Chapter 0

This preliminary chapter contains a short exposition of the set theory we shall need. It begins with a brief discussion of logic, so that set theory can be presented with precision, and continues with a review of the way in which mathematical objects are defined as sets. The chapter ends with four sections which treat specific set-theoretic topics.

Some of this material will be familiar to the reader and some of it probably will be new. We suggest that he read it through "lightly" at the beginning, and then refer back to it for details as needed.

§1. Logic: quantifiers

We begin with logic. A <u>statement</u> is a sentence which is true or false as it stands. Thus '4 + 3 = 5' and '1 < 2' are, respectively, false and true mathematical statements. Many sentences occurring in mathematics contain variables and are therefore not true or false as they stand, but become statements when the variables are given values. Simple examples are: '$x < 4$', '$x < y$', 'x is an integer', '$3x^2 + y^2 = 10$'. Such sentences will be called <u>statement frames.</u> If $P(x)$ is a frame containing the one variable 'x', such as '$x < 4$', then $P(5)$ is the sentence obtained by replacing 'x' in $P(x)$ by the numeral '5', and $P(1)$ and $P(5)$ are accordingly statements. Another way to obtain a statement is to assert that $P(x)$ is always true. That is, if we replace $P(x)$ by (For every x) $P(x)$, then this sentence is either true or false as it stands. It is true if $P(x)$ is true for every x, and is false if $P(x)$ is false for some x. Thus, 'For every x, $x^2 > 0$' is false, and 'For every x, $x^2 - 1 = (x - 1)(x + 1)$' is true. Symbolically we use the prefix '$(\forall x)$', which can be variously translated as 'for all x', 'for every x', or 'for each x'. This prefixing phrase is called a <u>universal quantifier,</u> and the process of applying it to convert a statement frame into a statement is <u>universal quantification.</u>

Frequently sentences containing variables are presented as being always true without explicitly writing in the universal quantifiers. Thus the associative law for the addition of numbers is often written

$$x + (y + z) = (x + y) + z ,$$

it being understood that the equation is to be true for all x, y and z. The actual statement being made is thus

$$(\forall x)(\forall y)(\forall z)(x + (y + z) = (x + y) + z).$$

The reader may have realized that the frame $P(x)$ can also be made into a statement by asserting that it is sometimes true. We write '$(\exists x)P(x)$' and read it 'There exists an x such that $P(x)$'. This process is called existential quantification. Notice that '$(\exists x)$' abbreviates the long phrase 'There exists an x such that'.

The statement '$(\forall x)(x < 4)$' still contains the variable 'x' of course, but 'x' is no longer free to be given values, and is now called a bound variable.* The notation '$P(x)$' is used only when 'x' is free in the sentence being discusse

Now suppose that we have a sentence $P(x, y)$ containing two free variables. Clearly two quantifiers are needed to obtain a statement from it, and we now come to a very important observation. If quantifiers of both types are used then the order in which they are written affects the meaning of the statement; $(\exists y)(\forall x)P(x, y)$ and $(\forall x)(\exists y)P(x, y)$ say different things. The first says that one y can be found that works for all x: "There exists a y such that for all x ---". The second says that for each x a y can be found that works: "For each x there exists y such that ---". But in the second case, it may very well happen that when x is changed the y that can be found will also have to be changed. The existence of a single y that works for all x is thus the stronger statement. For example, it is true that $(\forall x)(\exists y)(x < y)$ and false that $(\exists y)(\forall x)(x < y)$. The reader must be absolutely clear on this point; his whole mathematical future is at stake. The second statement says that there exists a y, call it y_0, such that $(\forall x)(x < y_0)$, that is, such that every number is less than y_0. This is false; $y_0 + 1$, in particular, is not less than y_0. The first

*Roughly speaking, quantified varables are bound and unquantified varables are free.

statement says that for each x we can find a corresponding y. And we can: take $y = x + 1$.

On the other hand, among a group of quantifiers of the same type the order does not affect the meaning. Thus '$(\forall x)(\forall y)$' and '$(\forall y)(\forall x)$' have the same meaning. We often abbreviate such clumps of similar quantifiers by using the quantification symbol only once, as in '$(\forall x, y)$', which can be read 'For every x and y'. Thus the strictly correct '$(\forall x)(\forall y)(\forall z)(x + (y + z) = (x + y) + z)$ receives the slightly more idiomatic rendition '$(\forall x, y, z)(x+(y+z)=(x+y)+z)$'. The situation is clearly the same for a group of existential quantifiers.

The beginning student generally feels that the prefixing phrases "For every x there exists a y such that" and "There exists a y such that for every x" are artificial sounding and unidiomatic. This is indeed the case, but this awkwardness is the price that has to be paid to fix the order of the quantifiers, so that the meaning of the quantified statement is clear and unambiguous. Quantifiers do occur in ordinary idiomatic discourse, but often their idiomatic occurrences house ambiguity. The following two sentences are good examples of such ambiguous idiomatic usage : "Every x is less than some y", and "Some y is greater than every x". If a poll were taken it would be found that most men on the street feel that these two sentences say the same thing, but half will feel that the common assertion is false and half will think it true!

§2. The logical connectives

When the word 'and' is inserted between two sentences the resulting sentence is true if both constituent sentences are true and otherwise is false. That is, the "truth value", T or F, of the compound sentence depends only on the truth values of the constituent sentences. We can thus describe the way 'and' acts in compounding sentences in the simple "truth table"

P	Q	P and Q
T	T	T
T	F	F
F	T	F
F	F	F

where 'P' and 'Q' stand for arbitrary statement frames. Words like 'and' are called logical connectives.

Another such connective is the word 'or'. Unfortunately, this word is used ambiguously in ordinary discourse. Sometimes it is intended in the exclusive sense, where 'P or Q' means that one of P and Q is true but not both, and sometimes in the inclusive sense that at least one is true and possibly both are true. Mathematics cannot tolerate ambiguity and in mathematics 'or' is always used in the latter way. We thus have the truth table

P	Q	P or Q
T	T	T
T	F	T
F	T	T
F	F	F

The above two connectives are binary, in the sense that they combine two sentences to form one new sentence. The word not applies to one sentence and really shouldn't be considered to be a connective at all; nevertheless it is called a unary connective. Its truth table is obviously

P	not P
T	F
F	T

.

In idiomatic usage the word 'not' is generally buried in the interior of a sentence. We say that 'x is not equal to y' rather than 'not x is equal to y'. However, for the purposes of logical manipulation the negation sign (the word 'not' or a symbol like $\sim$) precedes the sentence being negated. We shall, of course, continue to write '$x \neq y$', but keep in mind that this is idiomatic for 'not $x = y$'.

We come now to the troublesome 'if - then' connective, which we write either as 'if P then Q' or '$P \Rightarrow Q$'. This is almost always applied in the universally quantified context $(\forall x)(P(x) \Rightarrow Q(x))$, and its meaning is best unravelled by studying this usage. We consider 'If $x < 3$ then $x < 5$' to be a true sentence. More exactly, it is true for all x, so that the universal quantification $(\forall x)(x < 3 \Rightarrow x < 5)$ is a true statement. This conclusion forces us to agree that in particular '$2 < 3 \Rightarrow 2 < 5$', '$4 < 3 \Rightarrow 4 < 5$' and '$6 < 3 \Rightarrow 6 < 5$' are all true statements. The truth table for '$\Rightarrow$' thus contains the values entered below.

P	Q	$P \Rightarrow Q$
T	T	T
T	F	--
F	T	T
F	F	T

On the other hand we feel that it is not always true that $x < 7 \Rightarrow x < 5$, and therefore, in particular, that '$6 < 7 \Rightarrow 6 < 5$' is false. Thus the remaining row in the table above gives the value 'F' for $P \Rightarrow Q$.

Combinations of frame variables and logical connectives such as we have been considering are called truth functional forms. The elementary forms such as 'P $\Rightarrow$ Q' and '~P' can be further combined by connectives to construct composite forms such as '~(P$\Rightarrow$Q)' and '(P$\Rightarrow$Q) and (Q$\Rightarrow$P)' . The class of all truth functional forms can be defined recursively as follows. First, any form variable, such as 'P' or 'Q' is a truth functional form. Second, if 'φ' and 'ψ' are replaced by truth functional forms, then 'φ and ψ' , 'φ or ψ' , '$\varphi \Rightarrow \psi$' and '~φ' become truth functional forms. A frame has a given (truth functional)form if it can be obtained from that form by substitution. Thus 'x<y or ~(x<y)' has the form 'P or ~P'. Composite truth functional forms have truth tables that can be worked out by combining the above tables. For example, '~(P$\Rightarrow$Q)' has the table below, the truth function for the whole form being in the column under the connective which is applied last ('~' in this example)

P	Q	$\sim$	(P $\Rightarrow$ Q)
T	T	F	T
T	F	T	F
F	T	F	T
F	F	F	T

and thus is true only when P is true and Q is false.

A truth functional form such as 'P or ($\sim$P)' which is always true (i.e., has only 'T' in the final column of its truth table) is called a tautology, or a tautologous form. The reader can check that '(P & (P $\Rightarrow$ Q)) $\Rightarrow$ Q' and '((P $\Rightarrow$ Q) & (Q $\Rightarrow$ R)) $\Rightarrow$ (P $\Rightarrow$ R)' are also tautologous. Indeed, any valid principles of reasoning that does not involve quautifiers must be expressed by a tautologous form.

The 'if and only if' connective, designated '$\Leftrightarrow$' is defined by: 'P $\Leftrightarrow$ Q' is an abbreviation for '(P $\Rightarrow$ Q) & (Q $\Rightarrow$ P)'. Its truth table works out to be

P	Q	P $\Leftrightarrow$ Q
T	T	T
T	F	F
F	T	F
F	F	T

That is P $\Leftrightarrow$ Q is true if and only if P and Q have the same truth values.

Two truth functional forms A and B are said to be equivalent if and only if (the final columns of) their truth tables are the same, and, in view of the table for '$\Leftrightarrow$', we see that A and B are equivalent if and only if A $\Leftrightarrow$ B is tautologous. Replacing a statement exemplifying a form A by the version obtained by recasting it under an equivalent form B is a device much used in logical reasoning. Thus to prove a statement P true it suffices to prove the statement $\sim P$ is false, since 'P' and '$\sim(\sim P)$' are equivalent forms. Other important equivalences are:

$$\sim(P \text{ or } Q) \Longleftrightarrow (\sim P) \,\&\, (\sim Q)$$

$$(P \Rightarrow Q) \Longleftrightarrow Q \text{ or } (\sim P)$$

$$\sim(P \Rightarrow Q) \Longleftrightarrow P \,\&\, (\sim Q)$$

§3. Negations of quantifiers

The combinations '$\sim(\forall x)$' and '$(\exists x)\sim$' have the same meanings: something is not always true if and only if it is sometimes false. Similarly, '$\sim(\exists y)$' and '$(\forall y)\sim$' have the same meanings. These equivalences can be applied to move a negation sign past each quantifier in turn in a sequence of quantifiers, giving the important practical rule that

> in taking the negation of a statement beginning with a string of quantifiers we simply change each quantifier to the opposite kind and move the negation sign to the end of the string.

Thus $\sim(\forall x)(\exists y)(\forall z)P(x, y, z) \Longleftrightarrow (\exists x)(\forall y)(\exists z)\sim P(x, y, z)$.

§4. Other principles of quantification

Besides the equivalence '$\sim(\forall x)P(x) \Longleftrightarrow (\exists x)\sim P(x)$' and our conclusions about the commutativity of quantifiers, there are a few other basic principles of quantificational reasoning that we use constantly. We list them below; the reader should convince himself that each is valid.

(1) If P is a tautologous frame and 'x_1', ..., 'x_n' are its free variables, then $(\forall x_1, \ldots, x_n)P$ is true.

(2) $(\forall x)P$ & $(\forall x)(P \Longrightarrow Q) \Longrightarrow (\forall x)Q$.

(3) $(\forall x)P$ & $(\forall x)Q \Longrightarrow (\forall x)(P$ & $Q)$.

(4) If 'x' is not free in P then $(\exists x)P \Longrightarrow P$.

(5) $(\forall x)P \Longrightarrow (\exists x)P$.

(6) If x is not free in P then P & $(\exists x)Q \Longleftrightarrow (\exists x)(P$ & $Q)$.

(7) If x is not free in P then P or $(\exists x)Q \Longleftrightarrow (\exists x)(P$ or $Q)$.

§ 5. Sets

It is present day practice to define every mathematical object as a set of some kind or other, and we must certainly examine this fundamenta. notion, however briefly.

A set is a collection of objects considered itself as an entity. The objects in the collection are called the <u>elements</u> or <u>members</u> of the set. The symbol for " is a member of " is " $\in$ " (a sort of capital epsilon) , so that " $x \in A$ " is read " x is a member of A ", " x is an element of A ", " x belongs to A " or " x is in A ".

The concept of set is such that a set A is the same object as a set B if and only if A and B have exactly the same members. The identity symbol " = " is thus defined by : " $A = B$ " is an abbreviation for " $(\forall x)(x \in A \Longleftrightarrow x \in B)$ ". We repeat : the meaning of " = " is logical identity, so that $(A = B) \Longleftrightarrow (A$ <u>is</u> $B)$.

We say that a set A is a <u>subset</u> of a set B , or that A is <u>included</u> in B (or that B is a superset of A) if every element of A is an elemen of B. The symbol for indclusion is " $\subset$ " . Thus " $A \subset B$ " is an abbreviation for " $(\forall x)(x \in A \Rightarrow x \in B)$ " . And since " $P \Longleftrightarrow Q$ " has the same meaning as " $(P \Rightarrow Q)$ and $(Q \Rightarrow P)$ ", we see that

$$(A = B) \Longleftrightarrow (A \subset B) \text{ and } (B \subset A) .$$

This is a frequently used way of establishing set identity: we prove that $A = B$ by proving that $A \subset B$ and that $B \subset A$.

A set is defined by specifying its members. If the set is finite the members can actually be listed, and the notation used for the set is curly brackets surrounding a membership list. Thus $\{1, 4, 7\}$ is the set containing just the three numbers 1, 4 and 7. $\{x\}$ is the <u>unit class</u> of x, the set having only the one object x as a member, and $\{x, y\}$ is the <u>pair</u> class of x and y. This notation can be abused to name some infinite sets. Thus $\{2, 4, 6, 8, \ldots\}$ would certainly be considered to be the set of all even positive integers. But generally infinite sets are defined by statement frames. If $P(x)$ is a sentence containing the one free variable 'x', then $\{x : P(x)\}$ is the set of all x such that $P(x)$. In other words, $\{x : P(x)\}$ is that set such that

$$y \in \{x : P(x)\} \Longleftrightarrow P(y) .$$

For example, $\{x : x^2 < 3\}$ is the set of all real numbers x such that $x^2 < 3$, i.e., the open interval $(-\sqrt{3}, \sqrt{3})$, and $y \in \{x : x^2 < 3\} \Longleftrightarrow y^2 < 3$.

A statement frame $P(x)$ with one free variable x can be thought of as stating a <u>property</u> that an object x may or may not have, and $\{x : P(x)\}$ is the set of all objects having that property.

We need the <u>empty set</u>, $\emptyset$, in much the same way that we need zero in arithmetic. If $P(x)$ is <u>never</u> true then $\{x : P(x)\} = \emptyset$

For example, $\{x : x \neq x\} = \emptyset$.

We shall now digress for a moment to be sure that the reader understands the distinction between an object and a name of that object. A chair is not the same thing as the word 'chair' and the number 4 is a mathematical object that is not the same thing as the numeral '4'. The numeral '4' is a name of the number 4, as also are 'four', '2 + 2' and 'IV'. According to our present viewpoint 4 is taken to be some specific set. There is no need in this course to carry logical analysis this far, but some readers may be interested to know that we usually define 4 as $\{0, 1, 2, 3\}$. Similarly $2 = \{0, 1\}$, $1 = \{0\}$ and 0 is the empty set $\emptyset$.

In order to avoid overworking the word 'set' many synonyms are used such as 'class', 'collection', 'family' and 'aggregate'. Thus we might say, "Let a be a family of classes of sets". If a shoe store is a collection of pairs of shoes, then a chain of shoe stores is such a three level object.

§6. Restricted variables

A variable used in mathematics is not allowed to take all objects as values but only the members of a certain set, called the domain of the variable. The domain is sometimes explicitly indicated, but is often only implied. For example, the letter 'n' is customarily used to specify an integer, so that '$(\forall n)P(n)$' would automatically be read 'For every integer n, P(n)'. However, sometimes n is taken to be a positive integer. In case of possible ambiguity or doubt, we would write '$(\forall n \in \mathbb{Z})P(n)$' and read it "For all n in $\mathbb{Z}$, P(n)", where '$\mathbb{Z}$' is the standard symbol for the set of all integers.

Similarly, '$(\exists n \in \mathbb{Z})P(n)$' is read "There exists an n in $\mathbb{Z}$ such that P(n)". Notice that the symbol '$\in$' is here read as the preposition 'in'. The above quantifiers are called restricted quantifiers.

In the same way we have restricted set formation, both implicit and explicit, as in '$\{n : P(n)\}$' and '$\{n \in \mathbb{Z} : P(n)\}$', both of which are read "the set of all integers n such that P(n)".

Restricted variables can be defined as abbreviations of unrestricted variables by:

$$(\forall x \in A)P(x) \iff (\forall x)(x \in A \Rightarrow P(x))$$

$$(\exists x \in A)P(x) \iff (\exists x)(x \in A \;\&\; P(x))$$

$$\{x \in A : P(x)\} = \{x : x \in A \;\&\; P(x)\}$$

Although there is never any ambig ity in sentences containing explicitly restricted variables, it sometimes helps the eye to see the structure of the sentence if the restricting phrases are written in superscript position, as in $(\forall \epsilon^{>0})(\exists n^{\in \mathbb{Z}})$.

Some restriction was implicit on page 2. If the reader agreed that

$(\forall x)(x^2 - 1 = (x-1)(x+1))$ is true, he probably took x to be a ~~real number.~~ member of a ring with 1

§7. Ordered pairs and relations

The notion of an ordered pair is basic in mathematics. According to our general principle the ordered pair $\langle a, b\rangle$ is taken to be a certain set, but which of the many possible models we use is unimportant. The only essential fact is its characterizing property:

$$\langle x, y\rangle = \langle a, b\rangle \Longleftrightarrow x = a \text{ and } y = b .$$

Thus $\langle 1, 3\rangle \neq \langle 3, 1\rangle$.

