary B. Let f be a continuous function on B. There exists a solution of*

$$\Delta u = 0$$

on D such that $\lim u = f$ *along every radius.*

PROOF. Let u be defined by the invariant Poisson integral

$$u(z) = \int_B f(b) K(z, b) d(b) \,;$$

u is harmonic by Theorem 3. All that needs to be proved is the statement on boundary values. If R is a radius, there is a one parameter subgroup g_s of automorphisms of D such that $g_s(0) \in R$ for $s > 0$; this *radial translation* is unique up to multiplying s by a positive constant [6], but an explicit construction follows. By a change of variables,

$$u(g_s(0)) = \int_B f_{g_s}(b)\, db \ ,$$

where db is the measure invariant under the stability group of 0. Let NEXT

$$b_0 = \lim_{s\to\infty} g_s(0) \ ,$$

be the endpoint of the radius. It is only necessary to show that

$$\lim_{s\to\infty} \int_B f_{g_s}(b)\, db = f(b_0) \ .$$

This will follow when it is shown that there exists a subset A of lower dimension such that, if V is the complement of any neighborhood of b_0, $g_s V$ converges to A uniformly as $s \to -\infty$. The remainder of the proof consists of the demonstration of the last fact; it suffices to consider only irreducible domains, and, by homogeneity, only one point in each B.-S. boundary. Automorphism groups of the classical Cartan domains have been examined, without proofs, [11, pp. 151–161], and proofs of some of Siegel's statements have appeared [9], but the actual subgroups used here may easily be shown to consist of automorphisms by examining the infinitesimal generator of each.

For domains of Type III, using the Cayley transformation (18), the radial translation is

$$g_s(z) = e^{-s} z \ .$$

The point b_0 is the origin; its inverse under (18) is $(0, \cdots, 0, -i)$. As $s \to -\infty$, any point but this inverse converges to the set A, which is the set of exceptional boundary points for the Cayley transformation.

For domains of Type IV, or of Type I with $m = n$, thetran sformations

$$g_s(Z) = (Z \cosh s - I \sinh s)(-Z \sinh s + I \cosh s)^{-1} \tag{32}$$

translate 0 radially to I. If U is a point of either B.-S. boundary, it may be moved to a diagonal matrix by an automorphism of the stability group of 0 which leaves I fixed. The automorphism commutes with g_s, and each of the diagonal entries of the transformed matrix must have absolute value 1; the effect of g_s then follows from the 1-dimensional case. Let A be the subset of the B.-S. boundary satisfying $Z^2 = I$; if V is the complement of any neighborhood of I, $g_s V$ converges to A as $s \to -\infty$. Notice that the same thing is true if V is the complement of a neighborhood of I in $\overline{D}$. For the domains considered here, a Cayley transformation is available under which the radial translation is similar to that in the domains of type III.

For a domain of Type I with $m < n$, write $Z = (Z_1, Z_2)$, where Z_1 is an m by m matrix, and Z_2 is an m by $n - m$ matrix. The 1 parameter group

$$h_s(Z_1, Z_2) = (g_s(Z_1)\ ,\ g_s(Z_1)\, Z_2 \cosh s)\ ,$$

where g_s is given by (32), is a subgroup of automorphisms of the form

$$Z \to (AZ + B)(CZ + D)^{-1}\ ,$$

$$A = I \cosh s\ , \qquad B = (-\, I \sinh s,\ 0)\ ,$$

$$C = \begin{pmatrix} -\, I \sinh s \\ 0 \end{pmatrix}, \qquad D = \begin{pmatrix} I \cosh s & 0 \\ 0 & I \end{pmatrix},$$

described in [11, p. 152]. As was noticed, $g_s(Z_1)$ converges to A described above. Since if $U = (U_1, U_2)$ is on the B.-S. boundary on the domain now considered, $U_1 U_1^* + U_2 U_2^* = I$, if U is not $(I, 0)$, $h_s(U)$ converges to a point of the form $(Z_1, 0)$, where $Z_1 \in A$. This implies the result.

The domains of Type II fall into two subclasses, according to whether their matrices are even dimensional or odd. Each domain is naturally imbedded in a domain of the first type. In the even case, the B.-S. boundary B is the intersection of $\overline{D}$ with the B.-S. boundary B' of the domain D' of type I containing D. A typical element of B is

$$C = \begin{pmatrix} 0 & iI \\ -iI & 0 \end{pmatrix}, \tag{34}$$

where I is a unit matrix. The one parameter group

$$g_s(Z) = (Z \cosh s - C \sinh s)(-CZ \sinh s + I \cosh s)^{-1} \tag{35}$$

is a group of radial translations both of D and D'. Since all radial translations of D' are conjugate, there is a set $A = A' \cap B$ such that the exterior of any neighborhood of C converges to $A \in s$ it is translated by (35) with $s \to -\infty$. A is easily seen to be lower dimensional in B.

In the second case, when n is odd, the n by n matrix Z may be written

$$Z = \begin{pmatrix} Z_1 & Z_2 \\ Z_3 & Z_4 \end{pmatrix},$$

in block form, where Z_4 is $n - 1$ by $n - 1$. A typical element of the B.-S. boundary is given by

$$\begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix}, \tag{36}$$

with C the matrix of (34). A subgroup of radial translations is given by

$$h_s(Z) = (A_s Z + B_s)(C_s Z + D_s)^{-1}\ ,$$

where

$$A_s = D_s = \begin{pmatrix} I & 0 \\ 0 & I \cosh s \end{pmatrix}, \qquad B_s = C_s = \begin{pmatrix} I & 0 \\ 0 & -\, C \sinh s \end{pmatrix},$$

as in [11, p. 155]. The lower right block of $h_s(Z)$ is given by $g_s(Z_s)$, as in (35). As before, this converges to a point of A above, and since $I - Z^*Z$ has rank 1 on the B.-S. boundary, the other blocks of $h_s(Z)$ converge to 0, unless Z is given by (36). This completes the proof.

It should be remarked that the most important fact in each of these cases is that the trajectories of the one parameter family of transformations almost all start at the one point and end at the antipodal point. I conclude by remarking that the uniqueness of the solution to the Dirichlet problem does not follow from Theorem 2. The hypotheses of Lemma 1 require continuity and differentiability on the closure of the domain. Au niqueness theorem will be contained in a paper examining the geometry of these equations near the B.-S. boundary.

UNIVERSITY OF ILLINOIS

BIBLIOGRAPHY

1. S. BERGMAN, *Kernel function and extended classes . . . , Colloque sur les fonctions de plusieurs variables*, Centre Belge de Recherches Math., 1953, 135–157.

1a. ———, *Über die Abbildungen durch Paare von Funktionen . . .*, Jber. Deutschen. Math. Verein., 39 (1930), 24–27.

2. S. BOCHNER, *Group invariance of Cauchy's formula in several variables*, Ann. of Math., 45 (1944), 686–707.

3. ——— and W. T. MARTIN, Several complex variables, Princeton University Press, 1948.

4. ——— and K. YANO, Curvature and Betti numbers, Ann. of Math. Study no. 32, 1953.

5. A. BOREL, *Kählerian coset spaces of semisimple Lie groups*, Proc. Nat. Acad. Sci. U.S.A., 40 (1954), 1147–1151.

6. E. CARTAN, *Sur les domaines bornés homogènes . . .*, Abh. Math. Sem. Hamburg, 11 (1935) 116–162; Oeuvres, partie 1, 1259–1304.

7. HARISH-CHANDRA, *Representations of semisimple Lie groups*, VI, Amer. J. Math., 78 (1956), 564–628.

8. L. K. HUA, *On the theory of automorphic functions of a matrix variable*, Amer. J. Math., 66 (1944), 470–488.

8a. ———, *On the theory of functions of several complex variables*, I–III, (in Chinese), Acta Math. Sinica, 2 (1953), 288–323; 5 (1955), 1–25, 205–242.

9. H. KLINGEN, *Diskontinuierliche Gruppen in symmetrischen Räumen*, I, Math. Ann., 129 (1955), 345–369.

10. J. MITCHELL, *Potential theory in the geometry of matrices*, Trans. Amer. Math. Soc., 79 (1955), 401–422.

11. C. L. SIEGEL, Analytic functions of several complex variables, Inst. Adv. Study, Princeton, 1949.

12. G. ŠILOV (with I. GELFAND and D. RAIKOV), *Commutative normed rings*, Uspehi Mat. Nauk. N. S., 1 (1946), 48–146.

ГАРМОНИЧЕСКИЙ АНАЛИЗ ФУНКЦИЙ МНОГИХ КОМПЛЕКСНЫХ ПЕРЕМЕННЫХ В КЛАССИЧЕСКИХ ОБЛАСТЯХ

ХУА ЛО-КЕН

Издательство Иностранной Литературы

Москва 1959

Original Chinese text published by Science Press, Peking, 1958

Translated from the Russian
by Leo Ebner and Adam Korányi

Publication aided by grant NSF-GN 57
from the
NATIONAL SCIENCE FOUNDATION

Text composed on Photon, partly subsidized by NSF Grant G21913

International Standard Serial Number 0065-9282
International Standard Book Number 0-8218-1556-3
Library of Congress Card Number 63-16769

Third printing, revised, 1979
Reprinted 1986

Printed in the United States of America

The paper used in this journal is acid-free and falls within the guidelines
established to ensure permanence and durability

INTRODUCTION

I. **Classical domains.** By a classical domain we shall understand an irreducible bounded symmetric domain (in the space of several complex variables) of one of the following four types:

(1) The domain $\mathfrak{R}_{\mathrm{I}}$ of $m \times n$ matrices with complex entries satisfying the condition

$$I^{(m)} - Z\overline{Z'} > 0.$$

Here $I^{(m)}$ is the identity matrix of order m, $\overline{Z'}$ is the complex conjugate of the transposed matrix Z'. ($H>0$ for a hermitian matrix H denotes, as usual, that H is positive definite.)

(2) The domain $\mathfrak{R}_{\mathrm{II}}$ of symmetric matrices of order n (with complex entries) satisfying the condition

$$I^{(n)} - Z\overline{Z} > 0.$$

(3) The domain $\mathfrak{R}_{\mathrm{III}}$ of skew-symmetric matrices of order n (with complex entries) satisfying the condition

$$I^{(n)} + Z\overline{Z} > 0.$$

(4) The domain $\mathfrak{R}_{\mathrm{IV}}$ of n-dimensional ($n>2$) vectors

$$z = (z_1, z_2, \cdots, z_n)$$

(z_k are complex numbers) satisfying the conditions[2]

$$|zz'|^2 + 1 - 2zz' > 0, \qquad |zz'| < 1.$$

The complex dimension of these four domains is mn, $n(n+1)/2$, $n(n-1)/2$, n, respectively.

The author has shown (cf. L. K. Hua [3]) that $\mathfrak{R}_{\mathrm{IV}}$ can also be regarded as a homogeneous space of $2 \times n$ real matrices. Therefore, the study of all these domains can be reduced to a study of the geometry of matrices.

In 1935, E. Cartan [1] proved that there exist only six types of irreducible homogeneous bounded symmetric domains. Beside the above four types, there exist only two; their dimensions are 16 and 27. Of course

[2] Translator's note (n.b., unless otherwise noted, these words refer to the Russian translator). Here and throughout, the author considers a vector as a matrix of one row and n columns. So z' is a matrix of one column and n rows (the transpose of the matrix z).

these two types are rather special. The problem of the explicit description of these two types is still open.

The purpose of the present book is to study harmonic analysis on the classical domains. (The exact content of this harmonic analysis will be outlined later.)

II. **Characteristic manifolds.** Let $\mathfrak{R}$ be a bounded homogeneous domain in the $2n$-dimensional Euclidean space of n complex variables $z=(z_1, z_2, \cdots, z_n)$, and $f(z)$ an analytic function of z, regular in $\mathfrak{R}$. It is known that the maximum of the modulus of the function $f(z)$ is assumed on the boundary of $\mathfrak{R}$. Let $\mathfrak{C}$ be a manifold on the boundary of $\mathfrak{R}$ having the following properties:

(a) The modulus of every analytic function regular in $\mathfrak{R}$ assumes its maximum on $\mathfrak{C}$.

(b) For every point a on $\mathfrak{C}$ there exists a function $f(z)$, regular on $\mathfrak{R}$, such that the modulus of $f(z)$ assumes its maximum at $z=a$.