We do identify the model for the notion of relation (strictly speaking, the no of dyadic relation): a relation is simply a set of ordered pairs. If R is a relation then we say that x has the relation R to y, and write 'xRy', if and only if $\langle x, y\rangle \in R$. We also say that x corresponds to y under R, and call R a correspondence. The set of all first elements occurring in the ordered pairs of a relation R is called the domain of R, and is designated dom R, dom (R) or $\mathcal{D}(R)$. Thus:

$$\operatorname{dom} R = \{x : (\exists y) \langle x, y\rangle \in R\} .$$

The set of second elements is called the range of R:

$$\operatorname{range} R = \{y : (\exists x) \langle x, y\rangle \in R\} .$$

The inverse, R^{-1}, of a relation R is the set of ordered pairs obtained by reversing those of R:

$$R^{-1} = \{\langle x, y\rangle : \langle y, x\rangle \in R\} .$$

A statement frame $P(x, y)$ having two free variables determines a pair of mutually inverse relations R & S, called the graphs of P, by:

$$R = \{\langle x, y\rangle : P(x, y)\}, \quad S = \{\langle y, x\rangle : P(x, y)\}$$

A two variable frame together with a choice of which variable is considered to be first might be called a directed frame. Then a directed frame would have a uniquely determined relation for its graph.

The set $A \times B = \{\langle x, y\rangle : x \in A \ \& \ y \in B\}$ of all ordered pairs with first element in A and second element in B is called the Cartesian product of the sets A and B. A relation R is always a subset of dom R $\times$ range R.

If R is a relation and A is any set, the restriction of R to A, $R \restriction A$, is the subset of R consisting of those pairs with first element in A:

$$R \restriction A = \{\langle x, y\rangle : \langle x, y\rangle \in R \text{ and } x \in A\}$$

Thus $R \restriction A = R \cap (A \times \text{range } R)$, where $C \cap D$ is the intersection of the sets C and D.

If R is a relation and A is any set then the image of A under R, R[A], is the set of second elements of ordered pairs in R whose first elements are in A:

$$R[A] = \{y : (\exists x)(x \in A \ \& \ \langle x, y\rangle \in R)\}$$

Thus $R[A] = \text{range } (R \restriction A)$.

§8. Functions and mappings

A <u>function</u> is a relation f such that each domain element x is paired with only one range element y. This property can be expressed as below:

$$\langle x, y\rangle \in f \text{ and } \langle x, z\rangle \in f \Rightarrow y = z .$$

The y which is thus uniquely determined by f and x is designated $f(x)$:

$$y = f(x) \Longleftrightarrow \langle x, y\rangle \in f$$

One tends to think of a function as being active and a relation which is not a function as being more passive. A function f <u>acts on</u> an element x in its domain to give $f(x)$. We take x and <u>apply</u> f to it; indeed we often call a function an <u>operator</u>. On the other hand if R is a relation but not a function, then there is no particular y related to an element x in its domain and the pairing of x with y is viewed more passively.

A function f is generally defined by specifying its value $f(x)$ for each x in its domain. In this connection a stopped arrow notation is used to indicate the pairing. Thus we say, "Consider the function $x \mapsto x^2$ ", or, "Let $f: x \mapsto x^2$ ", or, "where $f: x \mapsto x^2$ ", which can be read, respectively, as; "Consider the function taking x into x^2 ", "Let f be the function mapping x into x^2 ", "where f is the function taking x into x^2 ". The domain must be understood for this to be meaningful.

If f is a function f^{-1} is of course a relation, but in general it is not a function. If f^{-1} is a function we say that f is <u>one-to-one</u>, and that f is a one-to-one correspondence between its domain and its range. Each $x \in \text{dom } f$ corresponds to only one $y \in \text{range } f$ (f is a function) <u>and</u> each $y \in \text{range } f$ corresponds to only one $x \in \text{dom } f$ (f^{-1} is a function).

The notation

$$f : A \longrightarrow B$$

is read "a(the) function f on A into B" or "f is a function on A into B." The notation implies that f is a function, that $\text{dom } f = A$ and that $\text{range } f \subset B$. Many people feel that the very notion of function should include all of these ingredients. That is, a function should be considered to be an ordered triple $\langle f, A, B \rangle$, where f is a function according to our more limited definition, A is the domain of f, and B is a superset of the range of f, called the <u>codomain</u> of f in this context. We shall use the terms 'map', 'mapping' and 'transformation' for such a triple, so that the notation $f : A \longrightarrow B$ in its totality presents a mapping. Moreover, when there is no question about what set is the codomain we shall often call the function f itself a mapping, since the triple $\langle f, A, B \rangle$ is then determined by f. The two arrow notations can be combined, as in: "Define $f : \mathbb{R} \longrightarrow \mathbb{R}$ by $x \mapsto x^2$.

A mapping $f : A \longrightarrow B$ is said to be <u>injective</u> if f is one-to-one, <u>surjective</u> if $\text{range } f = B$ and <u>bijective</u> if it is both injective and surjective. A bijective mapping $f : A \longrightarrow B$ is thus a one-to-one correspondence between its domain A and its codomain B. Of course a function is always surjective onto its range R, and the statement that f is surjective means that $R = B$ where B is the understood codomain.

§8a Product sets ; index notation

Definition. If S and A are sets then S^A is the set of all functions on A into S.

Thus $\mathbb{R}^{\mathbb{R}}$ is the set of all real-valued functions of one real variable and $S^{\mathbb{Z}^+}$ is the set of all infinite sequences in S (it being understood that an infinite sequence is nothing but a function with domain the set $\mathbb{Z}^+$ of all positive integers). Similarly, if we set $\bar{n} = \{1, \ldots, n\}$ then $S^{\bar{n}}$ is the set of all finite sequences of length n in S.

If B is a subset of A then its <u>characteristic function</u> in A is the function (usually designated χ_B) which has the constant value 1 on B and the constant value 0 on $B' = A-B$. The set of all characteristic functions of subsets of A is thus 2^A (since $2 = \{0, 1\}$). But because this set is in a natural one-to-one correspondence with the set of all subsets of A, we use the same symbol for the latter set. Thus 2^A is also interpreted as the set of all subsets of A.

We shall spend most of the remainder of this section discussing certain definitional ambiguities which mathematicians tolerate.

A finite sequence of l ngth n is equivalent to an ordered n-tuple, so that the set $S^{\bar{n}}$ is essentially the set S^n of all ordered n-tuples in S. However, there is a technical difference. We take note of it now and then dismiss it. The ordered triple $\langle x, y, z\rangle$ is usually defined to be the ordered pair $\langle\langle x, y\rangle, z\rangle$. The reason for this definition is probably that a function of two variables x and y is ordinarily considered to be a function of the single ordered pair variable $\langle x, y\rangle$ so that, for example, a real-valued function of two real variables is a subset of $(\mathbb{R}\times\mathbb{R})\times\mathbb{R}$. But we also consider such a function to be a subset of Cartesian 3-space $\mathbb{R}^3$. Therefore we define $\mathbb{R}^3$ as $(\mathbb{R}\times\mathbb{R})\times\mathbb{R}$; that is, we define the ordered triple $< x, y, z >$ as $<< x, y >, z >$.

The corresponding sequence of length 3 is the set $\{ < 1, x > , < 2, y >, < 3, z > \}$ which, of course, is a different object. But it would serve just as well as a model for an ordered triple, and mathematicians tend to slur over the distinction. We shall have more to say on this point later when we discuss natural isomorphisms ( ). For the moment we shall simply regard $\mathbb{R}^3$ and $\mathbb{R}^{\bar{3}}$ as being the same; an ordered triple is something which can be "viewed" as being either an ordered pair of which the first element is an ordered pair or as a sequence of length 3 (or, for that matter, as an ordered pair of which the <u>second</u> element is an ordered pair).

Similarly we pretend that Cartesian 4-space

$$\mathbb{R}^4 \text{ is } \mathbb{R}^{\bar{4}} \text{ or } \mathbb{R}^2\times\mathbb{R}^2 \text{ or } \mathbb{R}^1\times\mathbb{R}^3 = \mathbb{R}\times((\mathbb{R}\times\mathbb{R})\times\mathbb{R}) \text{ etc.}$$

Clearly we are in effect assuming an associative law for ordered pair formation that we don't really have. This is one of the prices we pay for the precision of set theory; in vaguer days there would have been a single fuzzy notion.

The device of indices is much used in mathematics and it also has ambigous implications which we should examine.

An indexed collection, as a set, is nothing but the range set of a function, the indexing function, and a particular indexed object, say x_i, is simply the value of that function at the domain element i. If the set of indices is I the indexed set is designated $\{ x_i : i \in I \}$ or $\{ x_i \}_{i \in I}$ (or $\{x_i\}_{i=1}^{\infty}$ in case $I = \mathbb{Z}^+$). However, this notation suggests that we view the indexed set as being obtained by letting the index run through the index set I and collecting the indexed objects. That is, an indexed set is viewed as being the set together with the indexing function. This ambivalence is reflected in the fact that the same notation frequently designates the mapping. Thus we refer to the sequence $\{ x_n \}_{n=1}^{\infty}$, where, of course, the sequence is the mapping $n \longrightarrow x_n$. We believe that if the reader examines his idea of a sequence he will find this ambiguity present. He doesn't mean just the set, nor just the mapping, but the mapping with emphasis on its range, or the range "together with" the mapping. But since set theory cannot reflect these nuances in any simple and graceful way we shall take an indexed set to be the indexing function. Of course, the same range object may be repeated with different indices ; there is no implication that an indexing is one-to-one. Notice also that indexing imposes no restriction on the set being indexed ; any set can at least be self-indexed (by the identity function) .

Except for the ambiguous $\{x_i : i \in I\}$ there is no universally used notation for the indexing function. Since x_i is the value of the function at i we might think of x_i as another way of writing $x(i)$ in which case we designate the function x. We certainly do this in the case of ordered n-tuplets when we say, "Consider the n-tuplet $x = < x_1, \ldots, x_n >$". On the other hand, there is no compelling reason to set $x_i = x(i)$. All that we are compelled to do is to set $x_i = f(i)$ if we have called the indexing function f.

We come now to the general definition of Cartesian product. Earlier we argued that the Cartesian product $A \times B \times C$ is the set of all ordered triples $x = < x_1, x_2, x_3 >$ such that $x_1 \in A$, $x_2 \in B$ and $x_3 \in C$. More generally $A_1 \times A_2 \times \ldots \times A_n$ is the set of ordered n-tuples $x = < x_1, \ldots, x_n >$ such that $x_i \in A_i$ for $i = 1, \ldots, n$. We also use the notation $\Pi_{i=1}^{n} A_i$ for the Cartesian product $A_1 \times \ldots \times A_n$, and if we interpret an ordered n-tuplet as a function on $\bar{n} = \{ 1, \ldots, n \}$, we have

$$\Pi_{i=1}^{n} A_i \text{ is the set of all functions } x \text{ with domain } \bar{n} \text{ such that } x_i \in A_i \text{ for all } i \in \bar{n} .$$

This rephrasal generalizes almost verbatim to give us the notion of the Cartesian product of an arbitrary indexed collection of sets.

DEFINITION. The Cartesian product $\Pi_{i \in I} S_i$ of the indexed collection of sets $\{ S_i : i \in I \}$ is the set of all functions f with domain I such that $f(i) \in S_i$ for all $i \in I$.

We can also use the notation $\Pi \{ S_i : i \in I \}$ for the product and f_i for the value $f(i)$.

§9. The Boolean operations

Let S be a fixed domain and let $\mathcal{F}$ be a family of subsets of S. The union of $\mathcal{F}$, or the union of all the sets in $\mathcal{F}$, is the set of all elements $x \in S$ such that x lies in at least one set in $\mathcal{F}$. We designate the union $\bigcup \mathcal{F}$, or $\bigcup_{A \in \mathcal{F}} A$, and thus have

$$\bigcup \mathcal{F} = \{x : (\exists A^{\in \mathcal{F}})(x \in A)\}$$

Often we consider the family $\mathcal{F}$ to be indexed. That is, we suppose given a set I (the set of indices) and a surjective mapping $A : I \longrightarrow \mathcal{F}$, so that $\mathcal{F} = \{ A_i : i \in I \}$.

Then the union of the indexed collection is designated $\bigcup_{i \in I} A_i$ or $\bigcup \{A_i : i \in I\}$. The device of indices has both technical and psychological advantages, and we shall generally use it.

If $\mathcal{F}$ is finite, and either it or the index set is listed, then a different notation is used for its union. If $\mathcal{F} = \{A, B\}$ we designate the union $A \cup B$, a notation that displays the listed names.

If $\mathcal{F} = \{A_i : i = 1, \ldots, n\}$, we generally write '$A_1 \cup A_2 \cup \ldots \cup A_n$' or '$\bigcup_{i=1}^{n} A_i$' for $\bigcup \mathcal{F}$.

The intersection of the indexed family $\{A_i\}_{i \in I}$, designated $\bigcap_{i \in I} A_i$, is the set of all points that lie in every A_i. Thus:

$$x \in \bigcap_{i \in I} A_i \iff (\forall_i^{\in I})(x \in A_i)$$

For an unindexed family $\mathcal{F}$ we use the notation $\bigcap \mathcal{F}$ or $\bigcap_{A \in \mathcal{F}} A$, and if $\mathcal{F} = \{A, B\}$ then $\bigcap \mathcal{F} = A \cap B$.

The complement, A', of a subset of S is the set of element $x \in S$ not in A: $A' = \{x^{\in S} ; x \notin A\}$. The law of de Morgan says that <u>the complement of an intersection is the union of the complements</u>.

$$(\bigcap_{i \in I} A_i)' = \bigcup_{i \in I} (A_i') .$$

This an immediate consequence of the rule for negating quantifiers. Since $\sim(\forall i) P(i) \Longleftrightarrow (\exists i) \sim P(i)$, we have $(\forall x)[\sim(\forall i)(x \in A_i) \Longleftrightarrow (\exists i)(x \notin A_i)]$, which says exactly that

$$(\forall x)(x \in (\bigcap_i A_i)' \Longleftrightarrow x \in \bigcup_i (A_i')) .$$

If we set $B_i = A_i'$ and take complements again, we obtain the dual form: $(\bigcup_{i \in I} B_i)' = \bigcap_{i \in I} (B_i')$

Other principles of quantification yield the laws:

$$B \cap (\bigcup_{i \in I} A_i) = \bigcup_{i \in I} (B \cap A_i)$$

(from $P \;\&\; (\exists x) Q(x) \Longleftrightarrow (\exists x)(P \;\&\; Q(x))$),

$$B \cup (\bigcap_{i \in I} A_i) = \bigcap_{i \in I} (B \cup A_i) ,$$

$$B \cap (\bigcap_{i \in I} A_i) = \bigcap_{i \in I} (B \cap A_i) ,$$

$$B \cup (\bigcup_{i \in I} A_i) = \bigcup_{i \in I} (B \cup A_i) .$$

In the case of two sets, these laws imply the familiar laws of set algebra:

$$(A \cup B)' = A' \cap B' , \quad (A \cap B)' = A' \cup B' \qquad \text{(de Morgan)}$$

$$(A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Even here, thinking in terms of indices makes the laws more intuitive. Thus

$$(A_1 \cap A_2)' = A_1' \cup A_2'$$

is obvious when thought of as an equivalence between 'not always in' and 'sometimes not in'.

The family $\mathcal{F}$ is <u>disjoint</u> $\Longleftrightarrow$ distinct sets in $\mathcal{F}$ have no elements in common, i.e., $\Longleftrightarrow (\forall X, Y \in \mathcal{F})(X \neq Y \Rightarrow X \cap Y = \emptyset)$. For an indexed family $\{A_i\}_{i \in I}$ the condition becomes $i \neq j \Rightarrow A_i \cap A_j = \emptyset$. If $\mathcal{F} = \{A, B\}$ we simply say that A and B are disjoint.

Given $f : U \longrightarrow V$ and an indexed family $\{B_i\}$ of subsets of V then we have the following important identities:

$$f^{-1}[\cup_i B_i] = \cup_i f^{-1}[B_i]$$

$$f^{-1}[\cap_i B_i] = \cap_i f^{-1}[B_i]$$

and, for a single set $B \subset V$,

$$f^{-1}[B'] = (f^{-1}[B])' .$$

For example, $x \in f^{-1}[\cap_i B_i] \Longleftrightarrow f(x) \in \cap_i B_i \Longleftrightarrow (\forall i)(f(x) \in B_i)$ $\Longleftrightarrow (\forall i)(x \in f^{-1}[B_i]) \Longleftrightarrow x \in \cap_i f^{-1}[B_i]$.

One, but not the other two, of these three identities remains valid when f is replaced by any relation R.

§10. Composition

If, we are given maps $f : A \longrightarrow B$ and $g : B \longrightarrow C$ then the <u>composition</u> of g with f, $g \circ f$, is the map of A into C defined by:

$$(g \circ f)(x) = g(f(x)) \quad \text{for all } x \in A .$$

Notice that the codomain of f has to be the domain of g in order for $g \circ f$ to be defined.

This operation is perhaps the basic binary operation of mathematics.

Lemma. Composition satisfies the associative law: $f \circ (g \circ h) = (f \circ g) \circ h$.

Proof. $(f \circ (g \circ h))(x) = f((g \circ h)(x)) = f(g(h(x))) = (f \circ g)(h(x)) = ((f \circ g) \circ h)(x)$, for all $x \in \text{dom } h$.

If A is a set, the identity map $I_A : A \longrightarrow A$ is the mapping taking every $x \in A$ into itself. Thus $I_A = \{\langle x, x \rangle : x \in A\}$. If f maps A into B then clearly

$$f \circ I_A = f = I_B \circ f .$$

If $g : B \longrightarrow A$ is such that $g \circ f = I_A$ then we say that g is a <u>left inverse</u> of f and that f is a <u>right inverse</u> of g.

Theorem. A mapping $f : A \longrightarrow B$ has a left inverse $\Longleftrightarrow$ f is injective, and f has a right inverse $\Longleftrightarrow$ f is surjective.

Proof. If $g \circ f = I_A$ then $f(x) = f(y) \Rightarrow x = g(f(x)) = g(f(y)) = y$ and f is injective. Conversely, if f is injective, with range R, then f^{-1} is a function with domain R. Define g on B by $g \restriction R = f^{-1}$ and $g \restriction (B - R)$ is any function on $B-R$ into A. Then $g \circ f = I_A$. Next, suppose that $f \circ h = I_B$. Then each

$y \in B$ can be written $y = f(h(y))$ and so f is surjective. Conversely, if f is surjective we can form a subset $C \subset A$ by choosing one point x_y from $f^{-1}[y]$ for each $y \in B$. Then the pairs $< y, x_y >$ form a function h : $B \longrightarrow A$ and $f \circ h = I_B$.

Lemma. If the mapping $f : A \longrightarrow B$ has both a right inverse h and a left inverse g they must necessarily be equal.

Proof. This is just algebraic juggling and works for any associative operation. We have

$$h = I_A \circ h = (g \circ f) \circ h = g \circ (f \circ h) = g \circ I_B = g .$$

In this case we call the uniquely determined map $g : B \longrightarrow A$ such that $f \circ g = I_B$ and $g \circ f = I_A$ the <u>inverse</u> of f. We then have:

<u>Corollary:</u> A mapping $f : A \longrightarrow B$ has an inverse if and only if it is bijective.

Now let $\mathfrak{S}(A)$ be the set of all bijections $f : A \longrightarrow A$. Then $\mathfrak{S}(A)$ is closed under the binary operation of composition and

(1) $f \circ (g \circ h) = (f \circ g) \circ h$ for all $f, g, h \in \mathfrak{S}$.

(2) There exists a unique $I \in \mathfrak{S}(A)$ such that $f \circ I = I \circ f = f$ for all $f \in \mathfrak{S}$.

(3) For each $f \in \mathfrak{S}$ there exists a unique $g \in \mathfrak{S}$ such that $f \circ g = g \circ f = I$.

Any set G closed under a binary operation having these properties is called a <u>group</u> with respect to that operation. Thus $\mathfrak{S}(A)$ is a group with respect to composition.

Composition can also be defined for relations as follows. If $R \subset A \times B$ and $S \subset B \times C$ then $S \circ R \subset A \times C$ is defined by:

$$\langle x, z \rangle \in S \circ R \iff (\exists y^{\in B})(\langle x, y \rangle \in R \;\&\; \langle y, z \rangle \in S)$$

If R and S are mappings this definition agrees with our earlier one.

§11. Partitions and equivalence relations

A partition of a set A is a disjoint family $\mathcal{F}$ of sets whose union is A. We call the elements of $\mathcal{F}$ 'fibers', and say that $\mathcal{F}$ fibers A or is a fibering of A. Passing from a set A to a fibering of A is one of the principal ways of forming new mathematical objects.

Any function automatically fibers its domain D into the sets D_y on which f has the constant value y. Here, of course, $D_y = f^{-1}[y] = \{x \in D : f(x) = y\}$ for each y in the range R of f, and it is clear that $y \neq z \Rightarrow D_y \cap D_z = \emptyset$ and that $D = \bigcup\{D_y : y \in R\}$. Also, the mapping $y \longrightarrow D_y$ is a one-to-one correspondence between R and the set $\mathcal{F}$ of fibers.

Different functions on the same domain can define the same fibering, but there is one "canonical" function associated with a fibering, namely, the "projection" mapping π from D to $\mathcal{F}$ which assigns to each element $x \in D$ the fiber $\pi(x)$ which contains it.

A function f also defines a natural relation $\sim$ on its domain: $x_1 \sim x_2 \Longleftrightarrow f(x_1) = f(x_2)$. Thus $x_1 \sim x_2 \Longleftrightarrow x_1$ and x_2 belong to the same fiber , or, in terms of the projection map π into the fiber space, $x_1 \sim x_2 \Longleftrightarrow \pi(x_1) = \pi(x_2)$.

This relation $\sim$ is a subset of $D \times D$ which is reflexive ($x \sim x$ for every $x \in D$), symmetric ($x \sim y \Rightarrow y \sim x$) and transitive ($x \sim y$ and $y \sim z \Rightarrow x \sim z$). Any relation $E \subset D \times D$ having these properties is said to be an equivalence relation on D. The most important fact to be noted in this section is that, conversely, any equivalence relation E on a set D arises in the above way from a function on D (or a fibering of D).

To see this, define $[x]$ as the set of elements $y \in D$ equivalent to x, $[x] = \{y : xEy\}$, and let $\mathcal{F}$ be the family of subsets of D obtained this way.

Thus $\mathcal{F}$ is the range of the mapping $f : D \longrightarrow \mathcal{F}$ defined by $f(x) = [x]$.

Theorem. $\mathcal{F}$ is a partition of D, E is its induced equivalence relation and f is its projection mapping.

Proof. We start by observing that

$$(*) \qquad xEy \Longleftrightarrow [x] = [y].$$

For xEy and $yEz \Rightarrow xEz$, so that $[y] \subset [x]$. Similarly $xEy \Rightarrow yEx \Rightarrow [x] \subset [y]$. Thus $xEy \Rightarrow [x] = [y]$. Conversely $y \in [y]$ (reflexivity) and $[x] = [y] \Rightarrow y \in [x] \Rightarrow xEy$. Thus E is the equivalence relation arising from the mapping f. Moreover, the same condition (*) says that $y \in [x] \Longleftrightarrow f(y) = f(x)$, so that $[x] = D_x = \pi(x)$, q. e. d.

The set $\mathcal{F}$ of fibers arising in this way from an equivalence relation E on a set D will be designated D/E and called the quotient of D by E.

The fundamental role this argument plays in mathematics is due to the fact that in many important situations equivalence relations occur as the primary object, and then are used to define partitions and functions. We give two examples.

Let $\mathbb{Z}$ be the integers (positive, negative and zero). A fraction can be considered to be an ordered pair $\langle m, n \rangle$ of integers with $n \neq 0$. The set of all fractions is thus $\mathbb{Z} \times (\mathbb{Z} - \{0\})$. Two fractions $\langle m, n \rangle$ and $\langle p, q \rangle$ are "equal" if and only if $mq = np$, and equality is checked to be an equivalence relation. The equivalence class $[\langle m, n \rangle]$ is the object taken to be the rational number m/n. Thus *the rational number system* $\mathbb{Q}$ is the set of fibers in a partition of $\mathbb{Z} \times (\mathbb{Z} - \{0\})$.

Next, we choose a fixed integer $p \in \mathbb{Z}$ and define a relation E on $\mathbb{Z}$ by $m E n \iff p$ divides $m-n$. E is an equivalence relation and the set $\mathbb{Z}_m$ of its equivalence classes is called the integers modulo p. It is easy to see that mEn if and only if m and n have the same remainder when divided by p, so that in this case there is an easily calculated function f, $f(m)$ being the remainder after dividing m by p, which defines the fibering. The set of possible remainders is $\{0, 1, \ldots, p-1\}$ so that $\mathbb{Z}_p$ contains p elements.

Functions can be "lifted" from a set D to a fibering of D by the following theorem.

Theorem. Let g be a function on D and let E be an equivalence relation on D. Then there exists a function $\overline{g}$ on the quotient space D/E such that $g = \overline{g} \circ \pi$ if and only if g is constant on the fibers of E.

Proof. If $g[A]$ is a unit class $\{x_A\}$ for each fiber A, we define $\overline{g}(A)$ as x_A and have, obviously, $g = \overline{g} \circ \pi$. The converse is even easier.

§12. Duality.

There is another elementary but very important phenomenon called duality which occurs in practically all branches of mathematics. Let $F : A \times B \longrightarrow C$ be any function of two variables. It is obvious that if x is held fixed in $F(x, y)$ then we obtain a function of y for each value of x. That is, for each x there is a function $f_x : B \longrightarrow C$ defined by $f_x(y) = F(x, y)$. And then $x \longmapsto f_x$ is a mapping f of A into C^B. Similarly, each $y \in B$ yields a function $g_y \in C^A$, where $g_y(x) = F(x, y)$.

Now suppose conversely that we are given a mapping $f : A \longrightarrow C^B$. For each $x \in A$ we designate the corresponding value of f in index notation as f_x, and define $F : A \times B \longrightarrow C$ by $F(x, y) = f_x(y)$. We are now clearly back where we started. The mappings $f : A \longrightarrow C^B$, $F : A \times B \longrightarrow C$ and $g : B \longrightarrow C^A$ are thus equivalent to each other, and can be thought of as being three different ways of viewing the same phenomenon. The extreme mappings f and g will be said to be dual to each other.

The mapping f is the indexed family of functions $\{f_x : x \in A\} \subset C^B$. Now suppose that $\mathcal{F} \subset C^B$ is an unindexed collection of functions on B into C. We can index $\mathcal{F}$ by some index set A if we want, but we can also treat $\mathcal{F}$ itself as though it were A. That is, we can define $F : \mathcal{F} \times B \longrightarrow C$ by $F(f, y) = f(y)$. Then $g : B \longrightarrow C^{\mathcal{F}}$ is defined by $g_y(f) = f(y)$. What is happening here is simply that in the expression $f(y)$ we regard both symbols as variables, so that $f(y)$ is a function on $\mathcal{F} \times B$. Then when we hold y fixed we have a function on $\mathcal{F}$, mapping $\mathcal{F}$ into C.

We shall see some important applications of this duality principle as our subject develops. For example, we shall regard an $m \times n$ matrix as also being a sequence of m column vectors, as well as a sequence of n row vectors.

In the same vein, a sequence $f_1, \ldots, f_n$ of functions on A into B can be regarded as a single function on A into $B^{\bar{n}}$, where we have set $\bar{n} = \{1, \ldots, n\}$ since n itself is taken to be the set $\{0, \ldots, n-1\}$. Also, we shall regard a finite dimensional vector space V as being its own second conjugate space $(V^*)^*$, so that V and V^* are in duality.

It is instructive to look at elementary Euclidean geometry from this point of view. Today we regard a straight line as being a set of geometric points. An older and more neutral view is to take points and lines as being two different kinds of primitive objects. Accordingly, let A be the set of all points (so that A is the Euclidean plane as we now view it) and let B be the set of all straight lines. Let F be the <u>incidence</u> function: $F(p, \ell) = 1$ if p and ℓ are incident (p is "on" ℓ, ℓ is "on" p) and $F(p, \ell) = 0$ otherwise. Thus F maps $A \times B$ into $\{0, 1\}$. Then for each $\ell \in B$ the function $g_\ell(p) = F(p, \ell)$ is the characteristic function of the set of points that we think of as being the line ℓ ($g_\ell(p) = 1$ if p is on ℓ and is 0 if p is not on ℓ.) Thus each line determines the set of points that are on it. But, dually, each point p determines the set of lines ℓ "on" <u>it</u>, through <u>its</u> characteristic function $f_p(\ell)$. Thus, in complete duality we can regard a line as being a set of points and a point as being a set of lines. This duality aspect of geometry is basic in projective geometry.

Chapter I

VECTOR SPACES

§ 1. Fundamental notions ; Function spaces.

The calculus of functions of more than one variable unites the calculus of one variable, which the reader presumably knows, with the theory of vector spaces, and the adequacy of its treatment depends directly on the extent to which vector space theory is really used. The theories of differential equations and differential geometry are similarly based on a mixture of calculus and vector space theory. Such "vector calculus" and its applications constitute the subject matter of this course, and in order for our treatment to be completely satisfactory we shall have to spend considerable time at the beginning studying vector spaces themselves. This we do in the first two chapters, the present chapter being devoted to some of the algebras of vector spaces and the next chapter to their limit theory.

First we give the abstract definition of a vector space and look at an important class of concrete examples. We then develop the notions of subspace, linear combination and linear transformation, and establish some of their most elementary properties.

A. A vector space is a collection of objects that can be added to each other and multiplied by numbers, the two operations being required to satisfy certain laws of algebra.

DEFINITION. Let V be a set and let $A : V \times V \to V$ and $S : \mathbb{R} \times V \to V$ be given mappings. Write '$\alpha + \beta$' for '$A(\alpha, \beta)$' and '$x\alpha$' for $S(x, \alpha)$'. Then V is a vector space with respect to A and S if and only if

A1	$\alpha + (\beta+\gamma) = (\alpha+\beta)+\gamma$	for all $\alpha, \beta, \gamma \in V$
A_2	$\alpha+\beta = \beta+\alpha$	for all $\alpha, \beta \in V$
A_3	There exists an element $O \in V$ such that $\alpha+O = \alpha$ for all $\alpha \in V$	
A_4	For every $\alpha \in V$ there exists $\beta \in V$ such that $\alpha+\beta = 0$	
S1	$(xy)\alpha = x(y\alpha)$	for all $x, y \in \mathbb{R}$, $\alpha \in V$
S2	$(x+y)\alpha = x\alpha + y\alpha$	for all $x, y \in \mathbb{R}$, $\alpha \in V$
S3	$x(\alpha+\beta) = x\alpha + x\beta$	for all $x \in \mathbb{R}$, $\alpha, \beta \in V$
S4	$1\alpha = \alpha$	for all $\alpha \in V$.

In contexts where it is clear (as it generally is) what operations are intended we refer simply to the vector space V.

The simplest example of a vector space is the set $V = \mathbb{R}^A$ of all real-valued functions on a set A, together with the natural operations of adding two functions and multiplying a function by a number. That is, $f+g$ is the function defined by $(f+g)(a) = f(a)+g(a)$, and cf is the function defined by $(cf)(a) = c(f(a))$. The laws A1-S4 follow at once from these definitions and the corresponding laws of algebra for the real number system. For example, the equation

$(x+y)f = xf + yf$ means that $((x+y)f)(a) = (xf)(a) + (yf)(a)$ for all $a \in A$.
But $((x+y)f)(a) = (x+y)(f(a)) = x(f(a)) + y(f(a)) = (xf)(a) + (yf)(a)$, where we have used the definition of scalar multiplication in $\mathbb{R}^A$, the distributive law in $\mathbb{R}$ and the definition of scalar multiplication in $\mathbb{R}^A$, in that order.

The set A can be anything at all. If $A = \mathbb{R}$ then $V = \mathbb{R}^{\mathbb{R}}$ is the space of all real valued functions of one real variable. If $A = \mathbb{R} \times \mathbb{R}$, then $V = \mathbb{R}^{\mathbb{R} \times \mathbb{R}}$ is the space of all real-valued functions of two real variables. If $A = \{1, \ldots, n\} = \bar{n}$, then $V = \mathbb{R}^{\bar{n}}$ is Cartesian n-space.

B. If one vector space is included in another vector space and has the same vector operations we call it a subspace of the larger space.

DEFINITION. A subset W of a vector space V which is itself a vector space under the (restrictions to W of the)operations of V is called a <u>subspace</u> of V.

LEMMA 1. If a subset W of a vector space V is non-empty and is closed under the operations of V then W is a subspace of V.

<u>Proof:</u> The closure of W under addition means simply that $\alpha+\beta$ is in W whenever α and β are both in W, and similarly for scalar multiplication.

The universally quantified laws A1, A2, and S1-S4 hold in W because they hold in the larger set V. In order to check A3 we notice that by hypothesis there is <u>some</u> $\beta \in W$. But then $0 = 0\beta \in W$, since W is closed under scalar multiplication. And, for every $\alpha \in W$, $-\alpha = (-1)\alpha \in W$ for the same reason. Therefore W is a vector space.

Consider for example $V = \mathbb{R}^A$ where A is the closed interval $[a,b] \subset \mathbb{R}$, and let $\mathcal{C}([a,b])$ be the set of <u>continuous</u> real-valued functions on $A = [a,b]$. Then $\mathcal{C}([a,b])$ is a subset of V which is closed unter the operations of V (i.e., $f+g$ and cf are continuous whenever f and g are), and so is a subspace of V.

Subspaces of the vector spaces $\mathbb{R}^A$ are called <u>function spaces</u>. Thus a function space is a collection of real-valued functions with common domain which is closed under addition and multiplication by scalars.

What we have defined so far ought to be called the notion of a <u>real</u> vector space and its subspaces. There is an analogous notion of a <u>complex</u> vector space, examplified by the space $\mathbb{C}^A$ of all complex-valued functions on A. In fact, if the reader knew what is meant by a field F, we could give a single general definition of a vector space over F, in which scalar multiplication is by the elements of F, the standard example being the space $V = F^A$ of all functions on A to F.

LEMMA 2. The intersection $W = W_1 \cap W_2$ of two subspaces of a vector space V is itself a subspace. In fact, the intersection $W = \bigcap_{i \in I} W_i$ of any collection $\{ W_i : i \in I \}$ of subspaces of V is itself a subspace.

<u>Proof:</u> If α and β are in W then they are in every W_i. But then $\alpha+\beta$ is in every W_i and so is in W. Thus W is closed under

addition. Similarly W is closed under scalar multiplication. Moreover, W is non-empty since O is in every W_i and so is in W. Thus W is a subspace, by Lemma 1.

COROLLARY. If A is any <u>subset</u> of a vector space V then there is a uniquely determined smallest <u>subspace</u> of V which includes A.

<u>Proof:</u> Let $\{W_i : i \in I\}$ be the collection of all the subspaces which include A and let W be the intersection of this collection. Then W is a subspace including A, and any other such subspace U is one of the W_i's and so includes W. Moreover, if $U \neq W$ then U is properly larger than W. Therefore W is the unique smallest subspace including A.

We call W the <u>linear span</u> of A, at the subspace <u>generated</u> by A, and designate it $L(A)$.

C. We gave above a "non-constructive" definition of the linear span $L(A)$ of a subset $A \subset V$. In order to see what the elements of $L(A)$ look like we must consider linear combinations. The associative and commutative laws for vector addition imply that a finite collection of vectors $\{\alpha_i\}_1^n$ has a uniquely determined sum which is independent of the order in which the α_i are taken and of the way in which they are grouped. Since $\bar{n} = \{1,\ldots,n\}$ has a natural ordering, the best way to express this independence is to take a finite <u>unordered</u> index set I, a corresponding indexed collection of vectors $\{\alpha_i : i \in I\}$ and then observe that there is a uniquely determined vector $\Sigma_{i\in I}\alpha_i$ which can be obtained by adding the α_i in any order and in any grouping. This "obvious" fact requires a rather fussy proof by mathematical induction, but we shall simply assume it. If $c = \{c_i : i \in I\}$ is a corresponding set of scalars then $\Sigma_{i\in I} c_i \alpha_i$ is called a <u>linear combination</u> of the α_i's. Thus $3 \sin x - \cos x + 2e^x$ is a linear combination of $\{\sin x, \cos x, e^x\}$ and any polynomial of degree at most 5 is a linear combination of the set of monomials $\{x^i\}_{i=0}^5$.

LEMMA 3. The linear span $L(\{\alpha_i\})$ of a finite set of vectors $\{\alpha_i : i \in I\}$ is exactly the set of all linear combinations of these vectors.

<u>Proof:</u> We have that

$$(\Sigma x_i \alpha_i) + (\Sigma y_i \alpha_i) = \Sigma (x_i + y_i)\alpha_i ,$$

because the left side equals $\Sigma(x_i\alpha_i + y_i\alpha_i)$ when it is regrouped by pairs, and then S2 gives the right side. Since S3 (and mathematical induction) clearly imply that $c\,\Sigma x_i \alpha_i = \Sigma (c x_i)\alpha_i$, we see that the set of all linear combinations of $\{\alpha_i : i \in I\}$ is a subspace W. W contains each α_i (why ?), and if V is a subspace including $\{\alpha_i\}$ then $W \subset V$ (why ?). Thus $W = L(\{\alpha_i\})$.

We define an element β to be a linear combination of an <u>infinite</u> indexed set $\{\alpha_i : i \in I\}$ if β is a linear combination of $\{\alpha_i : i \in I_1\}$ for some finite subset $I_1 \subset I$. This is the only possible definition since we cannot form infinite sums in a purely algebraic situation. However, when we add a notion of convergence we shall be able to consider infinite linear combinations ( ).