Such a manifold $\mathfrak{C}$ is called a characteristic manifold of the domain $\mathfrak{R}$. We should mention that $\mathfrak{C}$ is in general a proper subset of the boundary, and that the dimension of $\mathfrak{C}$ may be much less than $2n-1$. It is clear that $\mathfrak{C}$ is uniquely determined by $\mathfrak{R}$. It is easy to show that $\mathfrak{C}$ is closed, and that an analytic function which is regular in a neighborhood of each point of $\mathfrak{C}$ is uniquely determined by its values on $\mathfrak{C}$. Hence it follows that the real dimension of $\mathfrak{C}$ is not less than n. We shall denote by ξ the variable on $\mathfrak{C}$, and by $d\xi\overline{d\xi'}$ and $\dot{\xi}$ the metric and the element of volume of $\mathfrak{C}$.

Clearly, in the definition of $\mathfrak{C}$ it is enough to consider only linear functions instead of all analytic functions.

We describe the characteristic manifolds of the classical domains.

(1) $\mathfrak{C}_{\mathrm{I}}$ consists of the $m\times n$ matrices U satisfying the condition

$$U\overline{U'}=I^{(m)}.$$

(2) $\mathfrak{C}_{\mathrm{II}}$ consists of all symmetric unitary matrices of order n.

(3) $\mathfrak{C}_{\mathrm{III}}$ is defined differently for even and odd n. If n is even, then $\mathfrak{C}_{\mathrm{III}}$ consists of all skew-symmetric unitary matrices of order n. If n is odd, then $\mathfrak{C}_{\mathrm{III}}$ consists of all matrices of the form

$$UDU',$$

where U is an arbitrary unitary matrix and

$$D=\begin{pmatrix}0&1\\-1&0\end{pmatrix}\dotplus\cdots\dotplus\begin{pmatrix}0&1\\-1&0\end{pmatrix}\dotplus 0.$$

(4) $\mathfrak{C}_{IV}$ consists of the vectors $e^{i\theta}x$, where x is a real vector such that $xx'=1$.

The manifolds $\mathfrak{C}_{I}$, $\mathfrak{C}_{II}$, $\mathfrak{C}_{III}$ and $\mathfrak{C}_{IV}$ have real dimension $m(2n-m)$, $n(n+1)/2$, $n(n-1)/2+(1+(-1)^n)(n-1)/2$ and n, respectively.

These characteristic manifolds are homogeneous spaces. Furthermore, any point of $\mathfrak{C}$ can be carried into any other point of $\mathfrak{C}$ by a transformation leaving a given point of $\mathfrak{R}$ invariant. The general theory of harmonic analysis on homogeneous spaces has been developed earlier (cf. E. Cartan [1], H. Weyl [1]); however, the method presented in this book gives more precise and more useful results.

III. **Heuristic considerations.** Suppose that we have a sequence of analytic functions in $\mathfrak{R}$

$$\{\varphi_\nu(z)\}, \qquad \nu=0,1,2,\cdots,$$

such that any analytic function $f(z)$ in $\mathfrak{R}$ can be developed in a series

$$f(z)=\sum_{\nu=0}^{\infty} a_\nu \varphi_\nu(z),$$

convergent in $\mathfrak{R}$. We define the two infinite hermitian matrices

$$H_1=\left(\int_{\mathfrak{C}} \varphi_\nu(\xi)\overline{\varphi_\mu(\xi)}\,\dot{\xi}\right)_{\nu,\mu=0,1,2,\ldots}$$

and

$$H_2=\left(\int_{\mathfrak{R}} \varphi_\nu(z)\overline{\varphi_\mu(z)}\,\dot{z}\right)_{\nu,\mu=0,1,2,\ldots}$$

The basis $\{\varphi_\nu(z)\}$ can be chosen to be orthonormal, such that

$$\int_{\mathfrak{C}} \varphi_\nu(\xi)\overline{\varphi_\mu(\xi)}\,\dot{\xi}=\delta_{\nu\mu}$$

and

$$\int_{\mathfrak{R}} \varphi_\nu(z)\overline{\varphi_\mu(z)}\,\dot{z}=\lambda_\nu\delta_{\nu\mu}.$$

The eigenvalues $\lambda_0,\lambda_1,\lambda_2,\cdots$ are pseudoconformal invariants, i.e., they do not depend on the choice of the basis $\{\varphi_\nu(z)\}$ and are preserved under analytic mappings transforming $\mathfrak{R}$ and $\mathfrak{C}$ into $\mathfrak{R}_1$ and $\mathfrak{C}_1$, respectively.

The existence of a system $\{\varphi_\nu(z)\}$ is known from a theorem of H. Cartan

[1] on complete circular domains.[3]

Now setting

$$K(z, \overline{w}) = \sum_{\nu=0}^{\infty} \frac{\varphi_\nu(z)\overline{\varphi_\nu(w)}}{\lambda_\nu},$$

we obtain the Bergman kernel which has the following reproducing property. For any function $f(z)$ analytic in $\mathfrak{R}$ we have

$$f(z) = \int_{\mathfrak{R}} K(z, \overline{w}) f(w)\, \dot{w}.$$

Setting

$$H(z, \overline{\xi}) = \sum_{\nu=0}^{\infty} \varphi_\nu(z)\, \overline{\varphi_\nu(\xi)},$$

we obtain the Cauchy kernel of the domain $\mathfrak{R}$. This kernel has the reproducing property that for any analytic function $f(z)$ with a series development

$$f(\xi) = \sum_{\nu=0}^{\infty} a_\nu \varphi_\nu(\xi),$$

on $\mathfrak{C}$ we have

$$f(z) = \int_{\mathfrak{C}} H(z, \overline{\xi}) f(\xi)\, \dot{\xi}.$$

Setting

$$f(z) = u(z) H(z, \overline{w}),$$

we have

$$u(z) = \int_{\mathfrak{C}} \frac{H(z, \overline{\xi}) H(\xi, \overline{w})}{H(z, \overline{w})} u(\xi)\, \dot{\xi}.$$

The function

$$P(z, \xi) = \frac{H(z, \overline{\xi}) H(\xi, \overline{z})}{H(z, \overline{z})}$$

is called the Poisson kernel for the domain $\mathfrak{R}$. It is positive.

It is clear that the system of functions $\{\varphi_\nu(\xi)\}$, $\nu = 0, 1, 2, \cdots$ is not complete in the space of continuous functions on $\mathfrak{C}$. We complete it to a com-

[3] Translator's note. The domain $\mathfrak{R}$ in the space of several complex variables is said to be a circular domain (with center at the origin) if together with any point z in $\mathfrak{R}$ the point $ze^{i\varphi}$ is in $\mathfrak{R}$ for any real φ. If together with any point z in $\mathfrak{R}$ also the point $rze^{i\varphi}$ is in $\mathfrak{R}$ for any real φ and $0 \leqq r \leqq 1$, then $\mathfrak{R}$ is said to be a complete circular domain.

plete orthonormal system

$$\{\varphi_\nu(\xi)\}, \qquad \nu=0,\ \pm 1,\ \pm 2, \cdots,$$

and develop the function $P(z, \xi)$ into a Fourier series with respect to this new system

$$P(z, \xi)=\sum_{\nu=-\infty}^{\infty} \Phi_\nu(z) \overline{\varphi_\nu(\xi)}, \quad \Phi_\nu(z)=\int_{\mathfrak{C}} P(z, \xi) \varphi_\nu(\xi) \dot{\xi}.$$

If

$$\lim_{z \to \xi} \Phi_\nu(z)=\varphi_\nu(\xi),$$

then the functions on $\mathfrak{C}$ having a Fourier series development

$$\varphi(\xi)=\sum_{\nu=-\infty}^{\infty} c_\nu \varphi_\nu(\xi), \quad c_\nu=\int_{\mathfrak{C}} \varphi(\xi) \overline{\varphi_\nu(\xi)} \dot{\xi}$$

can be put in correspondence with the class of functions

$$\Phi(z)=\int_{\mathfrak{C}} P(z, \xi) \varphi(\xi) \dot{\xi}=\sum_{\nu=-\infty}^{\infty} c_\nu \Phi_\nu(z),$$

which we shall call harmonic functions in the domain $\mathfrak{R}$.

The harmonic functions can also be defined as solutions of a second order partial differential equation. This equation can be obtained from the following considerations.

The Bergman kernel yields a Riemannian metric on the space $\mathfrak{R}$:

$$d\bar{d} \ln K(z, \bar{z})=\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2}{\partial z_i \partial \bar{z}_j} \ln K(z, \bar{z}) dz_i d\bar{z}_j=\sum_{i, j=1}^{n} h_{ij} dz_i d\bar{z}_j.$$

Corresponding to the tensor h^{ij} we have a differential operator

$$\sum_{i=1}^{n} \sum_{j=1}^{n} h^{ij} \frac{\partial^2}{\partial z_i \partial \bar{z}_j},$$

which we call the Laplace operator of this space.

We can talk about a Dirichlet problem with respect to this operator.

THEOREM 4.1.1. *For a complete circular domain $\mathfrak{R}$ the system of functions*

$$(\beta_\nu^f)^{-\frac{1}{2}}\varphi_\nu^f(z), \quad \begin{array}{l} f=0,\ 1,\ 2,\ \ldots, \\ \nu=1,\ 2,\ \ldots,\ N_f, \end{array} \tag{4.1.6}$$

is a complete orthonormal system in the domain $\mathfrak{R}$. On the other hand, the system $\{\varphi_\nu^f(\xi)\}$ is orthonormal, but in general not complete in the space of functions which are continuous on $\mathfrak{C}$.

It is also well known that the series

$$\sum_{f=0}^{\infty}\sum_{\nu=1}^{N_f}\frac{\varphi_\nu^f(z)\,\overline{\varphi_\nu^f(w)}}{\beta_\nu^f}=K(z,\ \overline{w})$$

converges uniformly for any z and w which lie in the interior of $\mathfrak{R}$, representing there a function called the Bergman kernel.[12]

The sum of the series (if it converges)

$$\sum_{f=0}^{\infty}\sum_{\nu=0}^{N_f}\varphi_\nu^f(z)\,\overline{\varphi_\nu^f(\xi)}=H(z,\ \overline{\xi})$$

we shall call the Cauchy kernel for the domain $\mathfrak{R}$.

Finally, we shall call the function

$$P(z,\ \xi)=\frac{|H(z,\ \overline{\xi})|^2}{H(z,\ \overline{z})}$$

the Poisson kernel for the domain $\mathfrak{R}$.

This chapter deals with the direct methods of determination of these kernels.

4.2. **The Bergman kernel.** Let $\mathfrak{R}$ be a bounded domain which contains the origin, Γ a group of analytic mappings of $\mathfrak{R}$ onto itself, and Γ_0 a subgroup of Γ which leaves the origin fixed. It is well known (H. Cartan [1]) that an element of Γ_0 is fully determined by its linear terms in the neighborhood of the origin, i.e., the mapping of $\mathfrak{R}$ onto itself which has the form

$$w_i=\sum_{j=1}^{n}u_{ij}z_j+\sum_{\substack{m_1\ldots m_n\\ m_1+\ldots+m_n\geqslant 2}}a_{m_1\ldots m_n}^{(i)}z_1^{m_1}\ldots z_n^{m_n}, \tag{4.2.1}$$

[12] Translator's note. In the Russian literature the Bergman kernel is usually called the "kernel function of the domain". In this book such a designation would not be very appropriate since we are dealing with three kernels to which this designation could apply.

is fully determined if the matrix $(u_{ij})_1^n$ is given. As it is well known that Γ_0 is compact, it can be assumed without loss of generality that the matrices $(u_{ij}) = U$ which form the representation of Γ_0 are unitary. The letter U we shall also use for denoting the nonlinear transformation (4.2.1) itself, determined by the linear element U.