Notice that if $I_1 \subset I_2$ and if $\beta = \Sigma_{I_1} c_i\alpha_i$ then also $\beta = \Sigma_{I_2} c_i\alpha_i$ if we set $c_i = 0$ for $i \in I_2 - I_1$. Therefore if $\beta_1 = \Sigma_{I_1} c_i\alpha_i$ and $\beta_2 = \Sigma_{I_2} d_i\alpha_i$ then $\beta_1 + \beta_2 = \Sigma_J (c_i + d_i)\alpha_i$ where $J = I_1 \cup I_2$, $c_i = 0$ for $i \in I_2 - I_1$ and $d_i = 0$ for $i \in I_1 - I_2$. We thus have the corollary :

COROLLARY. The set of linear combinations of any infinite (indexed) set of vectors is its linear span.

For example, the set of all linear combinations of the infinite set of monomials $\{x^i\}_0^\infty$ is the vector space of all polynomials.

Any subset A of a vector space V can be considered to be self indexed (by the identity function) and a linear combination of the vectors in A is therefore obtained by choosing a finite subset $A_1 \subset A$ and a scalar x_α for each $\alpha \in A_1$, and then forming the sum $\Sigma_{\alpha \in A_1} x_\alpha \alpha$. If the reader feels more comfortable with the integers, he can choose a finite <u>sequence</u> $\{\alpha_i\}_1^n \subset A$ and an n-tuple $x = \{x_i\}_1^n$ of numbers, and then form the sum $\Sigma_{i=1}^n x_i\alpha_i$. In any case, the word "indexed" can be omitted from the corollary.

D. The general function space $\mathbb{R}^A$ and the subspace $\mathcal{C}([a, b])$ of $\mathbb{R}^{[a, b]}$ both have the property that in addition to being closed under the vector operations they are also closed under the operation of multiplying two functions together. The pointwise product of two functions is a function $((fg)(a) = f(a)g(a))$ and the product of two continuous functions is continuous. With respect to these three operations, addition, multiplication and scalar multiplication, they are called algebras. If the reader noticed this extra operation he may have wondered why, at least in the context of function spaces, we bother with the notion of vector space. Why not study the richer structure these spaces have as algebras ? The answer is that the vector operations are exactly the operations that are "preserved" by many of the most important mapping of functions. For example, define $T : \mathscr{C}([a,b]) \longrightarrow \mathbb{R}$ by $T(f) = \int_a^b f(t)dt$. Then the laws of the integral calculus say that $T(f+g) = T(f) + T(g)$ and $T(cf) = cT(f)$. Thus T "preserves" the vector operations. Or we can say that T "commutes" with the vector operations, since plus followed by T equals T followed by plus. Notice, however, that T does not preserve multiplication : it is not true in general that $T(fg) = T(f)T(g)$.

DEFINITION. If V and W are vector spaces then a mapping $T : V \longrightarrow W$ is a linear transformation if and only if $T(\alpha+\beta) = T(\alpha)+T(\beta)$ for all $\alpha, \beta \in V$, and $T(x\alpha) = xT(\alpha)$ for all $\alpha \in V$, $x \in \mathbb{R}$.

These two "preservation of operation" equations are often combined into a single "preservation of linear combination" equation :

$$T(x_1\alpha_1 + x_2\alpha_2) = x_1T(\alpha_1) + x_2T(\alpha_2) \quad \text{for all } \alpha_1, \alpha_2 \in V$$

and all $x_1, x_2 \in \mathbb{R}$. More generally we have

$$T(\Sigma_1^n x_i \alpha_i) = \Sigma_1^n x_i T(\alpha_i) .$$

Other examples of linear transformations will be given in the exercises ; we go on now to some elementary properties of linear transformations.

LEMMA 4. If $T : V \longrightarrow W$ is linear then $T^{-1}[Y]$ is a subspace of V whenever Y is a subspace of W and $T[X]$ is a subspace of W whenever X is a subspace of V .

<u>Proof:</u> If $\{\alpha_i\} \subset X$, the law $T(\Sigma x_i \alpha_i) = \Sigma x_i T(\alpha_i)$, read backward, shows that $L(T[X]) \subset T[L(X)] = T[X]$. Thus $T[X]$ is its own linear span, and so is a subspace. The other proof is almost the same and we omit it.

COROLLARY . The null set of T , $T^{-1}(0) = \{\alpha \in V : T(\alpha) = 0\}$ is a subspace of V and range of T , $T[V]$ is a subspace of W .

LEMMA 5. A linear mapping T is injective if and only if its null space is $\{0\}$.

<u>Proof:</u> To say that T is injective means that $T(\alpha) = T(\beta)$ if and only if $\alpha = \beta$, i.e., $T(\alpha - \beta) = 0$ if and only if $\alpha - \beta = 0$,

i.e., $T(\gamma) = 0$ if and only if $\gamma = 0$, i.e., the null space $N(T)$ is $\{0\}$.

DEFINITION. A linear map $T: V \longrightarrow W$ which is bijective is called an isomorphism. Two vector spaces V and W are isomorphic if and only if there exists an isomorphism between them.

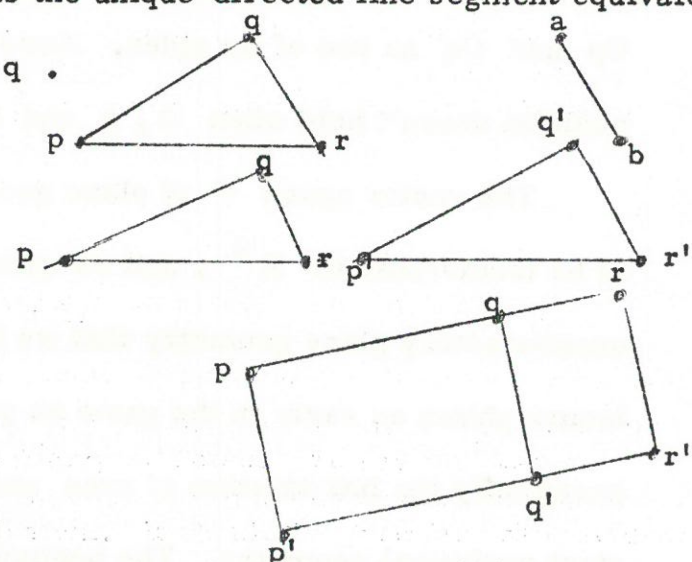
For example the map $< c_1, \ldots, c_n > \longrightarrow \Sigma_0^{n-1} c_{i+1} x^i$ is an isomorphism of $\mathbb{R}^n$ with the vector space of all polynomials of degree $< n$.

Isomorphic spaces "have the same form" and are identical as abstract vector spaces. That is, they cannot be distinguished from each other solely on the basis of vector properties which they do or do not have.

§ 2. Geometric vectors.

This section is not essential to the course. It concerns plane geometry and the transition from plane geometry to analytic geometry, which is part of our assumed background. However it points up the vector aspects of that transition, which may be unfamiliar to the reader, and should at least be read through. Since it is difficult to develop the theory of geometric vectors adequately on the basis of synthetic plane geometry we shall content ourselves with a brief description which will include no proofs.

A geometric vector is represented by a directed line segment. It cannot taken to be a directed line segment, since a pair of different directed line segments represent the same geometric vector if they are parallel, equal in length and similarly oriented. We say that two such directed line segments are equivalent. It should be geometrically evident that this does define an equivalence relation on the set of all line segments, and a geometric vector can be identified with an equivalence class. Let pq be the directed line segment from p to q and $\overrightarrow{pq}$ the corresponding geometric vector. We define the sum $\overrightarrow{pq} + \overrightarrow{ab}$ of the geometric vectors $\overrightarrow{pq}$ and $\overrightarrow{ab}$ as $\overrightarrow{pr}$ where qr is the unique directed line segment equivalent to ab and with initial point at q.



It follows from theorems in geometry that the sum is independent of which directed line segments represent the vectors, and thus is uniquely defined. The product $x(\overrightarrow{pq})$ is the vector $\overrightarrow{pr}$ where pr is parallel to pq, with length $|x|$ times the length of pq and similarly or oppositely oriented depending on whether x is positive or negative. Similarity theorems in geometry imply that again the geometric vector $x(\overrightarrow{pq})$ is independent of the directed line segment pq representing $\overrightarrow{pq}$. We thus have a pair of vector operations, and in terms of them the set V of all geometric vectors is a vector space, each of the laws A1-S3 being a theorem in plane geometry.

If we choose a special point O (the origin) in the plane then it is permissible to identify a geometric vector with that directed line segment emanating from O which represents it. Moreover, with the origin O thus fixed as the universal initial point, these segments are in turn uniquely specified by their terminal points. Thus the choice of O allows us to pair any point p with the geometric vector $\overrightarrow{Op}$ and the space of all geometric vectors corresponds to the plane itself. Addition is now given by the parallelogram rule : p+q = r if and only if r is the fourth vertex of the parallelogram having Op and Oq as two of its sides. However, this nice formulation of addition doesn't hold when O , p and q are collinear !

The vector space V of plane geometric vectors actually turns out to be isomorphic to $\mathbb{R}^2$, and we usually try to minimize the amount of unsatisfactory plane geometry that we have to do by establishing this isomorphism as early in the game as possible. This process is essentially the introduction of axes and coordinate systems by which we start analytical geometry. The beginning facts are still established in a generally unrigorous fashion. They are :

(1) The choice of an ordered pair of perpendicular axes and a unit segment determines a unique one-to-one coordinate correspondence between the Euclidean plane and $\mathbb{R}^2$;

(2) If we use the notation that $\underline{x} = \langle x_1, x_2 \rangle$ is the coordinate pair of the point x in the Euclidean plane then the directed line segments xy and zw are equivalent if and only if $\underline{y} - \underline{x} = \underline{w} - \underline{z}$ in the vector space $\mathbb{R}^2$.

(3) Three distinct points x, y and z are collinear if and only if $\underline{z} - \underline{x} = c(\underline{y} - \underline{x})$ for some real number c, in which case the segment xz is $|c|$ times as long as the segment xy, and x is between y and z if and only if and only if c is negative.

Assuming these facts the reader will see that $\overrightarrow{xy} \longrightarrow \underline{y} - \underline{x}$ is a one-to-one correspondence θ from V to $\mathbb{R}^2$, that $\theta(\overrightarrow{xy} + \overrightarrow{yz}) = \theta(\overrightarrow{xy}) + \theta(\overrightarrow{yz})$ and that $\theta(c\,\overrightarrow{xy}) = c\,\theta(\overrightarrow{xy})$. Thus θ is an isomorphism.

However, this conclusion assumes that we have proved that V is a vector space and therefore needs all of the dubious geometry that we sketched earlier. There is another way of proceeding which requires only that we establish (1) - (3) above. Then <u>without any further geometric arguments</u> we can conclude that the equivalence of directed line segments is indeed an equivalence relation so that the set V is defined, that our definition of the operations on V are meaningful and that $\psi = \theta^{-1}$ preserves operations (from $\mathbb{R}^2$ to V).

But now we can apply the following general and very useful lemma.

LEMMA 6. Let W be a vector space, let V be a set having two vector-like operations and let $T : W \rightarrow V$ be a bijection preserving the operations, i.e. satisfying $T(c\alpha + d\beta) = c\,T(\alpha) + dT(\beta)$. Then V is a vector space.

<u>Proof:</u> It follows from the fact that T preserves operations that T(0) is the zero for V, that $T(-\alpha)$ is the negative of $T(\alpha)$ and that the universally quantified laws all hold. The reader will be asked to make some of these calculations in the exercises.

§ 3. Product spaces, Hom(V, W) and quotient spaces.

A. Product spaces. If W is a real vector space and A is an arbitrary set then the set $V = W^A$ of all W-valued functions on A is a real vector space in exactly the same way that $\mathbb{R}^A$ is. Addition is the natural addition of functions: $(f+g)(a) = f(a) + g(a)$, and, similarly, $(xf)(a) = a(f(a))$ for every $f \in V$ and $x \in \mathbb{R}$. The laws A1-S4 follow just as before and for exactly the same reasons. For variety let us check the associative law for addition. The equation $f+(g+h) = (f+g)+h$ means that $(f+(g+h))(a) = ((f+g)+h)(a)$ for all $a \in A$. But

$(f+(g+h))(a) = f(a) + (g+h)(a) = f(a) + (g(a)+h(a)) = (f(a)+g(a))+h(a)$

$= (f+g)(a) + h(a) = ((f+g) + h)(a)$, where the middle equality in this chain of five holds by the associative law for W and the other four are all applications of the definition of addition. Thus the associative law for addition holds in W^A. The analogue of Cartesian n-space here is the set $W^{\bar{n}}$ of all n-tuples $\underline{\alpha} = \langle \alpha_1, \ldots, \alpha_n \rangle$ of elements in W; it is also designated W^n. By analogy with Cartesian n-space we call α_j the j^{th} coordinate of the n-tuple $\underline{\alpha}$. Generalizing again, we can call $f_a = f(a)$ the a^{th} coordinate of the function $f \in V = W^A$.

There is no reason at all why the same space W has to be used at each index as we did above. In fact, if $W_1, \ldots, W_n$ are any n vector spaces then the set of all n-tuples $\underline{\alpha} = \langle \alpha_1, \ldots, \alpha_n \rangle$ such that $\alpha_j \in W_j$ for $j = 1, \ldots, n$ is a vector space under the same definitions of the operations and for the same reasons. That is, the Cartesian product

$V = W_1 \times W_2 \times \dots \times W_n$ is again a vector-space of vector-valued functions.

Much later in the course we shall need the corresponding fact about a general Cartesian product of vector spaces. We remind the reader that if $\{W_i : i \in I\}$ is any indexed collection of vector spaces then their Cartesian product, $\Pi_{i \in I} W_i$, is defined as the set of all functions f with domain I such that $f(i) \in W_i$ for all $i \in I$ (see § 0.8a) .

THEOREM 1. The Cartesian product of a collection of vector spaces is itself a vector space under the natural operations of addition of two functions and multiplication of a function by a scalar (real number) .

Proof: The proofs of A1-S4 that we sampled earlier hold verbatim. They did not require that the functions being added have all their values in the same space, but only that the values at a given domain element i lie in the same space.

Notice that the function spaces $\mathbb{R}^A$, W^A, $\mathbb{R}^{\bar{n}}$, $W^{\bar{n}}$ and $\Pi_1^n W_i$ are all special cases of the function space $\Pi_{i \in I} W_i$.

These notions should be easily grasped conceptually, since they involve the same old idea that the natural addition of functions leads to a vector space of functions. The only difficulty is that the more general Cartesian product notions are hard to illustrate in a simple way apart from the eventual uses that we shall make of them. With V and W as vector spaces the function space W^V has the important subspace $\mathrm{Hom}(V, W)$ of all linear mappings from V to W , which we shall look at next, and the subspace of all differentiable mappings that we shall study in Chapter 3. Our use of $\prod_1^n W_i$ will be principally when the W_i are all subspaces of a given space W in our study of direct sums in § 4. The general Cartesian product vector space is needed when later on we consider the collection $\{V_p : p \in M\}$ of all tangent spaces V_p at the points p of a differentiable manifold M. Then a vector field on M is simply a function f in $\prod_{p \in M} V_p$ and the vector space of all continuous vector fields is a subspace of this Cartesian product.

B. Hom(V, W). Linear transformations have the very simple but important properties that the sum of two linear transformations is linear and the composition of two linear transformations is linear. These imprecise statements need bolstering by conditions on domains and codomains, but in essence they are the theme of this section. The proofs are simple formal algebraic arguments, but the objects being discussed will escalate in conceptual complexity.

If W is a vector space and A is any set, we know that the space $V = W^A$ of all functions $f : A \longrightarrow W$ is a vector space of functions (now vector-valued) in the same way that $\mathbb{R}^A$ is. If A is itself a vector space V, we naturally single out for special study the subset of W^V consisting of all linear mappings. We designate this subset Hom(V, W). The following elementary theorem summarizes its basic algebraic properties.

THEOREM 2. Hom(V, W) is a vector subspace of W^V. If $T \in \mathrm{Hom}(V_1, V_2)$ and $D \in \mathrm{Hom}(V_2, V_3)$ then $S \circ T \in \mathrm{Hom}(V_1, V_3)$. Moreover, composition is distributive over addition ; $(S_1 + S_2) \circ T = (S_1 \circ T) + (S_2 \circ T)$ and $S \circ (T_1 + T_2) = (S \circ T_1) + (S \circ T_2)$ under the obvious hypotheses on domains and codomains. Finally, composition commutes with scalar multiplication ; $c(S \circ T) = (cS) \circ T = S \circ (cT$

Proof: The whole theorem is an easy formality, but we shall write down two of the arguments as examples of how these things go. If $S, T \in \mathrm{Hom}(V, W)$ then $(S+T)(c_1\alpha_1 + c_2\alpha_2) = S(c_1\alpha_1 + c_2\alpha_2) + T(c_1\alpha_1 + c_2\alpha_2) = c_1S(\alpha_1) + c_2S(\alpha_2) + c_1T(\alpha_1) + c_2T(\alpha_2) = c_1(S+T)(\alpha_1) + c_2(S+T)(\alpha_2)$, so that $S+T$ is linear and Hom (V, W) is closed under addition. The reader should be sure he knows what hypothesis is being used at each step of the above continued equality. The closure of Hom(V, W) under scalar multiplication follows similarly. Thus Hom(V, W) is a subspace of W . Next, if S and T are linear we have $S \circ T(c_1\alpha_1 + c_2\alpha_2) = (S(T(c_1\alpha_1 + c_2\alpha_2)) = S(c_1T(\alpha_1) + c_2T(\alpha_2)) = c_1S(T(\alpha_1)) + c_2S(T(\alpha_2)) = c_1S \circ T(\alpha_1) + c_2S \circ T(\alpha_2)$, so that $S \circ T$ is linear. The two distributive laws will be left to the reader.

COROLLARY . If $T \in \mathrm{Hom}(V_1, V_2)$ then composition on the right by T is a linear transformation of $\mathrm{Hom}(V_2, V_3)$ into $\mathrm{Hom}(V_1, V_3)$. It is an isomorphism if T is an isomorphism.

Proof: The algebraic properties of composition stated in the theorem can be combined as follows :

$$(c_1S_1 + c_2S_2) \circ T = c_1(S_1 \circ T) + c_2(S_2 \circ T)$$

$$S \circ (c_1T_1 + c_2T_2) = c_1(S \circ T_1) + c_2(S \circ T_2) \quad .$$

The first equation says exactly that composition on the right by a fixed T is a linear transformation. If T is an isomorphism then composition by T^{-1} "undoes" composition by T and so is its inverse.

The second equation implies a similar corollary about composition on the left by a fixed S.

Now consider in particular the space $\mathrm{Hom}(V, V)$, which we may as well designate $\mathrm{Hom}(V)$. In addition to being a vector space it is also closed under composition, which we consider a multiplication operation. Since composition of functions is always associative (see 0.10) we thus have for multiplication the laws $A \circ (B \circ C) = (A \circ B) \circ C$, $A \circ (B + C) = (A \circ B) + (A \circ C)$, $(A + B) \circ C = (A \circ C) + (B \circ C)$ and $k(A \circ B) = (kA) \circ B = A \circ (kB)$. Any vector space which has in addition to the vector operations an operation of multiplication related to the vector operations in the above ways is called an algebra. Thus :

THEOREM 3. $\mathrm{Hom}(V)$ is an algebra.

We noticed earlier that certain real-valued function spaces were also algebras. $\mathbb{R}^A$ and $([0,1])$ were examples. In those cases multiplication was commutative, but in the case of $\mathrm{Hom}(V)$ multiplication is not commutative unless V is a trivial space ($V = \{0\}$ or V is isomorphic to $\mathbb{R}$) . We shall check this later on when we examine the finite dimensional theory in greater detail.

C. Product projections and injections. There are two related classes of simple linear mappings that are of basic importance in handling a Cartesian product space $\prod_{k \in K} W_k$. The first is the mapping $\pi_j : f \longrightarrow f(j)$ assigning to each element of $\prod_k W_k$ its value at j . Thus π_2 on $W_1 \times W_2 \times W_3$ is the mapping $< \alpha_1, \alpha_2, \alpha_3 > \longrightarrow \alpha_2$. We call f(j) the j^{th} coordinate of the "point" f and π_j the j^{th} coordinate mapping. The second is the mapping θ_j taking each vector $\alpha \in W_j$ into the function $f \in \prod_{k \in K} W_k$ defined by $f(j) = \alpha$ and $f(i) = 0$ for $i \neq j$. For example, θ_2 for $W_1 \times W_2 \times W_3$ is the mapping $\alpha \to < 0, \alpha, 0 >$ where $\alpha \in W_2$. We call θ_j the injection of W_j into $\prod_k W_k$. The linearity of π_j and θ_j follows directly from the definition of the vector operations for functions. The mappings π_j and θ_j are clearly connected, and the following projection-injection identities state their exact relationship :

$$\pi_j \circ \theta_j = I_j \text{ , and } \pi_j \circ \theta_i = 0_j \text{ if } i \neq j \text{ ;}$$

If K is finite then

$$\sum_{k \in K} \theta_k \circ \pi_k = I \text{ .}$$

Here, of course, I_j and 0_j are the identity and zero transformations on W_j and I is the identity on $\prod_k W_k$. In the case $\prod_{i=1}^{3} W_i$ we have

$\theta_2 \circ \pi_2(<\alpha_1, \alpha_2, \alpha_3>) = <0, \alpha_2, 0>$, and the identity simply says that $<\alpha_1, 0, 0> + <0, \alpha_2, 0> + <0, 0, \alpha_3> = <\alpha_1, \alpha_2, \alpha_3>$ for all $\alpha_1, \alpha_2, \alpha_3$. The coordinate projections π_j are useful in studying any product space, but because of the limitation in the above identity the injections θ_j are of interest principally in the case of finite products. The following theorem states a characteristic property of product spaces.

THEOREM 4. If $T_i \in \mathrm{Hom}(V, W_i)$ for each $i \in I$ then there exists a unique $T \in \mathrm{Hom}(V, \Pi_{i \in I} W_i)$ such that $T_i = \pi_i \circ T$ for all $i \in I$.

Proof: This theorem is true for any product space but we prove it only for a finite product, as an exercise in applying the above injection-projection identities. If T exists such that $T_i = \pi_i \circ T$ for each i , then

$$T = I \circ T = \sum(\theta_i \pi_i) \circ T = \sum \theta_i \circ (\pi_i \circ T) = \sum \theta_i \circ T_i \quad .$$

Thus T is uniquely determined as the map $\sum \theta_i \circ T_i$. Moreover, this T does have the required property, since

$$\pi_j \circ T = \pi_j \circ (\textstyle\sum_i \theta_i \circ T_i) = \sum_i (\pi_j \circ \theta_i) \circ T_i = I_j \circ T_j = T_j \quad .$$

There is a similar theorem in the other direction whose proof will be left to the reader.

THEOREM 5. If $T_i \in \mathrm{Hom}(W_i, V)$ for each i in a finite index set I , then there exists a unique $T \in \mathrm{Hom}(\Pi_i W_i, V)$ such that $T \circ \theta_j = T_j$ for each $j \in I$.

D. Quotient spaces ; affine subspaces. If N is a subspace of a vector space V and $\alpha \in V$ then the set $N + \alpha = \{ \xi + \alpha : \xi \in N \}$ is called either the coset of N containing α or the affine subspace of V through α and parallel to N . If N is given and fixed we shall use the notation $[\alpha] = N + \alpha$ (see 0.11).

THEOREM 6. The relation $\alpha \sim \beta \iff \beta - \alpha \in N$ is an equivalence relation and its fibers are the cosets of N . The vector operations on V lift to the space of fibers and make it a vector space, called the quotient space of V by N and designated V/N . The projection mapping π is a linear mapping of V onto V/N and the nullspace of π is N .

Proof: Checking that $\alpha \sim \beta$ is an equivalence relation is an easy mental exercise. Then, since $\beta - \alpha \in N \iff \beta \in N + \alpha = [\alpha]$, we see that $\alpha \sim \beta \iff \beta \in [\alpha]$ so that the fiber $\pi(\alpha)$ is the coset $[\alpha]$. The space of fibers is thus the set of all cosets of N .

We shall check that the operations of V lift to V/N by direct argument. First, if $\alpha_1 - \alpha_2 \in N$ and $\beta_1 - \beta_2 \in N$ then $(\alpha_1 + \beta_1) - (\alpha_2 + \beta_2) \in N$, since N is closed under addition. That is, $\alpha_1 \sim \alpha$ and $\beta_1 \sim \beta \Rightarrow \alpha_1 + \beta_1 \sim \alpha + \beta$ (see § 0.11) . This implies that the set sum

$\pi(\alpha_1) + \pi(\beta_1)$ is a subset of the coset $\pi(\alpha_1+\beta_1)$. Therefore, if we define $\boxplus : V/N \times V/N \to V/N$ by taking $A \boxplus B$ to be the unique coset including the set sum $A+B$, we have

$$\pi(\alpha) \boxplus \pi(\beta) = \pi(\alpha+\beta) .$$

Similarly, $\alpha_1 \sim \alpha \Longrightarrow c\alpha_1 \sim c\alpha$ for any scalar c, so that the set product $c\pi(\alpha_1)$ is a subset of the coset $\pi(c\alpha_1)$. If we define $\boxdot$ from $\mathbb{R} \times V/N$ to V/N by taking $c \boxdot A$ to be the unique coset including the set product cA, we have

$$c \boxdot \pi(\alpha) = \pi(c\alpha) .$$

Together, the two displayed equations above say that the projection mapping π preserves the vector operations, and it follows from Lemma 6 that V/N is a vector space. The zero vector of V/N is the coset N, so that $\pi(\alpha) = 0 \Longleftrightarrow \alpha \in N$. This completes the proof of the theorem.

Consider now the collection $\mathcal{Q}$ of all affine subspaces of a vector space V; $\mathcal{Q}$ is thus the set of all cosets of all vector subspaces of V. The properties we are interested in are very elementary and will be simply listed, with at most a hint at a proof.

(1) The intersection of any family of affine subspaces is either empty or is itself an affine subspace. In fact, if $\{A_i\}_{i \in I}$ is an indexed collection of affine subspaces and if A_i is a coset of the vector subspace W_i for each $i \in I$, then $\bigcap_{i \in I} A_i$ is either empty or a coset of the

vector subspace $\bigcap_{i\in I} W_i$.

To see this it is crucial to remember the different descriptions of a single affine subspace : if $A = \alpha + W$ where W is a vector subspace, then also $A = \beta + W$ for every $\beta \in A$. Therefore $\beta \in \bigcap_{i\in I} A_i$ implies $A_i = \beta + W_i$ for all i and $\bigcap A_i = \beta + \bigcap W_i$.

(2) If $A, B \in \mathcal{A}$ then $A + B \in \mathcal{A}$. That is, the set sum of any two affine subspaces is itself an affine subspace.

(3) $A \in \mathcal{A}$ and $T \in \mathrm{Hom}(V, V_1) \Longrightarrow T[A]$ is an affine subspace of V_1 .

(4) If B is an affine subspace of V_1 and $T \in \mathrm{Hom}(V, V_1)$ then $T^{-1}[B]$ is either empty or an affine subspace of V .

(5) For a fixed $\alpha \in V$ the <u>translation of</u> V <u>through</u> α is the mapping $S_\alpha : V \longrightarrow V$ defined by $S_\alpha(\xi) = \xi + \alpha$ for all $\xi \in V$. Translation is <u>not</u> linear ; for example, $S_\alpha(0) = \alpha$. It is clear, however, that translation carries affine subspaces into affine subspaces. Thus $S_\alpha(A) = A + \alpha$ and $S_\alpha(\beta + W) = (\alpha + \beta) + W$.

(6) <u>An affine transformation</u> of a vector space V into a vector space W is a linear mapping of V into W followed by a translation in W . Thus an affine transformation is of the form $\xi \longmapsto T(\xi) + \beta$, where $T \in \mathrm{Hom}(V, W)$ and $\beta \in W$. Notice that $\xi \longmapsto T(\xi + \alpha)$ is affine since $T(\xi + \alpha) = T(\xi) + \beta$, where $\beta = T(\alpha)$.

It follows from (3) and (5) that an affine transformation carries affine subspaces of V into affine subspaces of W .

(7) It is obvious that $S_\beta \circ S_\alpha = S_{(\beta+\alpha)}$, and we shall interpret composition of translations as an addition operation. Next, if nS_α is taken to mean $S_\alpha \circ S_\alpha \circ \ldots \circ S_\alpha$ to n composition summands, than $nS_\alpha = S_{n\alpha}$. Also, $(1/m)S_\alpha = S_{\alpha/m}$ in the sense that $S_{\alpha/m} \circ S_{\alpha/m} \circ \ldots S_{\alpha/m}$ to m summands equals S_α. Thus $xS_\alpha = S_{x\alpha}$ for x rational, and we simply define xS_α to be $S_{x\alpha}$ for x irrational.

But now we have two vector-like operations on the set $\mathcal{S}$ of all translations of V, and, since $S_{(x\alpha + y\beta)} = xS_\alpha \circ yS_\beta$, the mapping $\alpha \longmapsto S_\alpha$ of V onto $\mathcal{S}$ preserves the operations. It follows from Lemma 6 (first homework assignment) that $\mathcal{S}$ is a vector space. The zero of $\mathcal{S}$ is S_0, which is just the identity mapping I_V of V onto itself. Since S_α is not S_0 unless $\alpha = 0$, the mapping $\alpha \longmapsto S_\alpha$ is an isomorphism between the vector space V and the set of all translations of V. This has taken some time to describe but it is really pretty obvious.

§ 4. Direct sums.

A. Direct sums. LEMMA 7. If $V_1, \ldots, V_n$ are subspaces of the vector space V then the mapping $<\alpha_1, \ldots, \alpha_n> \longmapsto \Sigma_1^n \alpha_i$ is a linear transformation from $\Pi_1^n V_i$ to V.

Proof: If π is the mapping and $\underline{\alpha}$ is the n-tuple $<\alpha_1, \ldots, \alpha_n>$ then $\pi(\underline{\alpha}) + \pi(\underline{\beta}) = (\Sigma_1^n \alpha_i) + (\Sigma_1^n \beta_i)$. Regrouping by pairs this sum becomes $\Sigma_1^n(\alpha_i + \beta_i) = \pi(\underline{\alpha} + \underline{\beta})$. Thus π is additive. Also $x\pi(\underline{\alpha}) = x\Sigma_1^n \alpha_i = \Sigma_1^n x\alpha_{ii} = \pi(x\underline{\alpha})$, by S3 and induction. Thus π is linear.

DEFINITION. We shall say that the V_i's are independent if π is injective and that V is the direct sum of the V_i's if π is an isomorphism. We express the latter relationship by writing
$V = V_1 \oplus \ldots \oplus V_n = \bigoplus_1^n V_i$.

Thus $V = \bigoplus_{i=1}^n V_i$ if and only if π is injective and surjective, i.e., if and only if the subspaces $\{V_i\}_1^n$ are both independent and span V. A useful restatement is that each $\alpha \in V$ is uniquely expressible as a sum $\Sigma_1^n \alpha_i$ with $\alpha_i \in V_i$ for all i; α has some such expression because the V_i's span V, and the expression is unique by their independence.

For example, let $V = \mathcal{C}^2([-1,1])$, the space of functions on $[-1,1]$ having continuous second derivatives, let V_o be the space of constant functions, V_1 the space of linear functions of the form $f(t) = ct$, and R the space of functions $f \in V$ such that $f(0) = f'(0) = 0$. The two term Taylor expansion of f gives the unique representation $f(t) = a+bt + r(t)$, where $r(t) \in R$, and of course $a = f(0)$, $b = f'(0)$. That is $V = V_o \oplus V_1 \oplus R$.

On the other hand, if M is the subspace of polynomials of degree $\leqq 2$, then $V = M + R$ (why ?) but the sum is not direct (why ?). Since π is injective if and only if its nullspace is $\{0\}$ (Lemma 5, § 1.1D) we have :

LEMMA 8. The independence of the subspaces $\{V_i\}_1^n$ is equivalent to the property that if $\Sigma_1^n \alpha_i = 0$ and $\alpha_i \in V_i$ for all i then $\alpha_i = 0$ for all i.

The case of two subspaces is particularly simple.

LEMMA 9. The subspaces V_1 and V_2 of V are independent if and only if $V_1 \cap V_2 = \{0\}$.

<u>Proof:</u> If $\alpha_1 + \alpha_2 = 0$ then $\alpha_1 = -\alpha_2 \in V_1 \cap V_2$. Thus $\alpha_1 + \alpha_2 = 0$ will imply $\alpha_1 = \alpha_2 = 0$ if and only if $V_1 \cap V_2 = \{0\}$.

COROLLARY. $V = V_1 \oplus V_2$ if and only if $V = V_1 + V_2$ and $V_1 \cap V_2 = \{0\}$.

DEFINITION. If $V = V_1 \oplus V_2$ then V_1 and V_2 are called <u>complementary</u> subspaces, and each is a <u>complement</u> of the other.

Warning : A subspace M of V does not have a unique complementary subspace unless M is trivial (i.e., $M = \{0\}$ or $M=V$).

If we view $\mathbb{R}^3$ as coordinatized Euclidean 3-space then M is a proper subspace if and only if M is a plane containing the origin or M is a line through the origin. If M and N are proper subspaces one of which is a plane and the other a line not lying in that plane then M and N are complementary subspaces. Moreover these are the only non-trivial complementary pairs in $\mathbb{R}^3$. The reader will be asked to prove some of these facts in the exercises and they all will be clear by the end of § 6.

The following lemma is technically useful.

LEMMA 10. If V_1 and V_o are independent subspaces of V and $\{V_i\}_2^n$ are independent subspaces of V_o then $\{V_i\}_1^n$ are independent subspaces in V.

Proof: If $\alpha_i \in V_i$ for all i and $\Sigma_1^n \alpha_i = 0$ then setting $\alpha_o = \Sigma_2^n \alpha_i$ we have $\alpha_1 + \alpha_o = 0$ with $\alpha_o \in V_o$. Therefore $\alpha_1 = \alpha_o = 0$ by the independence of V_1 and V_o. But then $\alpha_2 = \alpha_3 = \ldots = \alpha_n = 0$, by the independence of $\{V_i\}_2^n$, and we are done (Lemma 8).

COROLLARY. $V = V_1 \oplus V_o$ and $V_o = \bigoplus_{i=2}^n V_i$ together imply that $V = \bigoplus_{i=1}^n V_i$.

B. Projections. If $V = \bigoplus_{i=1}^n V_i$, if π is the isomorphism $<\alpha_1, \ldots, \alpha_n> \longmapsto \alpha = \Sigma_1^n \alpha_i$ and if π_j is the projection map $<\alpha_1, \ldots, \alpha_n> \longmapsto \alpha_j$ of $\Pi_{i=1}^n V_i$ to V_j, then $(\pi_j \circ \pi^{-1})(\alpha) = \alpha_j$.

DEFINITION. We call α_j the j^{th} component of α and we call the linear map $P_j = \pi_j \circ \pi^{-1}$ the projection of V onto V_j (with respect to the given direct sum decomposition of V).

This use of the word "projection" is different from its use in the Cartesian product situation, and each is different from its use in 0.11 in the quotient space context. It will become apparent that these three uses are related, and the ambiguity causes little confusion since the proper meaning is always clear from the context.

THEOREM 7. The projections P_i are such that range $P_i = V_i$, $P_i \circ P_j = 0$ for $i \neq j$, and $\Sigma_1^n P_i = I$.

Proof. If β is any element of V_j then β is the sum of itself and the 0 vectors in all the other subspaces V_i. That is, $P_j(\beta) = \beta$ and $P_i(\beta) = 0$ if $i \neq j$. Thus range $(P_j) = V_j$ and $P_i \circ P_j = 0$. Finally, $\Sigma_1^n P_i(\alpha) = \Sigma_1^n \alpha_i = \alpha = I(\alpha)$, so that $\Sigma_1^n P_i = I$, and we are done.

The above projection properties are clearly the reflection in V of the projection-injection identities for $\Pi_1^n V_i$.

A converse theorem is also true.

THEOREM 8. If $\{P_i\}_1^n \subset \mathrm{Hom}\, V$ satisfy $\Sigma_1^n P_i = I$ and $P_i \circ P_j = 0$ for $i \neq j$ then $V = \bigoplus_{i=1}^n V_i$, where V_i is the range of P_i, and P_i is the corresponding projection on V_i.

Proof: The equation $\alpha = \Sigma_1^n P_i(\alpha)$ where $P_i(\alpha) \in V_i$ shows that $\{V_i\}_1^n$ span V. Next, for any $\alpha_j \in V_j$, $P_i(\alpha_j) = 0$ for $i \neq j$

(since $\alpha_j = P_j(\alpha)$ for some α , and so $P_i(\alpha_j) = P_i \circ P_j(\alpha) = 0$) . Also, $P_j(\alpha_j) = (I - \Sigma_{i \neq j} P_i)(\alpha_j) = I(\alpha_j) = \alpha_j$. Now consider $\alpha = \Sigma_1^n \alpha_i$ for any choice of $\alpha_i \in V_i$. Using the above results, we have $P_j(\alpha) = \Sigma_1^n P_j(\alpha_i) = \alpha_j$. Therefore $\alpha = 0$ implies $\alpha_j = 0$ for all j . That is, the subspaces V_i are independent. Therefore $V = \bigoplus_1^n V_i$. Finally, the equation $\alpha = \Sigma P_i(\alpha)$ with $P_i(\alpha) \in V_i$ shows that $P_i(\alpha)$ is the projection of α onto V_i .

COROLLARY . The projections P_i are idempotent ($P_i^2 = P_i$) , or, equivalently, each is the identity on its range.

<u>Proof:</u> $P_j^2 = P_j \circ (I - \Sigma_{i \neq j} P_i) = P_j \circ I = P_j$.