Let us now consider the set of cosets of Γ/Γ_0. All group transformations belonging to one and the same coset carry into the origin one and the same point a. The totality of all such points a forms in $\mathfrak{R}$ some set $\mathfrak{M}$. It is called a transitive set with respect to the group Γ which contains the origin. Thus any element of Γ is uniquely determined by a point a of $\mathfrak{M}$, and by the unitary matrix U of Γ_0. We shall write the transformations determined by the elements of Γ in the form

$$w = f(z, a, U), \quad a \in \mathfrak{M}, \quad U \in \Gamma_0. \tag{4.2.2}$$

Suppose

$$z = f(x, b, V), \quad b \in \mathfrak{M}, \quad V \in \Gamma_0, \tag{4.2.3}$$

is another transformation, and

$$w = f(f(x, b, V), a, U) = f(x, c, W) \tag{4.2.4}$$

is the product of the transformations (4.2.2) and (4.2.3). Setting $w = 0$, we at once obtain

$$a = f(c, b, V). \tag{4.2.5}$$

Differentiation of (4.2.4) yields

$$\frac{\partial f_i(x, c, W)}{\partial x_j} = \sum_{k=1}^{n} \frac{\partial f_i(z, a, U)}{\partial z_k} \frac{\partial f_k(x, b, V)}{\partial x_j}. \tag{4.2.6}$$

We shall denote the Jacobian of the transformation (4.2.2) by

$$J(z, a, U) = (a_{ij}), \quad a_{ij} = \frac{\partial f_i(z, a, U)}{\partial z_j}.$$

By setting $x = c$ in (4.2.6) we obtain $z = a$. Hence

$$J(c, c, W) = J(a, a, U) \cdot J(c, b, V).$$

By a change of notation we obtain

$$J(x, x, W) = J(z, z, U) J(x, b, V). \tag{4.2.7}$$

This formula is valid for x and z of $\mathfrak{M}$ which satisfy the relation

$$z = f(x, b, V). \tag{4.2.8}$$

If we have another transformation

$$u = f(x, b, V_0),$$

then the mapping of u into z leaves the origin unchanged. Hence

$$U_0 = \left(\frac{\partial z}{\partial u}\right)_{z=0}$$

is a unitary matrix. Whence follows that

$$\{J(x, b, V)\}_{x=b} = \left(\frac{\partial z}{\partial u}\right)_{z=0} \cdot \{J(x, b, V_0)\}_{x=b},$$

so that we have

$$J(b, b, V) = U_0 J(b, b, V_0), \tag{4.2.9}$$

where U_0 is the unitary matrix of Γ_0. Thus

$$\overline{J(z, z, V)'} \cdot J(z, z, V) = \overline{J(z, z, V_0)'} \cdot J(z, z, V_0). \tag{4.2.10}$$

This shows that $\overline{J'}J$ depends on the coset of Γ/Γ_0 but does not depend on the choice of representative of this coset. Hence we can write

$$|\det J(z, z, V)|^2 = Q(z, \bar{z}).$$

From (4.2.7) we obtain for z and x, lying in $\mathfrak{M}$ and satisfying the relation (4.2.8), the formula

$$Q(x, \bar{x}) = Q(z, \bar{z}) |\det J(x, b, V)|^2. \tag{4.2.11}$$

Bergman [1] has proved that under the transformation (4.2.8) the Bergman kernel of the domain $\mathfrak{R}$ changes according to the law

$$K(x, \bar{x}) = K(z, \bar{z}) |\det J(x, b, V)|^2. \tag{4.2.12}$$

Thus for z and x of $\mathfrak{M}$

$$\frac{K(x, \bar{x})}{Q(x, \bar{x})} = \frac{K(z, \bar{z})}{Q(z, \bar{z})}. \tag{4.2.13}$$

THEOREM 4.2.1. *If $\mathfrak{R}$ is a bounded circular domain, then for z lying in $\mathfrak{M}$, we have*

$$K(z, \bar{z}) = \frac{1}{\Omega} Q(z, \bar{z}),$$

where Ω is the complete volume of $\mathfrak{R}$.

PROOF. In view of §4.1, we can propose the following process of constructing an orthonormal system of functions.

We orthonormalize the terms

$$z_1^{a_1} \cdot z_2^{a_2} \cdot \ldots \cdot z_n^{a_n}, \qquad a_1 + \ldots + a_n = m,$$

for a given m, and we take the totality of all such functions for $m = 0, 1, 2, \cdots$. This totality forms a complete orthonormal system.

Among the functions $\varphi_\nu(z)$ obtained by this process, we have the constant $\Omega^{-1/2}$, whereas the other functions are homogeneous forms of order $m \geqq 1$. Hence

$$\varphi_0(z) = \Omega^{-\frac{1}{2}}, \qquad \varphi_\nu(0) = 0, \qquad \nu \geqq 1.$$

Therefore from the equation

$$K(z, \bar{z}) = \sum_{\nu=0}^{\infty} \varphi_\nu(z) \overline{\varphi_\nu(z)}$$

we obtain at once

$$K(0, 0) = \frac{1}{\Omega}.$$

On the other hand, by the definition of $Q(z, \bar{z})$ we have $Q(0, 0) = 1$. Hence the theorem follows from (4.2.13).

Assuming now that $\mathfrak{R}$ is a transitive domain (i.e., $\mathfrak{R} \equiv \mathfrak{M}$), let us ascertain the geometrical properties of $Q(z, \bar{z})$. From (4.2.7) we have

$$J(x, x, W)dx' = J(z, z, U)dz',$$

hence

$$\overline{dx} \cdot \overline{J(x, x, W)'} \cdot J(x, x, W) \cdot dx' = \overline{dz} \cdot \overline{J(z, z, U)'} \cdot J(z, z, U) \cdot dz'.$$

This invariant form can be considered as the metric of our space. The volume element in this metric is

$$|\det J(z, z, U)|^2 \dot{z} = Q(z, \bar{z}) \cdot \dot{z},$$

so that $Q(z, \bar{z})$ can be called the *volume density*.

From Theorem 4.2.1 we obtain the following proposition:

The Bergman kernel for any transitive circular region is equal to the ratio of the volume density to the Euclidean volume of the domain.

In the subsequent sections we shall determine the Bergman kernel for our four types of classical domains on the basis of the above considerations

only, without the use of complete orthonormal systems.[13]

4.3. Bergman kernels for the domains $\mathfrak{R}_{\mathrm{I}}$, $\mathfrak{R}_{\mathrm{II}}$ and $\mathfrak{R}_{\mathrm{III}}$.

1°. The group Γ for the domain $\mathfrak{R}_{\mathrm{I}}$ consists of the following transformations (see L. K. Hua [1]):

$$Z_1=(AZ+B)(CZ+D)^{-1}, \tag{4.3.1}$$

where A, B, C, D are matrices of dimensions $m\times m$, $m\times n$, $n\times m$ and $n\times n$, respectively, satisfying the relations

$$\bar{A}A'-\bar{B}B'=I^{(m)}, \quad \bar{A}C'=\bar{B}D', \quad \bar{C}C'-\bar{D}D'=-I^{(n)}.$$

For $m=n$, we assume, moreover, that

$$\det\begin{pmatrix} A & B \\ C & D\end{pmatrix}=+1. \tag{4.3.2}$$

Let us find the transformations which carry an arbitrary point $Z=P$ into the origin. By the definition of the domain $\mathfrak{R}_{\mathrm{I}}$ we have

$$I^{(m)}-\bar{P}P'>0.$$

Thus it follows from Theorem 2.1.2 that also

$$I^{(n)}-P'\bar{P}>0.$$

It is known that there exists an $m\times m$ matrix Q and $n\times n$ matrix R such that

$$\bar{Q}\left(I^{(m)}-\bar{P}P'\right)Q'=I^{(m)}, \quad \bar{R}\left(I^{(n)}-P'\bar{P}\right)R'=I^{(n)}. \tag{4.3.3}$$

The transformation

$$Z_1=Q(Z-P)(I^{(n)}-\bar{P}'Z)^{-1}R^{-1} \tag{4.3.4}$$

carries P into the origin. It is easy to see that this transformation is of the form (4.3.1).

Differentiating (4.3.4), we obtain

$$dZ_1=Q\left\{dZ\cdot(I-\bar{P}'Z)^{-1}+(Z-P)\,d\,(I-\bar{P}'Z)^{-1}\right\}R^{-1}.$$

We shall set $Z=P$. Then

$$dZ_1=Q\cdot dZ\cdot(I-\bar{P}'P)^{-1}R^{-1}=Q\cdot dZ\cdot\bar{R}',$$

i.e., at the point $Z=P$

$$\dot{Z}_1=\left|(\det Q)^m\cdot(\det\bar{R}')^n\right|^2\cdot\dot{Z}=\left\{\det(I-P\bar{P}')\right\}^{-m-n}\cdot\dot{Z}.$$

[13] Translator's note. The 4 domains in question, $\mathfrak{R}_{\mathrm{I}}$, $\mathfrak{R}_{\mathrm{II}}$, $\mathfrak{R}_{\mathrm{III}}$ and $\mathfrak{R}_{\mathrm{IV}}$, are defined by the author in the introduction to the book.

Hence

$$Q(Z, \bar{Z}) = \{\det(I - Z\bar{Z}')\}^{-(m+n)}.$$

Using the results of §4.2, we obtain the following theorem.

THEOREM 4.3.1. *The Bergman kernel of the domain* $\mathfrak{R}_{\mathrm{I}}$ *is*

$$\frac{1}{V(\mathfrak{R}_{\mathrm{I}})} \cdot \{\det(I - Z\bar{Z}')\}^{-(m+n)}, \tag{4.3.5}$$

where, by (2.2.2)

$$V(\mathfrak{R}_{\mathrm{I}}) = \frac{1!\,2!\ldots(m-1)!\,1!\,2!\ldots(n-1)!}{1!\,2!\ldots(m+n-1)!}\,\pi^{mn}.$$

2°. The group Γ of the domain $\mathfrak{R}_{\mathrm{II}}$ consists of transformations of the form

$$Z_1 = (AZ + B)(\bar{B}Z + \bar{A})^{-1}, \tag{4.3.6}$$

where

$$A'B = B'A, \quad \bar{A}A' - \bar{B}B' = I.$$

Suppose P is a point of $\mathfrak{R}_{\mathrm{II}}$. A matrix R can be found such that

$$\bar{R}(I - \bar{P}P')R' = I. \tag{4.3.7}$$

The transformation

$$Z_1 = R(Z - P)(I - \bar{P}Z)^{-1}\bar{R}^{-1}, \tag{4.3.8}$$

belonging to Γ, carries the point P into the origin.

Differentiation of (4.3.8) yields

$$dZ_1 = R\{dZ \cdot (I - \bar{P}Z)^{-1} + (Z - P)\,d(I - \bar{P}Z)^{-1}\}\,\bar{R}^{-1}.$$

Setting $Z = P$, we obtain

$$dZ_1 = R \cdot dZ \cdot (I - \bar{P}P)^{-1}\bar{R}^{-1} = R \cdot dZ \cdot R'.$$

Hence at the point $Z = P$

$$\dot{Z}_1 = |(\det R)^{n+1}|^2 \cdot \dot{Z} = \{\det(I - P\bar{P})\}^{-(n+1)} \cdot \dot{Z}.$$

Thus

$$Q(Z, \bar{Z}) = \{\det(I - Z\bar{Z})\}^{-(n+1)}.$$

THEOREM 4.3.2. *The Bergman kernel of the domain* $\mathfrak{R}_{\mathrm{II}}$ *is*

$$\frac{1}{V(\mathfrak{R}_{\mathrm{II}})} \cdot \{\det(I - Z\bar{Z})\}^{-(n+1)}, \tag{4.3.9}$$

where, by (2.3.2),

$$V(\mathfrak{R}_{II}) = \pi^{\frac{n(n+1)}{2}} \cdot \frac{2!\,4!\ldots(2n-2)!}{n!\,(n+1)!\ldots(2n-1)!}.$$

3°. The group Γ of the domain $\mathfrak{R}_{III}$ consists of transformations of the form

$$Z_1 = (AZ+B)(-\bar{B}Z+\bar{A})^{-1}, \tag{4.3.10}$$

where

$$A'B = -B'A, \quad \bar{A}'A - \bar{B}'B = I.$$

Suppose P is a point of $\mathfrak{R}_{III}$, i.e., $I + P\bar{P} > 0$. Then a matrix Q can be found such that

$$\bar{Q}(I+P\bar{P})Q' = I.$$

Then in Γ we have the transformation

$$Z_1 = Q(Z-P)(I+\bar{P}Z)^{-1}\bar{Q}^{-1}, \tag{4.3.11}$$

which carries the point P into the origin.