Since this can be rewritten $P_j(P_j(\alpha)) = P_j(\alpha)$ for every α in V , it says exactly that P_j is the identity on its range.

Again we have a converse.

THEOREM 9. If $P \in \mathrm{Hom}(V)$ is idempotent then V is the direct sum of its range and nullspace and P is the corresponding projection on its range.

<u>Proof:</u> Setting $Q = I - P$, we have $PQ = P - P^2 = 0$. Therefore V is the direct sum of the ranges of P and Q and P is the corresponding projection on its range by the above theorem. We therefore only need to know that the range of Q is the nullspace of P . But $P(\alpha) = 0$ if and only if $Q(\alpha) = \alpha$ (since $P+Q = I$), and since Q

is the identity on its range ($Q^2 = Q$) this holds if and only if $\alpha \in$ range Q. That is, $R(P) = R(Q)$, q.e.d.

If $V = M \oplus N$ and P is the corresponding projection on M we call P the projection on M along N. P is not determined by M alone, since M does not determine N.

C. Polynomials. The material in this and the next subsection will be used in our study of differential equations with constant coefficients and in the proof of the diagonalizability of a symmetric matrix. In linear algebra it is basic in almost any approach to the canonical forms of matrices.

If $p_1(t) = \Sigma_o^m a_i t^i$ and $p_2(t) = \Sigma_o^n b_j t^j$ are any two polynomials then $p(t) = p_1(t)\, p_2(t) = \Sigma_o^{m+n} c_k t^k$, where $c_k = \Sigma_{i+j=k}\, a_i b_j = \Sigma_{i=0}^{k}\, a_i b_{k-i}$. Now let T be any fixed element of $\mathrm{Hom}(V)$ and for any polynomial $q(t)$ let $q(T)$ be the transformation obtained by replacing t by T, t^2 by $T \circ T$, etc., in $q(t)$. Then the bilinearity of composition (Theorem 2) shows that $p_1(T) \circ p_2(T)$ has the same expansion as the polynomial $p_1(t)\, p_2(t)$, so that if $p(t) = p_1(t) p_2(t)$ then $p(T) = p_1(T) \circ p_2(T)$. In particular any two polynomials in T commute with each other under composition. Much more simply, the commutative law for addition implies that if

$$p(t) = p_1(t) + p_2(t) \quad \text{then} \quad p(T) = p_1(T) + p_2(T) .$$

The mapping $p(t) \longrightarrow p(T)$ from the algebra of polynomials into the algebra $\mathrm{Hom}(V)$ thus preserves addition, composition and (obviously) scalar multiplication. That is, it preserves all the operations of an algebra and is therefore what is called an (algebra) homomorphism.

The word "homomorphism" is a general term describing a mapping θ between two algebraic systems of the same kind such that θ preserves the operations of the system. Thus a homomorphism between vector spaces is simply a linear transformation and a homomorphism between groups is a mapping preserving the one group operation. An accessible, but not really typical, example is the logarithm function, which is a homomorphism from the multiplicative group of positive real numbers to the additive group of $\mathbb{R}$. The logarithm function is actually a bijective homomorphism and is therefore a group <u>isomorphism</u>.

If this were a course in algebra we would show that the division algorithm and the properties of the degree of a polynomial imply the following theorem.

THEOREM. If $p_1(t)$ and $p_2(t)$ are relatively prime polynomials then there exist polynomials $a_1(t)$ and $a_2(t)$ such that

$$a_1(t)\, p_1(t) + a_2(t)\, p_2(t) = 1 .$$

cf. $p|q = 1 \Rightarrow \exists\, a,b \ni ap+bq = 1$

Being relatively prime means having no common factors except constants. We shall assume this theorem and the results of the discussion preceding it in proving our next theorem.

D. The null space of $p(T)$.

DEFINITION . A subspace $M \subset V$ is invariant under $T \in \text{Hom}(V)$ if and only if $T[M] \subset M$ (i.e., $T \restriction M \in \text{Hom}(M)$) .

THEOREM 10. Given any $T \in \text{Hom } V$ and any polynomial p , the nullspace N of $p(T)$ is invariant under T . If $p = p_1 p_2$ is any factorization of p into relating prime factors and if N_1 and N_2 are the nullspace of $p_1(T)$ and $p_2(T)$ respectively, then $N = N_1 \oplus N_2$. Also, N_1 is the range of $p_2(T) \restriction N$ and N_2 is the range of $p_1(T) \restriction N$.

Proof: Since $T \circ p(T) = p(T) \circ T$ we have that

$$p(T)[T[N]] = T[p(T)[N]] = T(0) = 0$$

and so $T[N] \subset N$. Notice also that since $p(T) = p_1(T) \circ p_2(T)$ it follows that any α in N_2 is also in N , so that $N_2 \subset N$. Similarly , $N_1 \subset N$. We can therefore replace V by N and T by $T \restriction N$, and so can assume that $T \in \text{Hom } N$ and $p(T) = 0$.

Now choose polynomials a_1 and a_2 so that $a_1 p_1 + a_2 p_2 = 1$. Since $q \longrightarrow q(T)$ is an algebraic homomorphism we then have

$$a_1(T) \circ p_1(T) + a_2(T) \circ p_2(T) = I .$$

If $\alpha \in N_1$ then $\alpha = a_1(T)(p_1(T)(\alpha)) + a_2(T)p_2(T)(\alpha) = p_2(T)(a_2(T)(\alpha)) \in R_2$ $=$ range $p_2(T)$. Thus $N_1 \subset R_2$. But since $p_1(T) \circ p_2(T) = 0$ we also have $p_1(T)[R_2] = 0$ so that $R_2 \subset N_1$. Thus $R_2 = N_1$ and, similarly $R_1 = N_2$.

Suppose next that $\alpha \in N_1 \cap N_2$. Then $\alpha = a_1(T)(p_1(T)(\alpha)) + a_2(T)(p_2(T)(\alpha)) = 0 + 0 = 0$. That is, $N_1 \cap N_2 = \{0\}$. Finally, for any $\alpha \in N$, $\alpha = p_1(T)(a_1(T)(\alpha)) + p_2(T)(a_2(T)(\alpha)) = \alpha_1 + \alpha_2$ where $\alpha_1 \in R_1$ and $\alpha_2 \in R_2$. Thus $R_1 + R_2 = N$. Putting the four equations $R_2 = N_1$, $R_1 = N_2$, $N_1 \cap N_2 = \{0\}$ and $R_1 + R_2 = N$ together we get in particular that $N = N_1 \oplus N_2$, which finishes the proof of the theorem.

An equivalent but more algebraic proof begins by noticing that the maps $P_i = a_i(T) \circ p_i(T)$, $i = 1, 2$, are complementary projections. See the exercises.

E. <u>On solving a linear equation.</u> Many important problems in mathematics are in the following general form. A linear operator $T : V \longrightarrow W$ is given, and, for given $\eta \in W$ the equation $\eta = T(\xi)$ is to be solved for $\xi \in V$. In our terms the condition that there exist a solution is exactly that η be in the range space of T . In special circumstances this condition can be given more or less useful equivalent alternate formulations. Let us suppose that we know how to recognize $R(T)$, in which case we may as well make it the new codomain and so assume that T is surjective. There still remains the problem of what we mean by solving the equation. The universal principle running through all the important instances of the problem is that a solution process calculates a right inverse to T ; that is a linear operator $S : W \longrightarrow V$ such that $T \circ S = I_W$, the identity on W . Thus a solution process picks one solution vector $\xi \in V$ for each $\eta \in W$ in such a way that the solving ξ varies linearly with η . Granting that this is what is meant by solving, we have the following fundamental reformulation.

THEOREM 11. The linear right inverses to surjective T are in 1-1 correspondence with the subspaces of V complementary to $N = N(T)$. If S is a right inverse the corresponding complementary subspace is $R = \text{range}(S)$. Conversely if R is a subspace of V complementary to N then $T \upharpoonright R : R \longrightarrow W$ is an isomorphism, and S is its inverse.

Proof: If S is a right inverse of T and $P = ST$ then $N(P) = N$, since S is injective, and $R(P) = R(S)$ since T is surjective. Moreover, $P^2 = S(TS)T = SI_W T = ST = P$, so that P is a projection. Thus $V = N \oplus R$, where $R = R(P) = R(S)$, by Theorem 9 . If we consider S as being from W to R then of course $S^{-1} = T \upharpoonright R$.

Conversely, if R is a complement of N then $T \upharpoonright R$ is certainly injective, since $R \cap N = \{0\}$. Its surjectivity follows from the surjectivity of T : for every $\beta \in W$ there exists $\alpha \in V$ such that $T(\alpha) = \beta$, and then $\alpha = \eta + \rho$, with $\eta \in N$ and $\rho \in R$, gives $\beta = T(\eta) + T(\rho) = T(\rho)$. Thus $T \upharpoonright R : R \longrightarrow W$ is an isomorphism and its inverse S is a right inverse of T .

§ 5. Bases

A. <u>The linear combination map.</u> For a fixed finite sequence of vectors $\{\alpha_i\}_1^n$ in V the linear combination formula defines a linear map of $\mathbb{R}^n$ to V.

THEOREM 12. For a fixed n-tuple $\underline{\alpha} = < \alpha_1, \ldots, \alpha_n > \in V^n$ the mapping $T_{\underline{\alpha}} : \mathbb{R}^n \longrightarrow V$ defined by $\underline{x} \longmapsto \Sigma_1^n x_i \alpha_i$ is linear.

<u>Proof:</u> The formulas $\Sigma c_i \alpha_i + \Sigma d_i \alpha_i = \Sigma(c_i + d_i)\alpha_i$ and $k \Sigma c_i \alpha_i = \Sigma(kc_i)\alpha_i$ by which we showed in Lemma 3 that the set of all linear combinations of $\{\alpha_i\}$ is a vector space, are precisely the formulas showing the linearity of the mapping $\underline{c} = < c_1, \ldots, c_n > \longmapsto \Sigma_1^n c_i \alpha_i$. After noticing this Lemma 3 becomes a corollary of Lemma 4.

The theorem above was stated for ordered n-tuples but it is equally valid for any finite index set. In fact the equationsfrom Lemma 3 cited above were over an arbitrary finite index set.

In the indexing $i \longmapsto \alpha_i$ the vector α_j corresponds to the index j, but under the linear map $T_{\underline{\alpha}}$ the vector α_j corresponds to the n-tuple δ^j which has the value 1 at j and the value 0 elsewhere, so that $\Sigma_i \delta_i^j \alpha_i = \alpha_j$. This function δ^j is called a <u>Kronecker delta function.</u> It is clearly the characteristic function χ_B of the one-point set $B = \{j\}$.

Notice that the symbol "δ^j" is ambiguous in not specifying the domain just as "χ_B" is ambiguous; the same symbol names different functions depending on the domain S, the only requirement on S being that it be a superset of B. Also, for any index set I the set $\{\delta^j : j \in I\}$ of all Kronecker functions on I is equivalent (duality; § 0.11) to the single function of two variables $\{\delta_i^i\}$ which is the characteristic function of the diagonal in $I \times I$: $\delta_i^j = 1$ if $i = j$ and $\delta_i^j = 0$ if $i \neq j$.

B. Independence and bases.

DEFINITION. The finite indexed set $\{\alpha_i : i \in I\}$ is independent if and only if the above mapping $T_{\underline{\alpha}}$ is injective and $\{\alpha_i\}$ is a basis for V if and only if $T_{\underline{\alpha}}$ is an isomorphism (onto V). In this situation we call $\{\alpha_i : i \in I\}$ an ordered basis if $I = \bar{n}$ for some positive integer n.

Thus $\{\alpha_i : i \in I\}$ is a basis if and only if it is independent and its linear span is V. Or, again, $\{\alpha_i : i \in I\}$ is a basis if and only if for each $\xi \in V$ there exists a unique indexed "coefficient" set $\underline{x} \in \mathbb{R}^I$ such that $\xi = \Sigma x_i \alpha_i$. The numbers x_i always exist because $\{\alpha_i : i \in I\}$ spans V, and $\underline{x}$ is unique because $T_{\underline{\alpha}}$ is one-to-one.

The form of these definitions is dictated by our interpretation of the linear combination formula as the formula of a linear mapping. The more usual definition of independence is now a simple corollary.

LEMMA 11. The independence of the finite indexed set $\{\alpha_i : i \in I\}$ is equivalent to the property that $\Sigma x_i \alpha_i = 0$ only if all the coefficients x_i are 0.

Proof: This is the property that the nullspaces of $T_{\underline{\alpha}}$ consist only of 0 , and is equivalent to the injectivity of $T_{\underline{\alpha}}$, i.e., to the independence of $\{\alpha_i\}$, by Lemma 5.

If $\{\alpha_i\}_1^n$ is an ordered basis for V the unique n-tuple $\underline{x}$ such that $\xi = \Sigma_1^n x_i\alpha_i$ is called the coordinate n-tuple of ξ (with respect to the basis $\{\alpha_i\}$) and x_i is the i^{th} coordinate of ξ. We call $x_i\alpha_i$ (and sometimes x_i) the i^{th} component of ξ. The mapping $T_{\underline{\alpha}}$ will be called a basis isomorphism and its inverse $T_{\underline{\alpha}}^{-1}$, which assigns the unique n-tuple $\underline{x}$ to each vector $\xi \in V$, is a coordinate isomorphism. The linear functional $\xi \longmapsto x_j$ is the j^{th} coordinate functional ; it is the composition of the coordinate isomorphism $\xi \longmapsto \underline{x}$ with the functional $\underline{x} \longrightarrow x_j$ which we called a coordinate functional in the special case of the function space $\mathbb{R}^n$. (But the latter functional can be thought of as obtained from the composition of the identity coordinate isomorphism on $\mathbb{R}^n$ with itself so that the two definitions are consistent.)

Above we took the index set I to be $\bar{n} = \{ 1, \ldots, n \}$ and used the language of n-tuples. The only difference for an arbitrary finite index set is that we speak of a coordinate function $\underline{x} = \{ x_i : i \in I \}$ instead of a coordinate n-tuple.

LEMMA 12. The Kronecker functions $\{ \delta^j \}_{j=1}^n$ form a basis for $\mathbb{R}^n$.

Proof: Since $\Sigma_1^n x_i \delta^i(j) = x_j$ by the definition of δ^i we see that $\Sigma_1^n x_i\delta^i$ is the n-tuple $\underline{x}$ itself so that the mapping $\underline{x} \longrightarrow \Sigma_1^n x_i\delta^i$ is the identity mapping, a trivial isomorphism.

Among all possible indexed bases for $\mathbb{R}^n$ the Kronecker basis is thus singled out by the fact that its coordinate isomorphism is the identity, and for this reason it is called the standard basis or the natural basis of $\mathbb{R}^n$.

THEOREM 13. If $T \in \text{Hom}(V, W)$ is an isomorphism and $\{\alpha_i : i \in I\}$ is a basis for V then $\{T(\alpha_i) : i \in I\}$ is a basis for W.

Proof: If $\theta : \underline{x} \mapsto \Sigma x_i \alpha_i$ is the given basis isomorphism then $T \circ \theta : \mathbb{R}^I \longrightarrow W$ is an isomorphism and $T \circ \theta(\underline{x}) = T(\Sigma x_i \alpha_i) = \Sigma x_i T(\alpha_i)$. Thus, $\underline{x} \mapsto \Sigma x_i T(\alpha_i)$ is an isomorphism ; i.e., $\{T(\alpha_i)\}$ is a basis.

In the original situation we can view the basis $\{\alpha_i\}$ as the image of the standard basis $\{\delta^i\}$ under the basis isomorphism ! In any case, any isomorphism $\theta : \mathbb{R}^I \longrightarrow V$ becomes a basis isomorphism for the basis $\alpha_j = \theta(\delta^j)$. Conversely, we notice that if $\varphi : \mathbb{R}^I \longrightarrow V$ and $\theta : \mathbb{R}^I \longrightarrow W$ are basis isomorphisms over the same index set then $\theta \circ \varphi^{-1} : V \longrightarrow W$ is an isomorphism.

C. The existence of a basis. We shall now prove the existence of ordered bases for vector spaces V that are finite dimensional, in the sense of being spanned by a finite set of vectors. The situation for infinite dimensional V will be discussed briefly in E . We build up a basis one vector at a time by using the following lemma.

LEMMA 13. If $\{\beta_i\}_1^n$ is an independent ordered subset of V and

THEOREM 13a. If $\{\alpha_i\}_1^m$ is a basis for V_1 and $\{\alpha_i\}_{m+1}^n$ is a basis for V_2, and if V_1 and V_2 are complementary subspaces of a vector space V then $\{\alpha_i\}_1^n$ is a basis for V. Conversely, if $\{\alpha_i\}_1^n$ is a basis for V and $V_1 = L(\{\alpha_i\}_1^m)$, $V_2 = L(\{\alpha_i\}_{m+1}^n)$ then V_1 and V_2 are complementary.

Proof: It is clear that $\{\alpha_i\}_1^n$ spans V since its span includes both V_1 and V_2 and so $V_1 + V_2 = V$. Suppose, then, that $\Sigma_1^n x_i\alpha_i = 0$. Setting $\xi_1 = \Sigma_1^m x_i\alpha_i$ and $\xi_2 = \Sigma_{m+1}^n x_i\alpha_i$ it follows that $\xi_1 + \xi_2 = 0$ and $\xi_i \in V_i$, $i = 1,2$. But then $\xi_1 = \xi_2 = 0$ since V_1 and V_2 are complementary. Finally, $x_i = 0$ for $i = 1, \ldots, m$ because $\{\alpha_i\}_1^m$ is independent, and $x_i = 0$ for $i = m+1, \ldots, n$ because $\{\alpha_i\}_{m+1}^n$ is independent. Therefore $\{\alpha_i\}_1^n$ is a basis for V. We leave the converse argument as an exercise.

COROLLARY. If $V = \bigoplus_1^n V_i$ and B_i is a (self-indexed) basis for V_i then $B = \bigcup_1^n B_i$ is a basis for V.

Proof: If we count off $B_i \cup B_2$ starting with all the elements of B_1 we see from the theorem that $B_1 \cup B_2$ is a basis for $V_1 \oplus V_2$. Proceeding inductively we see that $\bigcup_{i=1}^j B_i$ is a basis for $\bigoplus_{i=1}^j V_i$ for $j = 2, \ldots, n$ and the corollary is the case $j = n$.

if $L(\{\beta_i\}) \neq V$ then $\{\beta_i\}_1^{n+1}$ is independent for any choice of β_{n+1} from $V - L(\{\beta_i\}_1^n)$.

<u>Proof:</u> Suppose that $\Sigma_1^{n+1} c_i\beta_i = 0$, with β_{n+1} chosen as above. Then $c_{n+1} = 0$, for otherwise this equation can be solved for β_{n+1}, $\beta_{n+1} = -\Sigma_1^n (c_i/c_{n+1})\beta_i$, contradicting the fact that β_{n+1} was chosen outside of the linear span of $\beta_1, \ldots, \beta_n$. But with $c_{n+1} = 0$ we have $\Sigma_1^n c_i\beta_i = 0$ and therefore all c_i are zero by the independence of $\{\beta_i\}_1^n$. Therefore $\{\beta_i\}_1^{n+1}$ is independent.