By differentiation of (4.3.11) we have

$$dZ_1 = Q\{dZ\cdot(I+\bar{P}Z)^{-1} + (Z-P)\,d(I+\bar{P}Z)^{-1}\}\,\bar{Q}^{-1}.$$

For $Z=P$, we obtain

$$dZ_1 = Q\cdot dZ\cdot(I+\bar{P}P)^{-1}\cdot\bar{Q}^{-1} = Q\cdot dZ\cdot Q'.$$

Hence at the point $Z=P$

$$\dot{Z}_1 = |(\det Q)^{n-1}|^2\,\dot{Z} = \{\det(I+\bar{P}P)\}^{-n+1}\cdot\dot{Z}.$$

Therefore

$$Q(Z,\bar{Z}) = \{\det(I+Z\bar{Z})\}^{-n+1}.$$

THEOREM 4.3.3. *The Bergman kernel of the domain* $\mathfrak{R}_{III}$ *is*

$$\frac{1}{V(\mathfrak{R}_{III})}\{\det(I+Z\bar{Z})\}^{-n+1},$$

where, by (2.4.2),

$$V(\mathfrak{R}_{III}) = \pi^{\frac{n(n-1)}{2}} \cdot \frac{2!\,4!\ldots(2n-4)!}{(n-1)!\,n!\ldots(2n-3)!}.$$

4.4. **The Bergman kernel for the domain** $\mathfrak{R}_{IV}$. The group Γ of the domain $\mathfrak{R}_{IV}$ consists of transformations of the form

$$w=\left\{\left[\left(\tfrac{1}{2}(zz'+1),\ \tfrac{i}{2}(zz'-1)\right)A'+zB'\right]\begin{pmatrix}1\\ i\end{pmatrix}\right\}^{-1}$$
$$\times\left\{\left(\tfrac{1}{2}(zz'+1),\ \tfrac{i}{2}(zz'-1)\right)C'+zD'\right\}, \tag{4.4.1}$$

where A, B, C and D are real matrices of dimensions 2×2, $2\times n$, $n\times 2$ and $n\times n$, respectively, satisfying the relations

$$\begin{pmatrix}A & B\\ C & D\end{pmatrix}\begin{pmatrix}I^{(2)} & 0\\ 0 & -I^{(n)}\end{pmatrix}\begin{pmatrix}A & B\\ C & D\end{pmatrix}'=\begin{pmatrix}I^{(2)} & 0\\ 0 & -I^{(n)}\end{pmatrix} \tag{4.4.2}$$

and

$$\det\begin{pmatrix}A & B\\ C & D\end{pmatrix}=+1. \tag{4.4.3}$$

We shall now find the transformations of Γ which carry the point z_0 into the origin. Proceeding from the vector z_0 we shall construct the $2\times n$ matrix X_0 as follows:

$$X_0=2\begin{pmatrix}z_0z_0'+1 & i\,(z_0z_0'-1)\\ \overline{z_0z_0'}+1 & -i\,(\overline{z_0z_0'}-1)\end{pmatrix}^{-1}\begin{pmatrix}z_0\\ \bar{z}_0\end{pmatrix}=2A_0^{-1}\begin{pmatrix}z_0\\ \bar{z}_0\end{pmatrix}$$
$$=\frac{1}{1-|z_0z_0'|^2}\begin{pmatrix}z_0+\bar{z}_0-(\overline{z_0z_0'}\cdot z_0+z_0z_0'\cdot\bar{z}_0)\\ i\,(z_0-\bar{z}_0)+i\,(\overline{z_0z_0'}\cdot z_0-z_0z_0'\cdot\bar{z}_0)\end{pmatrix}. \tag{4.4.4}$$

This matrix is evidently real. We have

$$I-X_0X_0'=\bar{A}_0^{-1}\left(\bar{A}_0A_0'-4\overline{\begin{pmatrix}z_0\\ \bar{z}_0\end{pmatrix}}\begin{pmatrix}z_0\\ \bar{z}_0\end{pmatrix}'\right)\cdot A_0'^{-1}$$
$$=2\bar{A}_0^{-1}\begin{pmatrix}1+|z_0z_0'|^2-2\bar{z}_0z_0' & 0\\ 0 & 1+|z_0z_0'|^2-2\bar{z}_0z_0'\end{pmatrix}\cdot A_0'^{-1}. \tag{4.4.5}$$

Hence

$$(I-X_0X_0')^{-1}=\frac{1}{2\,(1+|z_0z_0'|^2-2\bar{z}_0z_0')}A_0'\bar{A}_0=\frac{1}{1+|z_0z_0'|^2-2\bar{z}_0z_0'}$$
$$\times\begin{pmatrix}(z_0z_0'+1)(\overline{z_0z_0'}+1) & i\,(z_0z_0'-\overline{z_0z_0'})\\ i\,(z_0z_0'-\overline{z_0z_0'}) & (z_0z_0'-1)(\overline{z_0z_0'}-1)\end{pmatrix}=A'A, \tag{4.4.6}$$

where

$$A=\frac{1}{2}\left(1+|z_0z_0'|^2-2\bar{z}_0z_0'\right)^{-\frac{1}{2}}\begin{pmatrix}-i\left(z_0z_0'-\overline{z_0z_0'}\right) & z_0z_0'+\overline{z_0z_0'}-2\\ z_0z_0'+\overline{z_0z_0'}+2 & i\left(z_0z_0'-\overline{z_0z_0'}\right)\end{pmatrix}. \tag{4.4.7}$$

We shall choose a D which satisfies the condition $D(I^{(n)}-X_0'X_0)D'=I^{(n)}$; then the transformation

$$w=\left\{\left[\left(\frac{1}{2}(zz'+1),\ \frac{i}{2}(zz'-1)\right)A'-zX_0'A'\right]\begin{pmatrix}1\\ i\end{pmatrix}\right\}^{-1}$$
$$\times\left\{zD'-\left(\frac{1}{2}(zz'+1),\ \frac{i}{2}(zz'-1)\right)X_0D'\right\} \tag{4.4.8}$$

has the form (4.4.1) and carries the point z_0 into the origin.

In addition,

$$\det A=\det D=\frac{1-|z_0z_0'|^2}{1+|z_0z_0'|^2-2\bar{z}_0z_0'}. \tag{4.4.9}$$

Differentiation of (4.4.8) yields $(z=(z^{(1)},\cdots,z^{(n)}))$

$$dw=\left\{dz\cdot D'-\left(\sum_{p=1}^{n}z^{(p)}dz^{(p)},\ i\sum_{p=1}^{n}z^{(p)}dz^{(p)}\right)X_0D'\right\}$$
$$\times\left\{\left[\left(\frac{1}{2}(zz'+1),\ \frac{i}{2}(zz'-1)\right)A'-zX_0'A'\right]\begin{pmatrix}1\\ i\end{pmatrix}\right\}^{-1}$$
$$+\left\{zD'-\left(\frac{1}{2}(zz'+1),\ \frac{i}{2}(zz'-1)\right)X_0D'\right\}$$
$$\times d\left\{\left[\left(\frac{1}{2}(zz'+1),\ \frac{i}{2}(zz'-1)\right)A'-zX_0'A'\right]\begin{pmatrix}1\\ i\end{pmatrix}\right\}^{-1}.$$

Setting $z=z_0$, we obtain

$$dw=\left\{dz\cdot D'-\left(\sum_{p=1}^{n}z_0^{(p)}dz^{(p)},\ i\sum_{p=1}^{n}z_0^{(p)}dz^{(p)}\right)X_0D'\right\}$$
$$\times\left\{\left[\left(\frac{1}{2}(z_0z_0'+1),\ \frac{i}{2}(z_0z_0'-1)\right)A'-z_0X_0'A'\right]\begin{pmatrix}1\\ i\end{pmatrix}\right\}^{-1}.$$

i.e.,

$$dw=\{dz\cdot D'-d(zz')\cdot(1,i)X_0D'\}$$
$$\times\left\{\left[\left(\frac{1}{2}(z_0z_0'+1),\ \frac{i}{2}(z_0z_0'-1)\right)A'-z_0X_0'A'\right]\begin{pmatrix}1\\ i\end{pmatrix}\right\}^{-1} \tag{4.4.10}$$

Using (4.4.4) and (4.4.7), we have

$$dw=-i\,dz\cdot\left\{I-2\frac{z_0'\bar{z}_0-\overline{z_0z_0'}\cdot z_0'z_0}{1-|z_0z_0'|^2}\right\}\cdot D'\cdot\left(1+|z_0z_0'|^2-2\bar{z}_0z_0'\right)^{-\frac{1}{2}}. \tag{4.4.11}$$

From (4.4.9) we obtain

$$\det\left(\frac{\partial w}{\partial z}\right)_{z=z_0}$$
$$=\det\left\{I-2\frac{z_0'\bar{z}_0-\overline{z_0z_0'}\cdot z_0'z_0}{1-|z_0z_0'|^2}\right\}\cdot\frac{1-|z_0z_0'|^2}{(1+|z_0z_0'|^2-2\bar{z}_0z_0')}$$
$$\times(1+|z_0z_0'|^2-2\bar{z}_0z_0')^{-\frac{n}{2}},$$

or

$$\left|\det\left(\frac{\partial w}{\partial z}\right)_{z=z_0}\right|^2=(1+|z_0z_0'|^2-2\bar{z}_0z_0')^{-n}. \tag{4.4.12}$$

Here the identity

$$\det\left\{I-2\frac{z_0'\bar{z}_0-\overline{z_0z_0'}\cdot z_0'z_0}{1-|z_0z_0'|^2}\right\}=\frac{1+|z_0z_0'|^2-2\bar{z}_0z_0'}{1-|z_0z_0'|^2} \tag{4.4.13}$$

is used, which follows from the relation $\det(I-u'\bar{v})=1-\bar{v}u'$ (see Theorem 2.1.2).

Thus we arrived at the theorem:

THEOREM 4.4.1. *The Bergman kernel of the domain* $\mathfrak{R}_{\mathrm{IV}}$ *is*

$$\frac{1}{V(\mathfrak{R}_{\mathrm{IV}})}(1+|zz'|^2-2\bar{z}z')^{-n},$$

where, by (2.5.7),

$$V(\mathfrak{R}_{\mathrm{IV}})=\frac{\pi^n}{2^{n-1}\cdot n!}.$$

4.5. **The Cauchy kernel.** Let us now pass to the study of the Cauchy kernel

$$\sum_{f=0}^{\infty}\sum_{\nu=1}^{N_f}\varphi_\nu^f(z)\,\overline{\varphi_\nu^f(\xi)}=H(z,\bar{\xi}). \tag{4.5.1}$$

(Here z belongs to $\mathfrak{R}$, and ξ to $\mathfrak{C}$.)

Suppose Γ_0 is a group of motions of $\mathfrak{R}$ which leave the origin unchanged. We shall assume that $\mathfrak{C}$ is transitive with respect to Γ_0, i.e., that any two points of $\mathfrak{C}$ can be carried into each other by a transformation which belongs to Γ_0.

THEOREM 4.5.1. *The series*

This transformation carries $\mathfrak{C}$ into itself. We shall set

$$\zeta = f(\xi, a, U). \tag{4.6.4}$$

Then

$$\dot{\zeta} = B(\xi, a, U)\dot{\xi}. \tag{4.6.5}$$

(This formula is valid, since the real dimensionality of $\mathfrak{C}$ is equal to the complex dimensionality of $\mathfrak{R}$, thus enabling us to effect the passage from the boundary to the interior of the domain by a simple replacement of real parameters by complex ones.)

We know that on $\mathfrak{C}$ there exists an orthonormal system $\{\varphi_\nu^f(\xi)\}$ which we shall denote, for the sake of simplicity, by $\{\varphi_\nu(\xi)\}$. Then

$$\int_{\mathfrak{C}} \varphi_\mu(\zeta)\overline{\varphi_\nu(\zeta)\dot{\zeta}} = \int_{\mathfrak{C}} \varphi_\mu(f(\xi))\overline{\varphi_\nu(f(\xi))}\,|B(\xi, a, U)|\,\dot{\xi} = \delta_{\mu\nu}.$$

Hence the system

$$\psi_\mu(\xi) = \varphi_\mu(f(\xi, a, U))\,B^{\frac{1}{2}}(\xi, a, U)$$

is also orthonormal. Let us prove that the system $\{\psi_\mu(\xi)\}$ satisfies the conditions of Theorem 4.6.2. It is evident that $\psi_\mu(z)$ is analytic in $\mathfrak{R}$ and on its boundary, and

$$\sum_{\nu=0}^{\infty} \psi_\nu(z)\overline{\psi_\nu(\xi)} = \sum_{\nu=0}^{\infty} \varphi_\nu(w)\overline{\varphi_\nu(\zeta)} \cdot B^{\frac{1}{2}}(z, a, U)\,\overline{B^{\frac{1}{2}}(\xi, a, U)}. \tag{4.6.6}$$

Let us denote by

$$z = f^{-1}(w, a, U)$$

the inverse transformation of (4.6.3). To any function $\psi(z)$ which is analytic in $\mathfrak{R}$ and on its boundary, we shall set in correspondence the function

$$\varphi(w) = \psi\left(f^{-1}(w, a, U)\right) B^{-\frac{1}{2}}(z, a, U),$$

also analytic in $\mathfrak{R}$ and on its boundary. Since $\varphi(w)$ can be expanded in the series

$$\varphi(w) = \sum_{\nu=0}^{\infty} a_\nu \varphi_\nu(w),$$

we also obtain

$$\psi(z) = \sum_{\nu=0}^{\infty} a_\nu \psi_\nu(z).$$

This shows that condition (3) of Theorem 4.6.2 is satisfied. Thus, from (4.6.6) follows

$$H(z, \bar{\xi}) = H(w, \bar{\zeta}) \cdot B^{\frac{1}{2}}(z, a, U)\, \overline{B^{\frac{1}{2}}(\xi, a, U)}.$$

Since $H(0, \bar{\zeta}) = [V(\mathfrak{C})]^{-1}$, we obtain the assertion of the theorem after replacing a by z.

REMARK. The orthonormal system $\{\psi_\mu(\xi)\}$ considered above is not complete on $\mathfrak{C}$. There also exists a complete orthonormal system on $\mathfrak{C}$ (see H. Weyl [1]); it can be obtained by supplementing $\{\psi_\mu(\xi)\}$ by some system of functions $\psi_{-\nu}(\xi)$, $\nu = 1, 2, \cdots$. Here we are considering completeness in the space of functions which are continuous on $\mathfrak{C}$, i.e., if $g(\xi)$ is a continuous function on $\mathfrak{C}$, then from

$$\int_{\mathfrak{C}} g(\xi)\, \overline{\varphi_\nu(\xi)}\, \dot{\xi} = 0 \qquad (\nu = \pm 1, \pm 2, \ldots)$$

it follows that $g(\xi) \equiv 0$.

4.7. **The Cauchy kernels for classical domains.** By applying Theorem 4.6.3 it is possible to directly obtain the Cauchy kernels for classical domains.

1°. In $\mathfrak{R}_{\mathrm{I}}$, the characteristic manifold $\mathfrak{C}_{\mathrm{I}}$ is determined by the condition $U\bar{U}' = I$. Let us first assume that $m = n$. Then the dimensionality of the characteristic manifold equals half the dimensionality of $\mathfrak{R}_{\mathrm{I}}$, and hence

$$H(Z, \bar{U}) = \frac{1}{V(\mathfrak{C}_{\mathrm{I}})} \cdot \{\det(I - Z\bar{U}')\}^{-n}, \tag{4.7.1}$$

where by Theorem 3.1.1

$$V(\mathfrak{C}_{\mathrm{I}}) = \frac{(2\pi)^{\frac{n(n+1)}{2}}}{1!\,2!\ldots(n-1)!}.$$

If $m \neq n$, then, assuming for definiteness that $m < n$, we have

$$H(Z, \bar{U}) = \frac{1}{V(\mathfrak{C}_{\mathrm{I}})} [\det(I - Z\bar{U}')]^{-n}, \tag{4.7.2}$$

where (see §5.4)

$$V(\mathfrak{C}_{\mathrm{I}}) = \frac{(2\pi)^{mn - \frac{m(m-1)}{2}}}{(n-m)!\,(n-m+1)!\ldots(n-1)!}.$$

It is simpler to obtain expression (4.7.2) from (4.7.1) than directly from Theorem 4.6.3. In fact, we have

$$f(Z)=\frac{1}{V(\mathfrak{U}_n)}\int_{\mathfrak{U}_n} f(U_n)\left[\det\left(I-Z\overline{U}'_n\right)\right]^{-n}\dot{U}_n, \tag{4.7.3}$$

where the integral is taken over the set of all unitary matrices of order n. Setting

$$Z=\begin{pmatrix} Z_1 \\ 0 \end{pmatrix}, \text{ where } Z_1 - m\times n\text{- matrix}, \quad U_n=\begin{pmatrix} U_{mn} \\ V \end{pmatrix},$$

we obtain

$$f\left[\begin{pmatrix} Z_1 \\ 0 \end{pmatrix}\right]=\frac{1}{V(\mathfrak{U}_n)}\int_{U_{mn}}\int_{V} f\left[\begin{pmatrix} U_{mn} \\ V \end{pmatrix}\right]\left[\det\left(I-Z_1\overline{U}'_{mn}\right)\right]^{-n}\dot{U}_{mn}\dot{V}$$

$$=\frac{1}{V(\mathfrak{U}_n)}\int_{U_{mn}}\left[\det\left(I-Z_1\overline{U}'_{mn}\right)\right]^{-n}\dot{U}_{mn}\int_{V} f\left[\begin{pmatrix} U_{mn} \\ V \end{pmatrix}\right]\dot{V}, \tag{4.7.4}$$

where V runs through the set of all $(n-m)\times n$ matrices which satisfy the conditions

$$V\overline{V}'=I^{(n-m)}, \quad U_{mn}\overline{V}'=0. \tag{4.7.5}$$

For a fixed matrix U_{mn} we also find two unitary matrices, namely, the $m\times n$ matrix P and the $n\times n$ matrix Q, such that

$$PU_{mn}Q=\left(I^{(m)}, 0\right).$$

Hence it follows that $(I^{(m)}, 0)\overline{Q'V'}=0$, i.e., $VQ=(0, W)$, where W is the unitary matrix of order $n-m$. Therefore

$$\int_{V} f\left[\begin{pmatrix} U_{mn} \\ V \end{pmatrix}\right]\dot{V}=\int_{W} f\left[\begin{pmatrix} U_{mn} \\ (0, W) \end{pmatrix}\right]\dot{W}.$$

By (4.7.3), where n has been replaced by $n-m$, and $Z=0$, we have

$$\int_{V} f\left[\begin{pmatrix} U_{mn} \\ V \end{pmatrix}\right]\dot{V}=V(\mathfrak{U}_{n-m})\, f\left[\begin{pmatrix} U_{mn} \\ 0 \end{pmatrix}\right].$$

From (4.7.4) it now follows that

$$f\left[\begin{pmatrix} Z_1 \\ 0 \end{pmatrix}\right]=\frac{V(\mathfrak{U}_{n-m})}{V(\mathfrak{U}_n)}\int_{U_{mn}} f\left[\begin{pmatrix} U_{mn} \\ 0 \end{pmatrix}\right]\left[\det\left(I-Z\overline{U}'_{mn}\right)\right]^{-n}\dot{U}_{mn},$$

from which we obtain formula (4.7.2) for the Cauchy kernel.

2°. The characteristic manifold of the domain $\mathfrak{R}_{\mathrm{II}}$ is determined from the condition $U\overline{U}=I$, i.e., it coincides with the set of symmetric unitary matrices. Since the dimensionality of $\mathfrak{C}_{\mathrm{II}}$ is half the dimensionality of $\mathfrak{R}_{\mathrm{II}}$, Theorem 4.6.3 yields at once

$$H(Z,\overline{U})=\frac{1}{V(\mathfrak{C}_{\mathrm{II}})}\cdot[\det(I-Z\overline{U})]^{-\frac{n+1}{2}}, \tag{4.7.6}$$

where (see §3.5)

$$V(\mathfrak{C}_{\mathrm{II}})=2^{\frac{n(3n+1)}{4}}\cdot\pi^{\frac{n(n+1)}{4}}\cdot\frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}\cdot\prod_{\nu=1}^{n-1}\frac{\Gamma\left(\frac{n-\nu}{2}+1\right)}{\Gamma(n-\nu+1)}.$$

3°. The characteristic manifold of the domain $\mathfrak{R}_{\mathrm{III}}$ coincides with the set of matrices K defined by the equality (3.6.16).[14]

For even n

$$H(Z,\overline{K})=\frac{1}{V(\mathfrak{C}_{\mathrm{III}})}\cdot[\det(I+Z\overline{K})]^{-\frac{n-1}{2}}, \tag{4.7.7}$$

where (see §3.7)

$$V(\mathfrak{C}_{\mathrm{III}})=\frac{1}{2^{\nu-1}}\cdot(8\pi)^{\frac{n(n-1)}{4}}\cdot\prod_{s=1}^{n-1}\frac{\Gamma\left(\frac{s}{2}\right)}{\Gamma(s)},\quad \nu=\frac{n}{2}.$$

For odd n

$$H(Z,\overline{K})=\frac{1}{V(\mathfrak{C}_{\mathrm{III}})}\cdot\{\det(I+Z\overline{K})\}^{-\frac{n}{2}}, \tag{4.7.8}$$

where

$$V(\mathfrak{C}_{\mathrm{III}})=2\pi\frac{1}{2^{\nu-1}}\cdot(8\pi)^{\frac{n(n+1)}{4}}\cdot\prod_{s=1}^{n}\frac{\Gamma\left(\frac{s}{2}\right)}{\Gamma(s)},\quad \nu=\frac{n+1}{2}.$$

Equality (4.7.7) is obtained directly from Theorem 4.6.3, whereas equality (4.7.8) will now be derived from (4.7.7). Replacing n by $n+1$, we shall write Cauchy's formula with the kernel (4.7.7)

$$f(Z)=\frac{1}{V}\cdot\int_K f(K)\{\det(I+Z\overline{K})\}^{-\frac{n}{2}}\dot{K}, \tag{4.7.9}$$

[14] Note by the translator into English. This equality, found in Chapter III, defines the set of skew-symmetric matrices K.

where Z and K are matrices of order $(n+1)$, and $V=V(\mathfrak{C}_{\mathrm{III}}^{(n+1)})$. We shall set

$$Z=\begin{pmatrix}0 & 0\\ 0 & Z_1\end{pmatrix}, \text{ where } Z_1 \text{ is an } n\times n \text{ matrix,}$$

and accordingly:

$$K=\begin{pmatrix}0 & k\\ -k' & K_1\end{pmatrix},$$

where K_1 is an $n\times n$ matrix and k is a $1\times n$ matrix. Then

$$f\left(\begin{pmatrix}0 & 0\\ 0 & Z_1\end{pmatrix}\right)=\frac{1}{V}\int\limits_{K_1}\{\det(I+Z_1\bar{K}_1)\}^{-\frac{n}{2}}\dot{K}_1\int\limits_{k} f\left(\begin{pmatrix}0 & k\\ -k' & K_1\end{pmatrix}\right)\dot{k}. \quad (4.7.10)$$

Since K is a unitary matrix, it follows that

$$k\bar{k}'=1,\ k\bar{K}_1'=0,\ k'\bar{k}+K_1\bar{K}_1'=I^{(n)}.$$

Since the matrix $I-K_1\bar{K}_1'=k'\bar{k}$ has rank one and, furthermore,

$$(I-K_1\bar{K}_1')^2=k'\bar{k}k'\bar{k}=k'\bar{k}=I-K_1\bar{K}_1',$$

a unitary matrix U can be found such that

$$U(I-K_1\bar{K}_1')\bar{U}'=[1,\ 0,\ \ldots,\ 0],$$

i.e.,

$$UK_1\bar{K}_1'\bar{U}'=[0,\ 1,\ \ldots,\ 1]=\begin{pmatrix}0 & 0\\ 0 & F\end{pmatrix}\overline{\begin{pmatrix}0 & 0\\ 0 & F\end{pmatrix}}',$$

where

$$F=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\dot{+}\ \cdots\ \dot{+}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}.$$

Since the element in the upper left-hand corner of the matrix $UK_1\bar{K}_1'\bar{U}'$ is equal to zero, the first row of the matrix UK_1 consists of zeros only, i.e.,

$$UK_1=\begin{pmatrix}0,\ 0,\ \ldots,\ 0\\ *\end{pmatrix}.$$

From the fact that UK_1U' is a skew-symmetric matrix it follows that

$$UK_1U'=\begin{pmatrix}0 & 0\\ 0 & Q\end{pmatrix},$$

where Q is a skew-symmetric matrix of order $(n-1)$ and $Q\bar{Q}'=I=F\bar{F}'$.