THEOREM 14. Any finite spanning set $\{\alpha_i\}_1^m$ includes a basis.

<u>Proof:</u> We can suppose that the α_i are all non-zero. We define a subsequence inductively by running through the α_i's and at each step choosing the first α_j which is independent of those already chosen. Thus $i_1 = 1$, and if $L(\{\alpha_{i_1}, \ldots, \alpha_{i_j}\}) \neq V$, then i_{j+1} is the smallest i greater than i_j such that $\alpha_{i_{j+1}}$ is not in $L(\{\alpha_{i_1}, \ldots, \alpha_{i_j}\})$. If α_{i_n} is the last vector we can choose this way we have a subsequence

$$\beta_j = \alpha_{i_j} \quad \text{for } j = 1, \ldots, n$$

which is independent by repeated applications of the lemma. Also,

$$L(\{\beta_1, \ldots, \beta_j\}) = L(\{\alpha_1, \alpha_2, \ldots, \alpha_{ij}\}),$$

by induction on j. It is true for $j = 1$ since $i_1 = 1$, and if it is true for j then it follows for $j+1$ by the choice of i_{j+1}. In particular

$$L(\{\beta_k\}_1^n) = L(\{\alpha_{i_k}\}_1^n)$$

and since i_{n+1} cannot be chosen this common span must be V. Thus $\{\beta_k\}_1^n$ is an ordered basis for V.

THEOREM 15. If $\{\beta_i\}_1^m$ is independent and $\{\alpha_j\}_1^n$ spans V then the sequence $\{\beta_i\}_1^m$ can be extended to a basis.

Proof: We simply put the β_i's before the α_j's in a single sequence and proceed as above. The first m choices will be exactly $\beta_1, \ldots, \beta_m$ because they are independent.

D. The existence of linear transformations. If we follow a coordinate isomorphism by a linear combination map we get the following existence theorem, which we state only in n-tuple form.

THEOREM 16. If $\{\beta_i\}_1^n$ is an ordered basis for the vector space V and if $\{\alpha_i\}_1^n$ is any n-tuple of vectors in a vector space W then there exists $S \in \mathrm{Hom}(V, W)$ such that $S(\beta_i) = \alpha_i$ for $i = 1, \ldots, n$.

Proof: If $\theta : \underline{x} \mapsto \Sigma x_i \beta_i$ is the basis isomorphism and $T : \underline{x} \mapsto \Sigma x_i \alpha_i$ the linear combination map then $S = T \circ \theta^{-1}$ is linear and $S(\beta_i) = T(\theta^{-1}\beta_i) = T(\delta^i) = \alpha_i$ for $i = 1, \ldots, n$.

The above transformation is in fact uniquely determined by its values at the basis elements $\{\beta_i\}$ but uniqueness only depends on having a spanning set.

LEMMA 14. If $B \subset V$ and $L(B) = V$ then any $T \in \mathrm{Hom}(V,W)$ is uniquely determined by its values on B .

Proof: We have to show that if $T = S$ on B then $T = S$. But any $\xi \in V$ is a linear combination $\xi = \Sigma_1^n x_i \beta_i$ with $\{\beta_i\} \subset B$, and therefore

$$T(\xi) = \Sigma_1^n x_i T(\beta_i) = \Sigma_1^n x_i S(\beta_i) = S(\xi) , \text{ q.e.d.}$$

More elegantly, we have to show the linear map $T \longmapsto T \restriction B$ to be injective, i.e., $N(T)$ to be 0 . But if $T[B] = \{0\}$ then $T[V] = T[L(B)] = L(T[B]) = L(\{0\}) = \{0\}$, and so $T = 0$.

It is natural to ask how the unique S above varies with the n-tuple $\{\alpha_i\}$. The answer is : linearly.

THEOREM 17. Let $\{\beta_i\}_1^n$ be a fixed ordered basis for the vector space V and for each n-tuple $\underline{\alpha} = \{\alpha_i\}_1^n$ chosen from the vector space W let $S_{\underline{\alpha}} \in \mathrm{Hom}(V,W)$ be the unique transformation defined above. Then the map $\underline{\alpha} \longmapsto S_{\underline{\alpha}}$ is an isomorphism from W^n to $\mathrm{Hom}(V,W)$.

Proof: We know so far that the mapping is uniquely defined on W^n . It is linear:

$$T_{c\underline{\alpha}+d\underline{\gamma}}(\xi) = T_{c\underline{\alpha}+d\underline{\gamma}}(\Sigma x_i\beta_i) = \Sigma x_i(c\alpha_i + d\gamma_i) = c\,\Sigma x_i\alpha_i + d\,\Sigma x_i\gamma_i$$

$$= c\,T_{\underline{\alpha}}(\Sigma x_i\beta_i) + dT_{\underline{\gamma}}(\Sigma x_i\beta_i) = c\,T_{\underline{\alpha}}(\xi) + dT_{\underline{\gamma}}(\xi) .$$

Thus $T_{c\underline{\alpha}+d\underline{\gamma}} = cT_{\underline{\alpha}} + dT_{\underline{\gamma}}$. It is injective : if $T_{\underline{\alpha}} = 0$ then

$T_{\underline{\alpha}}(\beta_i) = \alpha_i = 0$ for all i and so $\underline{\alpha} = 0$. Finally, it is surjective : if $T \in \mathrm{Hom}(V, W)$ and $\alpha_i = T(\beta_i)$, $i = 1, \ldots, n$, then $T = T_{\underline{\alpha}}$ by the above uniqueness theorem again.

E. Infinite bases. Most vector spaces do not have finite bases and it is natural to try to extend the above discussion to index sets I that may be infinite. The Kronecker functions $\{ \delta^i : i \in I \}$ have the same definitions but they no longer span $\mathbb{R}^I$. By definition f is a linear combination of the functions δ^i if and only if f is of the form $\Sigma_{i \in I_1} c_i \delta^i$ where I_1 is a finite subset of I . But then $f = 0$ outside of I_1 . Conversely if $f \in \mathbb{R}^I$ is 0 except on a finite set I_1 then $f = \Sigma_{i \in I_1} f(i)\delta^i$. The linear span of $\{\delta^i : i \in I\}$ is thus exactly the set of all functions of $\mathbb{R}^I$ that are zero except on a finite set. We shall designate this subspace $\mathbb{R}_I$.

If $\{\alpha_i : i \in I\}$ is an indexed set of vectors in V and $f \in \mathbb{R}_I$ then the sum $\Sigma_{i \in I} f(i)\alpha_i$ becomes meaningful if we adopt the reasonable convention that the sum of an arbitrary number of 0's is 0 . Then $\Sigma_{i \in I} = \Sigma_{i \in I_0}$ where I_0 is any finite subset of I outside of which f is zero.

With this convention, $T_{\underline{\alpha}} : f \longmapsto \Sigma_i f(i)\alpha_i$ is a linear map of $\mathbb{R}_I$ to V, as in Theorem 12. And with the same convention $\Sigma_{i \in I} f(i)\alpha_i$ is an elegant expression for the general linear combination of the vectors α_i . Instead of choosing a finite subset I_1 and numbers c_i for just those indices i in I_1, we define c_i for all $i \in I$ but with the stipulation that $c_i = 0$ for all but a finite number of indices. That is, we take $\underline{c} = \{c_i : i \in I\}$ as a function in $\mathbb{R}_I$.

We make the same definitions of <u>independence</u> and <u>basis</u> as before. Then $\{\alpha_i : i \in I\}$ is a basis for V if and only if $T_{\underline{\alpha}} : \mathbb{R}_I \longrightarrow V$ is an isomorphism, i.e., if and only if for each $\xi \in V$ there exists a unique $\underline{x} \in \mathbb{R}_I$ such that $\xi = \Sigma_i x_i \alpha_i$.

The above sum is always finite (despite appearances) and the above notion of basis is purely algebraic. However, infinite bases in this sense are not very useful in analysis, and we shall therefore concentrate for the present on spaces that have finite bases (i.e., are finite dimensional). Then in one important context later on we shall discuss infinite bases where the sums are genuinely infinite by virtue of limit theory.

§ 6. Bilinearity

A. Bilinear mappings. The notion of a bilinear mapping is important to the understanding of linear algebra because it is the vector setting for the duality principle (0.12).

DEFINITION. If U, V and W are vector spaces then a mapping $\omega: \langle \xi, \eta \rangle \mapsto \omega(\xi, \eta)$ from $U \times V$ to W is bilinear if it is linear in each variable when the other variable is held fixed.

That is, if we hold ξ fixed then $\eta \mapsto \omega(\xi, \eta)$ is linear (and so belongs to $\mathrm{Hom}(V, W)$), and if η is held fixed then similarly $\omega(\xi, \eta)$ is in $\mathrm{Hom}(U, W)$ as a function of ξ. This is not the same notion as linearity on the product vector space $U \times V$. For example, $\langle x, y \rangle \mapsto x+y$ is a linear mapping from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$ but it is not bilinear. If y is held fixed the mapping $x \mapsto x+y$ is affine but not linear unless y is 0. On the other hand, $\langle x, y \rangle \mapsto xy$ is a bilinear mapping from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$ but it is not linear. If y is held fixed the mapping $x \mapsto yx$ is linear. But the sum of two ordered couples does not map to the sum of their images :

$$\langle x, y \rangle + \langle u, v \rangle = \langle x+u, y+v \rangle \mapsto (x+u)(y+v)$$

which is not the sum of the images, $xy + uv$.

The linear meaning of bilinearity is partly explained in the following theorem.

THEOREM 18. If $\omega : U \times V \longrightarrow W$ is bilinear then ω is equivalent by duality to a linear mapping from U to Hom(V, W) and also to a linear mapping from V to Hom(U, W).

<u>Proof:</u> For each fixed $\eta \in V$ let ω_η be the mapping $\xi \longmapsto \omega(\xi, \eta)$. That is, $\omega_\eta(\xi) = \omega(\xi, \eta)$. Then $\omega_\eta \in \mathrm{Hom}(U, W)$ by the bilinear hypothesis. The mapping $\eta \longmapsto \omega_\eta$ is thus from V to Hom(U, W), and <u>its</u> linearity is due to the linearity of ω in η when ξ is held fixed:

$$\omega_{c_1\eta_1 + c_2\eta_2}(\xi) = \omega(\xi, c_1\eta_1 + c_2\eta_2) = c_1\omega(\xi, \eta_1) +$$

$$c_2\omega(\xi, \eta_2) = c_1\omega_{\eta_1}(\xi) + c_2\omega_{\eta_2}(\xi), \quad \text{so that}$$

$$\omega_{c_1\eta_1 + c_2\eta_2} = c_1\omega_{\eta_1} + c_2\omega_{\eta_2}.$$

Similarly, if we define ω^ξ by $\omega^\xi(\eta) = \omega(\xi, \eta)$ then $\xi \longmapsto \omega^\xi$ is a linear mapping of U to Hom(V, W). Conversely, if $T : U \longrightarrow \mathrm{Hom}(V, W)$ is linear then the function ω defined by $\omega(\xi, \eta) = (T(\xi))(\eta)$ is bilinear. Moreover $\omega^\xi = T(\xi)$ so that T is the mapping $\xi \longmapsto \omega^\xi$.

We shall see that bilinearity occurs frequently. Sometimes the reinterpretation provided by the above theorem provides new insights and at other times it seems less helpful.

For example, the composition map $\langle S, T \rangle \longmapsto S \circ T$ is bilinear, and the corollary of Theorem 2 to the effect that composition on the right by a fixed T is a linear map is simply part of an explicit statement of the bilinearity. But the linear map $T \mapsto$ composition by T is a fantastic object that we have no need for except in the case $W = \mathbb{R}$.

On the other hand, the linear combination formula $\Sigma_1^n x_i \alpha_i$ and Theorem 17 do receive new illumination.

THEOREM 19. The mapping $\omega(\underline{x}, \underline{\alpha}) = \Sigma_1^n x_i \alpha_i$ is bilinear from $\mathbb{R}^n \times V^n$ to V and the mapping $\underline{\alpha} \mapsto \omega_{\underline{\alpha}}$ is therefore a linear mapping from V^n to $\mathrm{Hom}(\mathbb{R}^n, V)$.

Proof: The linearity of ω in $\underline{x}$ for a fixed $\underline{\alpha}$ is exactly the assertion of Theorem 12, and its linearity in $\underline{\alpha}$ for a fixed $\underline{x}$ is seen in the same way. Then $\underline{\alpha} \mapsto \omega_{\underline{\alpha}}$ is linear by Theorem 18.

Composing $\omega_{\underline{\alpha}}$ with the fixed inverse of the basis isomorphism $\theta : \underline{x} \mapsto \Sigma x_i \beta_i$ we obtain the linear map $\underline{\alpha} \mapsto \omega_{\alpha} \circ \theta^{-1} = S_{\underline{\alpha}}$ of Theorem 17. Of course its bijectivity still has to be established. From our present point of view we would probably do this first for $\underline{\alpha} \mapsto \omega_{\underline{\alpha}}$ where $\omega_{\underline{\alpha}}(\underline{x}) = \Sigma_1^n x_i \alpha_i$, and then transfer the results to the map $\underline{\alpha} \mapsto S_{\underline{\alpha}}$ by composing $\omega_{\underline{\alpha}}$ with the fixed coordinate isomorphism θ^{-1}.

Canonical

B. Natural isomorphisms. Often we find two vector spaces related to each other in such a way that a particular isomorphism between them is singled out. This phenomenon is hard to pin down in general terms but easy to describe by examples.

Duality is one source of such "natural" isomorphisms. For example, an $m \times n$ matrix $\{t_{ij}\}$ is a real-valued function of the two variables $<i,j>$, and as such is an element of the Cartesian space $\mathbb{R}^{\bar{m} \times \bar{n}}$. We can also view $\{t_{ij}\}$ as a sequence of n column vectors in $\mathbb{R}^m$. This is the dual point of view where we hold j fixed and obtain a function of i for each j , and from this point of view $\{t_{ij}\}$ is an element of $(\mathbb{R}^{\bar{m}})^{\bar{n}}$. This correspondence between $\mathbb{R}^{\bar{m} \times \bar{n}}$ and $(\mathbb{R}^{\bar{m}})^{\bar{n}}$ is clearly an isomorphism, and is an example of a natural isomorphism.

We review next the various ways of looking at Cartesian n-space itself. One standard way of defining an ordered n-tuplet is by induction. The ordered triplet $<x,y,z>$ is defined as the ordered pair $<<x,y>,z>$, and the ordered n-tuplet $<x_1,\ldots,x_n>$ as $<<x_1,\ldots,x_{n-1}>, x_n>$. Thus $\mathbb{R}^n$ is defined inductively by setting $\mathbb{R}^1 = \mathbb{R}$ and $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}$.

The ordered n-tuplet can also be defined as the function on $\bar{n} = \{1,\ldots,n\}$ which assigns x_i to i . Then

$$<x_1,\ldots,x_n> = \{<1,x>,\ldots,<n,x_n>\}$$

and and Cartesian n-space is $\mathbb{R}^{\bar{n}} = \mathbb{R}^{\{1,\ldots,n\}}$.

Finally, we often wish to view Cartesian (n+m)-space as the Cartesian product of Cartesian n-space with Cartesian m-space, so that now we take $<x_1,\ldots,x_{n+m}>$ as $<<x_1,\ldots,x_n>, <x_{n+1},\ldots,x_{n+m}>>$ and $\mathbb{R}^{n+m}$ as $\mathbb{R}^n \times \mathbb{R}^m$.

Here, again, if we pair two different models for the same n-tuplet we have an obvious natural isomorphism between the corresponding models for Cartesian n-space.

Finally, the characteristic properties of Cartesian product spaces given in Theorems 4 and 5 yield natural isomorphisms. Theorem 4 says that an n-tuple of linear maps $\{T_i\}_1^n$ on a common domain V is equivalent to a single n-tuple valued map T , where $T(\xi) = < T_1(\xi), \ldots, T_n(\xi) >$ for all $\xi \in V$. (This is duality again ! $T_i(\xi)$ is a function of the two variables i and ξ .) And it is not hard to see that this identification of T with $\{T_i\}_1^n$ is an isomorphism.

Similarly, Theorem 5 identifies an n-tuple of linear maps $\{T_i\}_1^n$ into a common codomain V with a single linear map T of an n-tuple variable and this identification is a natural isomorphism from $\Pi_1^n \operatorname{Hom}(W_i, V)$ to $\operatorname{Hom}(\Pi_1^n W_i, V)$.

An arbitrary isomorphism between two vector spaces identifies them in a transient way. For the moment we think of them as representing the same abstract space, but only as long as the isomorphism is before us. If we shift to a different isomorphism between them we obtain a new temporary identification. Natural isomorphisms , on the other hand, effect permanent identifications, and we think of paired objects as being two aspects of the same object in a deeper sense. Thus we think of a matrix as "being" either a sequence of row vectors, or a sequence of column vectors, or a single function of two integer indices.

Independent of basis

C. We can now make the ultimate dissection of the theorems centering around the linear combination formula. The axioms S1-S3 say <u>exactly</u> that the scalar product $x\alpha$ is bilinear. More precisely, they say that the mapping $S : \langle x, \alpha \rangle \longmapsto x\alpha$ from $\mathbb{R} \times W$ to W is bilinear. In the language of Theorem 18 $x\alpha = \omega_\alpha(x)$ and from that theorem we conclude that the mapping $\alpha \longmapsto \omega_\alpha$ is a linear mapping from W to $\mathrm{Hom}(\mathbb{R}, W)$. It is easily checked to be an isomorphism. If $\omega_\alpha = 0$ then $0 = \omega_\alpha(1) = 1 \cdot \alpha = \alpha$, so that it is injective. And if $T \in \mathrm{Hom}(\mathbb{R}, W)$ then $T(x) = xT(1) = x\alpha$ where $\alpha = T(1)$ so that $T = \omega_\alpha$ and the mapping is surjective.

This isomorphism between W and $\mathrm{Hom}(\mathbb{R}, W)$ extends to an isomorphism from W^n to $(\mathrm{Hom}(\mathbb{R}, W))^n$, which in turn is naturally isomorphic to $\mathrm{Hom}(\mathbb{R}^n, W)$ by the second Cartesian product isomorphism. Thus W^n is naturally isomorphic to $\mathrm{Hom}(\mathbb{R}^n, W)$, the mapping being $\underline{\alpha} \longmapsto T_{\underline{\alpha}}$ where $T_{\underline{\alpha}}(\underline{x}) = \Sigma_1^n x_i \alpha_i$. To get the form of Theorem 17 we compose with the isomorphism $T \longmapsto T \circ \theta^{-1}$ from $\mathrm{Hom}(\mathbb{R}^n, W)$ to $\mathrm{Hom}(V, W)$, where θ is a basis isomorphism $\underline{x} \longmapsto \Sigma x_i \beta_i$.