But if Q is a unitary skew-symmetric matrix, then a unitary matrix V_0 can be found such that $V_0 Q V_0' = F$. Hence it is also possible to find a unitary matrix U_0 such that

$$\overline{U}_0' K_1 \overline{U}_0 = \begin{pmatrix} 0 & 0 \\ 0 & F \end{pmatrix}.$$

Setting $h = k\overline{U}_0$, we obtain

$$h \begin{pmatrix} 0 & 0 \\ 0 & F \end{pmatrix}' = k\overline{U}_0 U_0' \overline{K}_1' U_0 = 0.$$

This means that

$$h = [e^{i\theta}, 0, \ldots, 0].$$

Thus the inner integral in formula (4.7.10) is equal to

$$\int_0^{2\pi} f\left[\begin{pmatrix} 0 & hU_0' \\ -U_0 h' & K_1 \end{pmatrix}\right] d\theta = 2\pi f\left[\begin{pmatrix} 0 & 0 \\ 0 & K_1 \end{pmatrix}\right]$$

and we obtain (4.7.8) from (4.7.10).

(We used the formula

$$\frac{1}{2\pi}\int_0^{2\pi} \varphi(z_1 e^{i\theta}, \ldots, z_n e^{i\theta})\, d\theta = \varphi(0, 0, \ldots, 0),$$

which holds for $\varphi(z)$ analytic in a closed circular domain with its center at the origin and z lying inside the domain.)

4°. The characteristic manifold of the domain $\mathfrak{R}_{\mathrm{IV}}$ consists of vectors of the form $e^{i\theta}x$, where $0 \leqq \theta \leqq \pi$, and $x = (x_1, \cdots, x_n)$ is a real vector which satisfies the condition $xx' = 1$.

$$H(z, \theta, x) = \frac{1}{V(\mathfrak{C}_{\mathrm{IV}})\,[(x - e^{-i\theta}z)(x - e^{-i\theta}z)']^{n/2}}, \tag{4.7.11}$$

It is easy to calculate the magnitude of the volume $V(\mathfrak{C}_{\mathrm{IV}})$:

$$V(\mathfrak{C}_{\mathrm{IV}}) = \frac{2\pi^{\frac{n}{2}+1}}{\Gamma\left(\frac{n}{2}\right)}.$$

4.8. **The Poisson kernel for circular domains.** Suppose that $\mathfrak{R}$, just as in §4.5, is a star-shaped circular domain, and $\mathfrak{C}$ its characteristic manifold, transitive with respect to the group Γ_0 of motions of $\mathfrak{R}$ which leave the origin unchanged. Then, by Theorem 4.6.1, there exists a Cauchy kernel

for the domain $\mathfrak{R}$, and Cauchy's formula holds for any function $f(z)$ which is analytic in $\mathfrak{R}$ and on its boundary.

Setting, in particular,

$$f(z) = H(z, \overline{w})\, g(z),$$

where $g(z)$ is an arbitrary function which is analytic in $\mathfrak{R}$ and on its boundary, we have

$$H(z, \overline{w})\, g(z) = \int_{\mathfrak{C}} H(z, \overline{\xi})\, H(\xi, \overline{w})\, g(\xi)\, \dot{\xi}.$$

For $w = z$, we obtain Poisson's formula

$$g(z) = \int_{\mathfrak{C}} P(z, \xi)\, g(\xi)\, \dot{\xi}, \tag{4.8.1}$$

where the kernel

$$P(z, \xi) = \frac{H(z, \overline{\xi})\, H(\xi, \overline{z})}{H(z, \overline{z})} \tag{4.8.2}$$

has occured above under the name of the Poisson kernel of the domain $\mathfrak{R}$.

Up to now we have established that formula (4.8.1) is valid for analytic $g(z)$; yet it can be extended to other classes of functions too (see §5.8). For any continuous function $u(\xi)$ the integral

$$u(z) = \int_{\mathfrak{C}} P(z, \xi)\, u(\xi)\, \dot{\xi} \tag{4.8.3}$$

defines a certain function. It can be proved (see §5.8) that $u(z) \to u(\xi)$ for $z \to \xi$. Functions of the form (4.8.3) we shall call harmonic functions in $\mathfrak{R}$. It is reasonable to expect that if there exists a complete orthonormal system $\{\psi_\nu(\xi)\}$ on $\mathfrak{C}$, then the set of functions which are harmonic in $\mathfrak{R}$ is the closure of the linear span of the system $\{\psi_\nu(z)\}$ (see §5.10).

If $\mathfrak{R}$ satisfies the conditions of Theorem 4.6.3, then the Poisson kernel can be written in the following simple form:

$$P(z, \xi) = \frac{1}{V(\mathfrak{C})} \cdot |B(\xi, z, U)|. \tag{4.8.4}$$

In conclusion let us list the Poisson kernels for the classical domains.

(1) For $\mathfrak{R}_{\mathrm{I}}$

$$P(Z, U) = \frac{1}{V(\mathfrak{C}_{\mathrm{I}})} \cdot \frac{[\det(I - Z\overline{Z}')]^n}{|\det(I - Z\overline{U}')|^{2n}}, \tag{4.8.5}$$

where $U \in \mathfrak{C}_{\mathrm{I}}$. In particular, for $m = n$, one can also write

$$P(Z, U) = \frac{1}{V(\mathfrak{C}_{\mathrm{I}})} \cdot \frac{[\det(I - Z\bar{Z}')]^n}{|\det(Z - U)|^{2n}}.$$

(2) For $\mathfrak{R}_{\mathrm{II}}$

$$P(Z, U) = \frac{1}{V(\mathfrak{C}_{\mathrm{II}})} \cdot \frac{[\det(I - Z\bar{Z})]^{\frac{n+1}{2}}}{|\det(I - Z\bar{U})|^{n+1}}, \tag{4.8.6}$$

where $U \in \mathfrak{C}_{\mathrm{II}}$.

(3) For $\mathfrak{R}_{\mathrm{III}}$ with even n

$$P(Z, K) = \frac{1}{V(\mathfrak{C}_{\mathrm{III}})} \cdot \frac{[\det(I + Z\bar{Z})]^{\frac{n-1}{2}}}{|\det(I + Z\bar{K})|^{n-1}}, \tag{4.8.7}$$

and with odd n

$$P(Z, K) = \frac{1}{V(\mathfrak{C}_{\mathrm{III}})} \cdot \frac{[\det(I + Z\bar{Z})]^{\frac{n}{2}}}{|\det(I + Z\bar{K})|^{n}}. \tag{4.8.8}$$

In both cases $K \in \mathfrak{C}_{\mathrm{III}}$.

(4) For $\mathfrak{R}_{\mathrm{IV}}$

$$P(z, \xi) = \frac{1}{V(\mathfrak{C}_{\mathrm{IV}})} \cdot \frac{(1 + |zz'|^2 - 2\bar{z}z')^{\frac{n}{2}}}{|(z - \xi)(z - \xi)'|^{n}}, \tag{4.8.9}$$

where $\xi \in \mathfrak{C}_{\mathrm{IV}}$.

Fourth printing, 2000
First MIT Press paperback edition, 1993

Originally published in Japanese in 1954 by Iwanami Shoten, Publishers, Tokyo, under the title *Iwanami Sūgaku Ziten*.

This book was set in Monolaser Times Roman by Asco Trade Typesetting Ltd., Hong Kong, and printed and bound by Arcata Graphics, Kingsport in the United States of America.

Library of Congress Cataloging-in-Publication Data

Iwanami sūgaku jiten (ziten). English.
Encyclopedic dictionary of mathematics.
Translation of: Iwanami sūgaku jiten.
Includes bibliographies and indexes.
1. Mathematics—Dictionaries. I. Itô, Kiyosi, 1915– . II. Nihon Sūgakkai. III. Title.
QA5.I8313 1986 510'.3'21 86-21092
ISBN 0-262-59020-4 (PB)

can be represented as homogeneous spaces of classical groups, using the notation introduced by E. Cartan. We denote by $M_u = G/K$ a simply connected irreducible symmetric Riemannian space of type (2), where G is a group that acts almost effectively on M_u and K is the subgroup given by $K = K_\theta^0$ for an involutive automorphism θ of G. For such an M_u, the space of type (4) that is dual to M_u is denoted by $M_\theta = G_\theta/K$. Clearly $\dim M_u = \dim M_\theta$. (For the dimension and rank of M_u and for those M_u that are represented as homogeneous spaces of simply connected †exceptional compact simple Lie groups → Appendix A, Table 5.III.) In this section (and also in Appendix A, Table 5.III), $O(n)$, $U(n)$, $Sp(n)$, $SL(n, \mathbf{R})$, and $SL(n, \mathbf{C})$ are the †orthogonal group of degree n, the †unitary group of degree n, the †symplectic group of degree $2n$, and the real and complex †special linear groups of degree n, respectively. Let $SO(n) = SL(n, \mathbf{R}) \cap O(n)$ and $SU(n) = SL(n, \mathbf{C}) \cap U(n)$. We put

$$I_{p,q} = \begin{pmatrix} -I_p & 0 \\ 0 & I_q \end{pmatrix}, \quad J_n = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

$$K_{p,q} = \begin{pmatrix} -I_p & 0 & 0 & 0 \\ 0 & I_q & 0 & 0 \\ 0 & 0 & -I_p & 0 \\ 0 & 0 & 0 & I_q \end{pmatrix},$$

where I_p is the $p \times p$ unit matrix.

Type AI. $M_u = SU(n)/SO(n)$ $(n > 1)$, where $\theta(s) = \bar{s}$ (with $\bar{s}$ the complex conjugate matrix of s). $M_\theta = SL(n, \mathbf{R})/SO(n)$.

Type AII. $M_u = SU(2n)/Sp(n)$ $(n > 1)$, where $\theta(s) = J_n s J_n^{-1}$. $M_\theta = SU^*(2n)/Sp(n)$. Here $SU^*(2n)$ is the subgroup of $SL(2n, \mathbf{C})$ formed by the matrices that commute with the transformation $(z_1, \ldots, z_n, z_{n+1}, \ldots, z_{2n}) \to (\bar{z}_{n+1}, \ldots, \bar{z}_{2n}, -\bar{z}_1, \ldots, -\bar{z}_n)$ in $\mathbf{C}^n$; $SU^*(2n)$ is called the **quaternion unimodular group** and is isomorphic to the commutator group of the group of all regular transformations in an n-dimensional vector space over the quaternion field $\mathbf{H}$.

Type AIII. $M_u = SU(p+q)/S(U_p \times U_q)$ $(p \geqslant q \geqslant 1)$, where $S(U_p \times U_q) = SU(p+q) \cap (U(p) \times U(q))$, with $U(p) \times U(q)$ being canonically identified with a subgroup of $U(p+q)$, and $\theta(s) = I_{p,q} s I_{p,q}$. This space M_u is a †complex Grassmann manifold. $M_\theta = SU(p,q)/S(U_p \times U_q)$, where $SU(p,q)$ is the subgroup of $SL(p+q, \mathbf{C})$ consisting of matrices that leave invariant the Hermitian form $z_1\bar{z}_1 + \ldots + z_p\bar{z}_p - z_{p+1}\bar{z}_{p+1} - \ldots - z_{p+q}\bar{z}_{p+q}$.

Type AIV. This is the case $q = 1$ of type AIII. M_u is the $(n-1)$-dimensional complex projective space, and M_θ is called a **Hermitian hyperbolic space**.

Type BDI. $M_u = SO(p+q)/SO(p) \times SO(q)$ $(p \geqslant q \geqslant 1, p > 1, p+q \neq 4)$, where $\theta(s) = I_{p,q} s I_{p,q}$. M_u is the †real Grassmann manifold formed by the oriented p-dimensional subspaces in $\mathbf{R}^{p+q}$. $M_\theta = SO_0(p,q)/SO(p) \times SO(q)$, where $SO(p,q)$ is the subgroup of $SL(n, \mathbf{R})$ consisting of matrices that leave invariant the quadratic form $x_1^2 + \ldots + x_p^2 - x_{p+1}^2 - \ldots - x_{p+q}^2$, and $SO_0(p,q)$ is the connected component of the identity element.

Type BDII. This is the case $q = 1$ of type BDI. M_u is the $(n-1)$-dimensional sphere, and M_θ is called a **real hyperbolic space**.

Type DIII. $M_u = SO(2n)/U(n)$ $(n > 2)$, where $U(n)$ is regarded as a subgroup of $SO(2n)$ by identifying $s \in U(n)$ with

$$\begin{pmatrix} \operatorname{Re} s & \operatorname{Im} s \\ -\operatorname{Im} s & \operatorname{Re} s \end{pmatrix} \in SO(2n),$$

and $\theta(s) = J_n s J_n^{-1}$. $M_\theta = SO^*(2n)/U(n)$. Here $SO^*(2n)$ denotes the group of all complex orthogonal matrices of determinant 1 leaving invariant the skew-Hermitian form $z_1\bar{z}_{n+1} - z_{n+1}\bar{z}_1 + z_2\bar{z}_{n+2} - z_{n+2}\bar{z}_2 + \ldots + z_n\bar{z}_{2n} - z_{2n}\bar{z}_n$; this group is isomorphic to the group of all linear transformations leaving invariant a nondegenerate skew-Hermitian form in an n-dimensional vector space over the quaternion field $\mathbf{H}$.

Type CI. $M_u = Sp(n)/U(n)$ $(n \geqslant 1)$, where $U(n)$ is considered as a subgroup of $Sp(n)$ by the identification $U(n) \subset SO(2n)$ explained in type DIII and $\theta(s) = \bar{s} (= J_n s J_n^{-1})$. $M_\theta = Sp(n, \mathbf{R})/U(n)$, where $Sp(n, \mathbf{R})$ is the real symplectic group of degree $2n$.

Type CII. $M_u = Sp(p+q)/Sp(p) \times Sp(q)$ $(p \geqslant q \geqslant 1)$, where $Sp(p) \times Sp(q)$ is identified with a subgroup of $Sp(p+q)$ by the mapping

$$\left(\begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} \right) \to \begin{pmatrix} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{pmatrix}$$

and $\theta(s) = K_{p,q} s K_{p,q}$. $M_\theta = Sp(p,q)/Sp(p) \times Sp(q)$. Here $Sp(p,q)$ is the group of complex symplectic matrices of degree $2(p+q)$ leaving invariant the Hermitian form $(z_1, \ldots, z_{p+q}) K_{p,q}\, {}^t(\bar{z}_1, \ldots, \bar{z}_{p+q})$; this group is interpreted as the group of all linear transformations leaving invariant a nondegenerate Hermitian form of index p in a $(p+q)$-dimensional vector space over the quaternion field $\mathbf{H}$. For $q = 1$, M_u is the quaternion projective space, and M_θ is called the **quaternion hyperbolic space**.

Among the spaces introduced here, there are some with lower p, q, n that coincide (as Riemannian spaces) (→ Appendix A, Table 5.III).

H. Space Forms

A Riemannian manifold of †constant curvature is called a **space form**; it is said to be **spherical**,

Euclidean, or **hyperbolic** according as the constant curvature K is positive, zero, or negative. A space form is a locally symmetric Riemannian space; a simply connected complete space form is a sphere if $K>0$, a real Euclidean space if $K=0$, and a real hyperbolic space if $K<0$. More generally, a complete spherical space form of even dimension is a sphere or a projective space, and one of odd dimension is an orientable manifold. A complete 2-dimensional Euclidean space form is one of the following spaces: Euclidean plane, cylinder, torus, †Möbius strip, †Klein bottle. Except for these five spaces and the 2-dimensional sphere, any †closed surface is a 2-dimensional hyperbolic space form (for details about space forms → [6]).

I. Examples of Irreducible Symmetric Bounded Domains

Among the irreducible symmetric Riemannian spaces described in Section H, those defined by irreducible symmetric Hermitian spaces are of types AIII, DIII, BDI ($q=2$), and CI. We list the irreducible symmetric bounded domains that are isomorphic to the irreducible Hermitian spaces defining these spaces. Positive definiteness of a matrix will be written $\gg 0$.

Type $\mathrm{I}_{m,m'}$ ($m' \geqslant m \geqslant 1$). The set of all $m \times m'$ complex matrices Z satisfying the condition $I_{m'} - {}^t\bar{Z}Z \gg 0$ is a symmetric bounded domain in $\mathbf{C}^{mm'}$, which is isomorphic (as a complex manifold) to the irreducible symmetric Hermitian space defined by M_θ of type AIII ($p=m$, $q=m'$).

Type II_m ($m \geqslant 2$). The set of all $m \times m$ complex †skew-symmetric matrices Z satisfying the condition $I_m - {}^t\bar{Z}Z \gg 0$ is a symmetric bounded domain in $\mathbf{C}^{m(m-1)/2}$ corresponding to the type DIII ($n=m$).

Type III_m ($m \geqslant 1$). The set of all $m \times m$ complex symmetric matrices satisfying the condition $I_m - {}^t\bar{Z}Z \gg 0$ is a symmetric bounded domain in $\mathbf{C}^{m(m+1)/2}$ corresponding to the type CI ($n=m$). This bounded domain is holomorphically isomorphic to the †Siegel upper half-space of degree m.

Type IV_m ($m \geqslant 1, m \neq 2$). This bounded domain in $\mathbf{C}^m$ is formed by the elements $(z_1, \ldots, z_m)$ satisfying the condition $|z_1|^2 + \ldots + |z_m|^2 < (1 + |z_1^2 + \ldots + z_m^2|)/2 < 1$, and corresponds to the type BDI ($p=m, q=2$).

Among these four types of bounded domains, the following complex analytic isomorphisms hold: $\mathrm{I}_{1,1} \cong \mathrm{II}_2 \cong \mathrm{III}_1 \cong \mathrm{IV}_1$, $\mathrm{II}_3 \cong \mathrm{I}_{1,3}$, $\mathrm{IV}_3 \cong \mathrm{III}_2$, $\mathrm{IV}_4 \cong \mathrm{I}_{2,2}$, $\mathrm{IV}_6 \cong \mathrm{II}_4$. (For details about these symmetric bounded domains → [2].) There are two more kinds of irreducible symmetric bounded domains, which are represented as homogeneous spaces of exceptional Lie groups.

J. Weakly Symmetric Riemannian Spaces

A generalization of symmetric Riemannian space is the notion of weakly symmetric Riemannian space introduced by Selberg. Let M be a Riemannian space. M is called a **weakly symmetric Riemannian space** if a Lie subgroup G of the group of isometries $I(M)$ acts transitively on M and there exists an element $\mu \in I(M)$ satisfying the relations (i) $\mu G \mu^{-1} = G$; (ii) $\mu^2 \in G$; and (iii) for any two points x, y of M, there exists an element m of G such that $\mu x = my$, $\mu y = mx$. A symmetric Riemannian space M becomes a weakly symmetric Riemannian space if we put $G = I(M)$ and $\mu =$ the identity transformation; as the element m in condition (iii) we can take the symmetry σ_p at the midpoint p on the geodesic arc joining x and y. There are, however, weakly symmetric Riemannian spaces that do not have the structure of a symmetric Riemannian space. An example of such a space is given by $M = G = SL(2, \mathbf{R})$ with a suitable Riemannian metric, where μ is the inner automorphism defined by

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

(Selberg [4]). On a weakly symmetric Riemannian space, the ring of all G-invariant differential-integral operators is commutative; this fact is useful in the theory of spherical functions (→ 437 Unitary Representations).

References

[1] S. Helgason, Differential geometry, Lie groups, and symmetric spaces, Academic Press, 1978.
[2] C. L. Siegel, Analytic functions of several complex variables, Princeton Univ. Press, 1950.
[3] J.-L. Koszul, Exposés sur les espaces homogènes symétriques, São Paulo, 1959.
[4] A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with application to Dirichlet series, J. Indian Math. Soc., 20 (1956), 47–87.
[5] E. Cartan, Sur certaines formes riemanniennes remarquables des géométries à groupe fondamental simple, Ann. Sci. Ecole Norm. Sup., 44 (1927), 345–467 (Oeuvres complètes, Gauthier-Villars, 1952, pt. I, vol. 2, 867–989).
[6] E. Cartan, Sur les domaines bornés homogènes de l'espace de n-variables complexes, Abh. Math. Sem. Univ. Hamburg, 11 (1936), 116–162 (Oeuvres complètes, Gauthier-Villars, 1952, pt. I, vol. 2, 1259–1305).

Schwinger - Hua - Wyler

Frank Dodd Tony Smith Jr - 2013 - vixra 1311.0088

In their book "Climbing the Mountain: The Scientific Biography of Julian Schwinger"
Jagdish Mehra and Kimball Milton said:

"... Schwinger ... always felt that the mathematics should emerge from the physics, not the other way around ...

[Julian Schwinger said in conversations and interviews with Jagdish Mehra, in Bel Air, California, March 1988]:

"... in 1966 ... Schwinger ... realized that he could base the whole machinery of particle physics on the abstraction of particle-creation and annihilation acts.
One can define a free action, say for a photon, in terms of propagation of virtual photons between photon sources, conserved in order to remove the scalar degree of freedom.
But a virtual photon can in turn act as a pair of electron-positron sources, through a `primitive interaction' between electrons and photons, essentially embodied in the conserved Dirac current.
So this multiparticle exchange gives rise to quantum corrections to the photon propagator, to vacuum polarization, and so on.
All this without any reference to renormalization or `high-energy speculations'. ...
The problem with conventional field theory is that it makes an implicit hypothesis that the physics is known down to zero distance ...
Source theory was ... that the physical quantities that you are interested in
were not the fields but the correlations between fields
and ... that the correlations between fields are really Green's functions
... which ... take into account not only how the particles behave
but how they are created.
The sources are the way of cataloging the various Green's functions.
The final point at which the theory asks to be compared with experiment ... involves just pure numbers, Green's functions and sources, not operator fields. ...
The whole point was to develop the space-time structure of a Green's function in general so it will be applicable both to stable particles and unstable particles.
...
Green's functions [were] universally recognized as carrying the information of physical interest ... one had differential equations for these Green's functions and then came the necessity of picking out of the vast infinity of solutions the physical ones of interest ...
This was enforced by appropriate boundary conditions, that the wave propagate outwards, that is, the idea of causality ... if you rotated the time axis into a complex space, then the boundary conditions ... would select just the physically acceptable states of the Green's function ... all representations of physical interest can be obtained from the ... Euclidean group ... attached [to]... the Lorentz group ... (the "unitary trick" of Weyl) ... a correspondence between the quantum theory of fields with its underlying Lorentz space, and a mathematical image in a Euclidean space ...".

The Schwinger Sources are finite regions in a Complex Domain spacetime corresponding to Green's functions of particle creation / annihilation.

What Complex Domains have Symmetries of Particle Physics ?

E8 8-dim Octonionic Spacetime (effective at high Planck-scale energies)
is by Triality isomorphic with the natural representation space of fundamental
First-Generation Fermion Particles (and AntiParticles)
so
Fermion Particles (and AntiParticles) are represented by Schwinger Sources
with Bounded Complex Domain structure of a Cartan domain.

David B. Lowdenslager in Annals of Mathematics 67 (1958) 467-484 said: "... For an irreducible Cartan domain ... there is only one linearly independent Riemannian metric ... the Bergman metric ... corresponding to ... Δ ... the Laplace-Beltrami operator ... solutions of ... Δ f = 0 ... are determined by their values ... on the ... Bergman-Shilov ... boundary B ... Let D be a classical Cartan domain, Δ an invariant Laplacian, and K a Poisson kernel for D. Then K as a function of D satisfies ΔK = 0 , for all b in B ...".

Steven G. Krantz in his book "Geometric Analysis of the Bergman Kernel and Metric" said:

"... the Bergman kernel ... K ... for Ω is related to the Green's function ... θ ... for the boundary value problem

$$\begin{aligned} \frac{\partial}{\partial \bar{z}_j} \triangle \beta &= 0 \qquad \text{on } \Omega, \quad j = 1, \ldots, k \\ \sum_{j=1}^{k} \frac{\partial \beta}{\partial \bar{z}_j} \cdot \overline{\alpha}_j &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

... in this way

$$K_\Omega(z,t) = \triangle_z \theta(z,t) \, .$$

...".

Armand Wyler, in his 1972 IAS Princeton preprint "The Complex Light Cone Symmetric Space of the Conformal Group", said: "... the bounded realization Dn of SO(n,2) / SO(n)xSO(2) ... allows to define ... the Bergman metric, the invariant differential operators and their elementary solutions (Green functions) ...[and]... the Shilov boundary Qn ...[as]... the quotient space C(Mn) / P(Mn) of the conformal group by the Poincare group ... and give ... eigenvalues of Casimir operators in the Lie algebra of C(Mn) ...".
In Wyler's approach, the elementary solutions of the invariant differential operators in the Bounded Complex Schwinger Source Domains are Schwinger Green's functions.

Using Schwinger-type Euclidean Spin(10) version of the Spin(8,2) Conformal Group, the Fermion Schwinger Sources correspond to the Symmetric space
the Lie Sphere Spin(10) / Spin(8)xU(1)
which has local symmetry of the Spin(8) gauge group with respect to which the first generation spinor fermions are seen as +half-spinor and -half-spinor spaces,
so the Fermion Schwinger Source Bounded Complex Domain D8 is of type IV8
which has Shilov Boundary Q8 = RP1 x S7 .

The Complex Domain of type IV8 is described by L. K. Hua in his book
"Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains" as

(4) The domain $\mathfrak{R}_{IV}$ of n-dimensional $(n>2)$ vectors

$$z=(z_1, z_2, \cdots, z_n)$$ BDI (q=2)

(z_k are complex numbers) satisfying the conditions[2]

$$|zz'|^2+1-2\bar{z}z'>0, \qquad |zz'|<1.$$

with Characteristic Manifold = Shilov Boundary = RP1 x S7

4°. The characteristic manifold of the domain $\mathfrak{R}_{IV}$ consists of vectors of the form $e^{i\theta}x$, where $0\leq\theta\leq\pi$, and $x=(x_1, \cdots, x_n)$ is a real vector which satisfies the condition $xx'=1$.

$$H(z, \theta, x)=\frac{1}{V(\mathfrak{C}_{IV})\,[(x-e^{-i\theta}z)(x-e^{-i\theta}z)']^{n/2}}, \tag{4.7.11}$$

It is easy to calculate the magnitude of the volume $V(\mathfrak{C}_{IV})$:

$$V(\mathfrak{C}_{IV})=\frac{2\pi^{\frac{n}{2}+1}}{\Gamma\left(\frac{n}{2}\right)}.$$

The Poisson kernel of a type IV Complex Domain is

(4) For $\mathfrak{R}_{IV}$

$$P(z, \xi)=\frac{1}{V(\mathfrak{C}_{IV})}\cdot\frac{(1+|zz'|^2-2\bar{z}z')^{\frac{n}{2}}}{|(z-\xi)(z-\xi)'|^n}, \tag{4.8.9}$$

where $\xi\in\mathfrak{C}_{IV}$.

and the Bergman kernel of a type IV Complex Domain is

THEOREM 4.4.1. *The Bergman kernel of the domain* $\mathfrak{R}_{IV}$ *is*

$$\frac{1}{V(\mathfrak{R}_{IV})}(1+|zz'|^2-2\bar{z}z')^{-n},$$

where, by (2.5.7),

$$V(\mathfrak{R}_{IV})=\frac{\pi^n}{2^{n-1}\cdot n!}.$$

How big are the Schwinger Sources ?

Schwinger Sources as described above are continuous manifold structures
of Bounded Complex Domains and their Shilov Boundaries
but
E8 Physics at the Planck Scale has spacetime condensing out of Clifford structures
forming a Leech lattice underlying 26-dim String Theory of World-Lines
represents 8 + 8 + 8 = 24-dim of fermion particles and antiparticles and of spacetime.

The automorphism group of a single 26-dim String Theory cell modulo the Leech lattice
is the Monster Group of order about 8 x 10^53.

When a fermion particle/antiparticle appears in E8 spacetime it does not remain a single
Planck-scale entity becauseTachyons create a cloud of particles/antiparticles.
The cloud is one Planck-scale Fundamental Fermion Valence Particle plus an effectively
neutral cloud of particle/antiparticle pairs forming a Kerr-Newman black hole.

That cloud constitutes the Schwinger Source.
Its structure comes from the 24-dim Leech lattice part of the Monster Group which is
2^(1+24) times the double cover of Co1, for a total order of about 10^26.

(Since a Leech lattice is based on copies of an E8 lattice and since there are 7 distinct E8 integral domain
lattices there are 7 (or 8 if you include a non-integral domain E8 lattice)mdistinct Leech lattices.
The physical Leech lattice is a superposition of them, effectively adding a factor of 8 to the order.)

The volume of the Kerr-Newman Cloud is on the order of 10^27 x Planck scale,
so the Kerr-Newman Cloud should contain about 10^27 particle/antiparticle pairs
and its size should be about 10^(27/3) x 1.6 x 10^(-33) cm =
= roughly 10^(-24) cm.

How do the Schwinger Sources fit into the E8 Lagrangian Structure ?

The fundamental high-energy E8 Lagrangian for Octonionic 8-dim SpaceTime is

$$\int_{ST} GRb + StMb + Spf \quad .$$

an integral over SpaceTime ST of a Gravity boson term GRb,
a Standard Model boson term StMb, and a Spinor fermion term Spf.

Consider the Spinor fermion term Spf based on Schwinger Source Fermions.

In the conventional picture, the spinor fermion term is of the form m S S* where m is the fermion mass and S and S* represent the given fermion.
Although the mass m is derived from the Higgs mechanism, the Higgs coupling constants are, in the conventional picture, ad hoc parameters, so that effectively the mass term is, in the conventional picuture, an ad hoc inclusion.

E8 Physics does not put in the mass m in an ad hoc way,
but
constructs the Lagrangian integral such that
the mass m emerges naturally from the geometry of the spinor fermions by setting the spinor fermion mass term as the volume of the Schwinger Source Fermions.

Effectively the integral over the Schwinger Source spacetime region
of its Kerr-Newman cloud of virtual particle/antiparticle pairs plus the valence fermion
gives the volume of the Schwinger Source fermion and defines its mass,
which, since it is dressed with the particle/antiparticle pair cloud,
gives quark mass as constituent mass.

Note that in the process of breaking Octonionic 8-dim SpaceTime
down to Quaternionic (4+4)-dim M4 x CP2 Kaluza-Klein
all Fermions are treated similarly so that ratios of their masses remain the same.

What about Gauge Bosons ?

The fundamental high-energy E8 Lagrangian for Octonionic 8-dim SpaceTime is

$$\int_{ST} GRb + StMb + Spf \quad .$$

an integral over SpaceTime ST of a Standard Model boson term StMb,
a Gravity boson term GRb, and a Spinor fermion term Spf.

What are the Schwinger Sources for the gauge boson terms StMb and GRb ?

The GRb bosons live in one of the two D4 Lie SubAlgebras of the E8 Lie Algebra.

The StMb bosons live in the other of the two D4 Lie SubAlgebras of the E8 Lie Algebra.

The process of breaking Octonionic 8-dim SpaceTime
down to Quaternionic (4+4)-dim M4 x CP2 Kaluza-Klein
creates differences in the way gauge bosons "see" M4 Physical SpaceTime.

Joseph Wolf (Journal of Mathematics and Mechanics 14 (1965) 1033) showed that
there are only 4 equivalence classes of 4-dimensional Riemannian Symmetric Spaces
with Quaternionic structures, with the following repesentatives:

S4 = 4-sphere = Spin(5) / Spin(4) where Spin(5) is the Schwinger-Euclidean version of the Anti-DeSitter Group that gives MacDowell-Mansouri Gravity

CP2 = complex projective 2-space = SU(3) / U(2) with the SU(3) of the Color Force

S2 x S2 = SU(2)/U(1) x SU(2)/U(1) with two copies of the SU(2) of the Weak Force

S1 x S1 x S1 x S1 = U(1) x U(1) x U(1) x U(1) = 4 copies of the U(1) of the EM Photon
(1 copy for each of the 4 covariant components of the Photon)

The GRb bosons (Schwinger-Euclidean versions) live in
a Spin(5) subalgebra of the Spin(6) Conformal subalgebra of D4 = Spin(8).
They "see" M4 Physical spacetime as the 4-sphere S4
so that their part of the Physical Lagrangian is

$$\int_{S4} \mathrm{GRb} \quad .$$

an integral over SpaceTime S4.
The Schwinger Sources for GRb bosons are the Complex Bounded Domains and
Shilov Boundaries for Spin(5) MacDowell-Mansouri Gravity bosons.
However, due to Stabilization of Condensate SpaceTime
by virtual Planck Mass Gravitational Black Holes,
for Gravity, the effective force strength that we see in our experiments
is not just composed of the S4 volume and the Spin(5) Schwinger Source volume,
but is suppressed by the square of the Planck Mass.
The unsuppressed Gravity force strength is the Geometric Part of the force strength.

The Standard Model SU(3) Color Force bosons live in
a SU(3) subalgebra of the SU(4) subalgebra of D4 = Spin(8).
They "see" M4 Physical spacetime as the complex projective plane CP2
so that their part of the Physical Lagrangian is

$$\int_{CP2} (\mathrm{SU(3)\ part\ of\ StM})b \quad .$$

an integral over SpaceTime CP2.
The Schwinger Sources for SU(3) bosons are the Complex Bounded Domains and
Shilov Boundaries for SU(3) Color Force bosons.
The Color Force Strength is given by
the SpaceTime CP2 volume and the SU(3) Schwinger Source volume.
Note that since the Schwinger Source volume is dressed with the particle/antiparticle
pair cloud, the calculated force strength is
for the characteristic energy level of the Color Force (about 245 MeV).

The Standard Model SU(2) Weak Force bosons live in
a SU(2) subalgebra of the U(2) local group of CP2 = SU(3) / U(2)
They "see" M4 Physical spacetime as two 2-spheres S2 x S2
so that their part of the Physical Lagrangian is

$$\int_{S2xS2} (SU(2) \text{ part of } StM)b \quad .$$

an integral over SpaceTime S2xS2.
The Schwinger Sources for SU(2) bosons are the Complex Bounded Domains and
Shilov Boundaries for SU(2) Weak Force bosons.
However, due to the action of the Higgs mechanism,
for the Weak Force, the effective force strength that we see in our experiments
is not just composed of the S2xS2 volume and the SU(2) Schwinger Source volume,
but is suppressed by the square of the Weak Boson masses.
The unsuppressed Weak Force strength is the Geometric Part of the force strength.

The Standard Model U(1) Electromagnetic Force bosons (photons) live in
a U(1) subalgebra of the U(2) local group of CP2 = SU(3) / U(2)
They "see" M4 Physical spacetime as four 1-sphere circles S1xS1xS1xS1 = T4
(T4 = 4-torus) so that their part of the Physical Lagrangian is

$$\int_{T4} (U(1) \text{ part of } StM)b \quad .$$

an integral over SpaceTime T4.
The Schwinger Sources for U(1) photons
are the Complex Bounded Domains and Shilov Boundaries for U(1) photons.
The Electromagnetic Force Strength is given by
the SpaceTime T4 volume and the U(1) Schwinger Source volume.

What are the results of Wyler-type calculations for Schwinger Sources ?

The Schwinger Source calculations using the Wyler approach give the following results, details of which can be found at http://vixra.org/abs/1310.0182 and my web sites.
Since calculations are for ratios of particle masses and force strengths,
the Higgs mass and the Geometric Part of the Gravity force strength are set so that the ratios agree with conventional observation data.

Particle/Force	Tree-Level	Higher-Order
e-neutrino	0	0 for nu_1
mu-neutrino	0	9 x 10^(-3) eV for nu_2
tau-neutrino	0	5.4 x 10^(-2) eV for nu_3
electron	0.5110 MeV	
down quark	312.8 MeV	charged pion = 139 MeV
up quark	312.8 MeV	proton = 938.25 MeV
		neutron - proton = 1.1 MeV
muon	104.8 MeV	106.2 MeV
strange quark	625 MeV	
charm quark	2090 MeV	
tauon	1.88 GeV	
beauty quark	5.63 GeV	
truth quark (low state)	130 GeV	(middle state) 174 GeV
		(high state) 218 GeV
W+	80.326 GeV	
W-	80.326 GeV	
W0	98.379 GeV	Z0 = 91.862 GeV
Mplanck=1.217x10^19 GeV		
Higgs VEV (assumed)	252.5 GeV	
Higgs (low state)	126 GeV	(middle state) 182 GeV
		(high state) 239 GeV
Gravity Gg (assumed)	1	
(Gg)(Mproton^2 / Mplanck^2)		5 x 10^(-39)
EM fine structure	1/137.03608	
Weak Gw	0.2535	
Gw(Mproton^2 / (Mw+^2 + Mw-^2 + Mz0^2))		1.05 x 10^(-5)
Color Force at 0.245 GeV	0.6286	0.106 at 91 GeV