§ 7. Dimension.

A. The fundamental lemma. The concept of dimension rests on the fact that two different bases for the same space always contain the same number of elements. This number, which is then the number of elements in every basis for V, is called the dimension of V. It tells all there is to know about V to within isomorphism: there exists an isomorphism between two spaces if and only if they have the same dimension. We shall consider only finite dimensional spaces. If V is not finite dimensional, its dimension is an infinite cardinal number, a concept with which the reader is probably unfamiliar.

The following replacement lemma is fundamental.

LEMMA 15. If A and B are subsets of V such that A is independent and B spans V, then any $\alpha \in A-B$ can be used to replace some $\beta \in B-A$ in such a way that the new set still spans.

Proof: By hypothesis α is a finite linear combination on B, $\alpha = \Sigma_1^n c_i\beta_i$, where ~~all~~ $\exists\, c_i \neq 0$; and the β_i are distinct elements of B. If all β_i were in A this equation in the form $\alpha - \Sigma c_i\beta_i = 0$ would contradict the independence of A. Therefore some $\beta_{i_0} \notin A$. Let C be the set obtained from B by replacing β_{i_0} by α. Since the above equation can be solved for β_{i_0}, we have $\beta_{i_0} \in L(C)$. Thus $B \subset L(C)$, and so $V = L(B) \subset L(C)$. Therefore $L(C) = V$ and we are done.

Notice that the number of elements in A-C is one less than in A-B ; we removed from B an element not in A and replaced it by an element of A .

In the theorem below $\# A$ is the number of elements in A .

THEOREM 20. If A and B are finite subsets of a vector space V such that A is independent and B spans V then $\# A \leqq \# B$.

Proof: The proof is an induction on $\#(A-B)$. If $\#(A-B) = 0$, then $A \subset B$ and we are done. Suppose that the theorem is true if $\#(A-B) < n$ and suppose that we have a pair A and B for which $\#(A-B) = n$. If C is defined as in the lemma, then $\#(A-C) = n-1$, so that $\# A \leqq \# C$ by the above inductive hypothesis. But $\# C = \# B$, and therefore $\# A \leqq \# B$, q.e.d.

DEFINITION . V is finite dimensional if $V = L(A)$ for some finite set A .

THEOREM 21. If V is finite dimensional then there is a non-negative integer n , called the dimension of V , such that every basis has n elements.

Proof: The finite spanning set A includes a basis B (Theorem 14) so that V has at least one finite basis. By the above theorem, $\# C \leqq \# B$ for any independent set C . If C is also a basis, the same theorem gives us $\# B \leqq \# C$. Thus every two bases have the same (finite) number of elements.

THEOREM 22. Two finite dimensional vector spaces are isomorphic if and only if they have the same dimension.

Proof: If T is an isomorphism of V onto W and B is a basis for V then $T[B]$ is a basis for W by Theorem 13. Therefore $d(V) = \# B = \# T[B] = d(W)$. Conversely, if $d(V) = d(W) = n$ then V and W are each isomorphic to $\mathbb{R}^n$ and so to each other.

THEOREM 23. Every subspace M of a finite dimensional vector space V is finite dimensional.

Proof: Let $\mathcal{A}$ be the family of finite independent subsets of M. By Lemma 15, if $A \in \mathcal{A}$ then $\# A \leqq d(V)$. Thus $\{\# A : A \in \mathcal{A}\}$ is a finite set of integers, and we can choose $B \in \mathcal{A}$ such that $n = \# B$ is the maximum of this finite set. But then $L(B) = M$, because otherwise for any $\alpha \in M - L(B)$ we have $B \cup \{\alpha\} \in \mathcal{A}$ by Lemma 13 and $\#(B \cup \{\alpha\}) = n+1$, contradicting the maximal nature of n. Thus M is finitely spanned, q.e.d.

B. Dimensional identities. We now prove two basic dimensional identities. V will always be assumed finite dimensional.

LEMMA 16. If V_1 and V_2 are complementary subspaces of V then $d(V) = d(V_1) + d(V_2)$.

Proof: This follows at once from Theorem 13a.

LEMMA 17. Let U, W be any two subspaces of a vector space V and let U_1 be a complement of $U \cap W$ in U. Then U_1 is also a complement of W in $U+W$.

Proof: First, $U_1 + W = U_1 + ((U \cap W) + W) = (U_1 + (U \cap W)) + W) = U+W$. We have used the obvious fact that the sum of a vector space and a subspace is the vector space. Next, $U_1 \cap W = (U_1 \cap U) \cap W = U_1 \cap (U \cap W) = \{0\}$ because U_1 is a complement of $U \cap W$ in U. We thus have both $U_1 + W = U+W$ and $U_1 \cap W = \{0\}$, and the lemma follows from the Corollary to Lemma 9.

THEOREM 24. If U and W are subspaces of a finite dimensional vector space then $d(U+W) + d(U \cap W) = d(U) + d(W)$.

This is a corollary of the above two lemmas. We have $d(U) + d(W) = (d(U \cap W) + d(U_1)) + d(W) = d(U \cap W) + (d(U_1) + d(W)) = d(U \cap W) + d(U + W)$.

THEOREM 25. Let V be finite dimensional and let W be any vector space. Let $T \in \mathrm{Hom}(V, W)$ have nullspace N (in V) and range R (in W). Then R is finite dimensional and $d(V) = d(N) + d(R)$.

Proof: Let U be a complement of N in V. Then we know that $T \restriction U$ is an isomorphism onto R. (See Theorem 11.) Therefore R is finite dimensional and $d(R) + d(N) = d(U) + d(N) = d(V)$, by our first identity.

COROLLARY. If W is finite dimensional and $d(W) = d(V)$, then T

is injective if and only if it is surjective, so that in this case injectivity, surjectivity and bijectivity are all equivalent.

<u>Proof:</u> T is surjective if and only if $R = W$. But this is equivalent to $d(R) = d(W)$, and if $d(W) = d(V)$ then the theorem shows this in turn to be equivalent to $d(N) = 0$, i.e., to $N = \{0\}$, q.e.d.

THEOREM 26. If $d(V) = n$ and $d(W) = m$ then $Hom(V, W)$ is finite dimensional and its dimension is mn .

<u>Proof:</u> By Theorem 17 $Hom(V, W)$ is isomorphic to W^n which is the direct sum of the n-subspaces isomorphic to W under the injections θ_i , $i = 1, \ldots, n$. The dimension of W^n is therefore $\Sigma_1^n m = mn$ by the corollary to Theorem 13a.

Another proof of Theorem 26 will be available in § .

§ 8. The dual space

A. Dual bases. Throughout this section all spaces will be assumed finite dimensional. Many of the definitions and properties are valid for infinite dimensional spaces as well, but for such spaces there is a difference between purely algebraic situations and situations in which algebra is mixed with hypotheses of continuity. One of the blessings of finite dimensionality is the absence of this complication. As the reader has probably surmized from the number of special linear functionals we have met, particularly the coordinate functionals, the space Hom(V, ℝ) of all linear functionals on V plays a special role.

DEFINITION. The dual space (or conjugate space) V^* of the vector space V is the vector space Hom(V, ℝ) of all linear mappings from V to ℝ. Its elements are called linear functionals.

One naturally wonders how big a space V^* is, and we settle the question immediately.

THEOREM 27. Let $\{\beta_i\}_1^n$ be an ordered basis for V, and let ϵ_j be the corresponding j^{th} coordinate functional on V : $\xi \longmapsto x_j$, where $\xi = \Sigma_1^n x_i\beta_i$. Then $\{\epsilon_j\}_1^n$ is an ordered basis for V^*.

Proof: Let us first make the proof by a direct calculation.

(a) Independence. Suppose that $\Sigma_1^n c_j\epsilon_j = 0$, i.e., that $\Sigma_1^n c_j\epsilon_j(\xi) = 0$ for all $\xi \in V$. Taking $\xi = \beta_i$ and noticing that $\epsilon_j(\beta_i) = 0$

if $j \neq i$ and $\epsilon_i(\beta_i) = 1$ (i.e., $\epsilon_j(\beta_i) = \delta^i_j$) we see that the above sum reduces to $c_i = 0$, and this for all i. Therefore $\{\epsilon_j\}_1^n$ is independent.

(b) Spanning. Now let λ be any element of V^* and set $\ell_i = \lambda(\beta_i)$, for all i. Then $\lambda(\xi) = \lambda(\Sigma_1^n x_i\beta_i) = \Sigma_1^n \ell_i x_i = \Sigma_1^n \ell_i \epsilon_i(\xi)$, and so $\lambda = \Sigma \ell_i \epsilon_i$. This shows that $\{\epsilon_j\}_1^n$ spans V^*, and together with (a) that it is a basis.

DEFINITION. The basis $\{\epsilon_j\}$ for V^* is called the <u>dual</u> of the basis $\{\beta_i\}$ for V.

As usual, one of our fundamental isomorphisms is lurking behind all of this. By Theorem 17, if $\{\ell_i\}_1^n \in \mathbb{R}^n$ and if $\lambda \in V^*$ is the unique functional such that $\lambda(\beta_i) = \ell_i$, $i = 1, \ldots, n$, then the correspondence $\{\ell_i\} \longmapsto \lambda$ is an isomorphism of $\mathbb{R}^n$ onto V^*. The j^{th} standard basis element in $\mathbb{R}^n$ is the n-tuple δ^j ($\delta^j_i = 0$ if $i \neq j$ and $\delta^j_j = 1$) and the corresponding functional is clearly ϵ_j. The set $\{\epsilon_j\}_1^n$ is thus the image of the standard basis of $\mathbb{R}^n$ under the isomorphism, and is therefore a basis.

COROLLARY. $d(V^*) = d(V)$.

B. <u>The second conjugate space.</u> By choosing a basis in V we have set up the coordinate isomorphism from V to $\mathbb{R}^n$ and the above isomorphism from $\mathbb{R}^n$ to V^*, and have therefore defined an isomorphism between V and V^*. This isomorphism varies with the basis, and there is no natural isomorphism between V and V^* . However, with $(V^*)^*$ it is another matter.

THEOREM 28. The function $\omega: V \times V^* \longrightarrow \mathbb{R}$ defined by $\omega(\xi, f) = f(\xi)$ is bilinear, and the mapping $\xi \longrightarrow \omega^\xi$ from V to V^{**} is a natural isomorphism.

<u>Proof:</u> In this context we generally set $\xi^{**} = \omega^\xi$, so that ξ^{**} is defined by $\xi^{**}(f) = f(\xi)$ for all $f \in V^*$. The bilinearity of ω should be clear and Theorem 18 therefore applies. The reader might like to run through a direct check of the linearity of $\xi \longrightarrow \xi^{**}$, starting with $(c_1\xi_1 + c_2\xi_2)^{**}(f)$. There still is the question of the injectivity of this mapping.

If $\alpha \neq 0$ we can find $f \in V^*$ so that $f(\alpha) \neq 0$. One way is to make α the first vector of an ordered basis and take f as the first functional in the dual basis. Then $f(\alpha) = 1$. Thus if $\alpha \neq 0$ then $(\exists f)(\alpha^{**}(f) \neq 0)$, and so $\alpha^{**} \neq 0$, and the mapping $\xi \longrightarrow \xi^{**}$ is injective. It is bijective by the corollary to Theorem 25.

If we think of V^{**} as being naturally identified with V in this way the two spaces V and V^* are symmetrically related to each other. Each is the dual of the other. In the expression " $f(\xi)$" we think of <u>both</u> symbols as variables and then hold one or the other fixed for the two interpretations. In such a situation we often use a more symmetric symbolism, such as (f, ξ) to indicate our intention to treat both symbols as variables.

LEMMA 18. If $\{\lambda_i\}$ is the basis in V^* dual to the basis $\{\alpha_i\}$ in V, then $\{\alpha_i^{**}\}$ is the basis in V^{**} dual to the basis $\{\lambda_i\}$ in V^*.

Proof: We have $\alpha_i^{**}(\lambda_j) = \lambda_j(\alpha_i) = \delta_j^i$, showing that α_i^{**} is the i^{th} coordinate projection. In case the reader has forgotten, the basis expansion $f = \Sigma c_j \lambda_j$ implies that $\alpha_i^{**}(f) = f(\alpha_i) = (\Sigma c_j \lambda_j)(\alpha_i) = c_i$ so that α_i^{**} is the mapping $f \longrightarrow c_i$.

C. Orthogonality. It is in this dual situation that orthogonality first naturally appears. However, we shall save the term "orthogonal" for the later context in which V and V^* have been identified through a scalar product, and shall speak here of the annihilator of a set rather than its orthogonal complement.

DEFINITION. If $A \subset V$ the annihilator of A, A^o, is the set of all $f \in V^*$ such that $f(\alpha) = 0$ for all $\alpha \in A$. Similarly if $A \subset V^*$ then $A^o = \{\alpha \in V : f(\alpha) = 0 \text{ for all } f \in A\}$. If we view V as $(V^*)^*$, the second definition is included in the first.

The following properties are easily established and will be left as exercises.

(1) A^o is always a subspace.

(2) $A \subset B \Rightarrow B^o \subset A^o$.

(3) $(L(A))^o = A^o$.

(4) $(A \cup B)^o = A^o \cap B^o$.

(5) $L(A) \subset A^{oo}$.

We now add one more crucial dimensional identity to those of the last section.

THEOREM 29. If W is a subspace of V then $d(V) = d(W) + d(W^o)$.

<u>Proof:</u> Let $\{\beta_i\}_1^m$ be a basis for W, and extend it to a basis $\{\beta_i\}_1^n$ for V. Let $\{\lambda_i\}_1^n$ be the dual basis in V^* and set $U = L(\{\lambda_i\}_{m+1}^n)$. We claim that $W^o = U$. First if $j > m$ then $\lambda_j(\beta_i) = 0$, $i=1,\ldots, m$ and so $\lambda_j \in L(\{\beta_i\}_1^m)^o = W^o$. Thus $U \subset W^o$. Now suppose that $f \in W^o$ and let $f = \Sigma_{j=1}^n c_j\lambda_j$ be its (dual) basis expansion. Then, for each $i \leqq m$, $c_i = f(\beta_i) = 0$, so that $f = \Sigma_{m+1}^n c_j\lambda_j \in U$. Thus $W^o \subset U$. The two inclusions imply that $W^o = U$, as claimed. Since $n = m + (n-m)$ and $n-m = d(U)$, we are done.

COROLLARY. $A^{oo} = L(A)$ for every subset $A \subset V$.

<u>Proof:</u> Since $L(A)^o = A^o$ we have $d(L(A)) + d(A^o) = d(V)$, by the theorem. Also $d(A^o) + d(A^{oo}) = d(V^*) = d(V)$. Thus $d(A^{oo}) = d(L(A))$, and since $L(A) \subset A^{oo}$, by (5) above, we have $L(A) = A^{oo}$.

D. <u>The adjoint of T.</u> If $T \in \mathrm{Hom}(V, W)$ and $\ell \in W^*$ then of course $\ell \circ T \in V^*$, and right composition by a fixed $T \in \mathrm{Hom}(V,W)$ is a linear mapping of W^* into V^* by the corollary to Theorem 2. This mapping is called the <u>adjoint</u> of T and is designated T^*. Thus $T^* : W^* \longrightarrow V^*$ is the map $\ell \longmapsto \ell \circ T$.

$T^*\ell = \ell \circ T$

THEOREM 30. The mapping $T \longmapsto T^*$ is an isomorphism from the vector space $\mathrm{Hom}(V, W)$ to the vector space $\mathrm{Hom}(W^*, V^*)$. Also $(T \circ S)^* = S^* \circ T^*$ under the relevant hypotheses on domains and codomains.

<u>Proof:</u> Everything that we have said above through the linearity of $T \longmapsto T^*$ is a consequence of the bilinearity of $\omega(\ell, T) = \ell \circ T$. The map we have called T^* is simply ω_T and the linearity of $T \longmapsto T^*$ thus follows from Theorem 18. Again the reader might benefit from a direct linearity check, beginning with $(c_1 T_1 + c_2 T_2)^* (\ell)$.

To see that $T \longmapsto T^*$ is injective, we take any $T \neq 0$ and choose $\alpha \in V$ so that $T(\alpha) \neq 0$. We then choose $\ell \in W^*$ so that $\ell(T(\alpha)) \neq 0$. Since $\ell(T(\alpha)) = (T^*(\ell))(\alpha)$ we have verified that $T^* \neq 0$.

Next, if $d(V) = m$ and $d(W) = n$ then also $d(V^*) = m$ and $d(W^*) = n$ by the corollary of Theorem 27, and $d(\mathrm{Hom}(V, W)) = mn = d(\mathrm{Hom}(W^*, V^*))$ by Theorem 26. The injective map $T \longmapsto T^*$ is thus an isomorphism (Theorem 25, corollary).

Finally, $(T \circ S)^* \ell = \ell \circ (T \circ S) = (\ell \circ T) \circ S = S^*(\ell \circ T) = S^*(T^*(\ell)) = (S^* \circ T^*)\ell$, so that $(T \circ S)^* = S^* \circ T^*$.

The reader would probably guess that T^{**} becomes identified with T under the identification of V with V^{**}. The proof will be left as an exercise (starred, of course).

Finally we have the following elementary identity, which is like $e^{i\pi} = -1$ in being a seemingly mystical but in reality trivial relationship among a number of different concepts.

THEOREM 30. $R(T^*) = (N(T))^o$ and $N(T^*) = (R(T))^o$.

<u>Proof:</u> The following statements are definitionally equivalent in pairs as they occur : $\ell \in N(T^*)$, $T^*(\ell) = 0$, $\ell \circ T = 0$, $\ell(T(\xi)) = 0$ for all $\xi \in V$, $\ell \in (R(T))^o$. Therefore $N(T^*) = (R(T))^o$. The other proof is similar and will be left to the reader.

COROLLARY. $d(R(T^*)) = d(R(T))$ and $d(N(T^*)) = d(N(T))$.

<u>Proof:</u> The dimensions of $R(T)$ and $(N(T))^o$ are each $d(V)-d(N(T))$, by Theorems 25 and 29, and the second is $d(R(T^*))$ by the above theorem. Therefore, $d(R(T)) = d(R(T^*))$. The proof that $d(N(T)) = d(N(T^*))$ is practically identical.

§ 9. Matrices.

A. Matrices and linear transformations.

Presumably the reader has already acquired some information about matrices and their relationship to linear transformations from the exercises. But to be on the safe side we start again at the beginning.

By popular conception a matrix is a rectangular array of numbers

$$\begin{matrix} t_{11} & t_{12} \cdots & t_{1n} \\ t_{21} & & t_{2n} \\ \vdots & & \\ t_{m1} & & t_{mn} \end{matrix}$$

Notice that the first index numbers the rows and the second index numbers the columns. If there are m rows and n columns in the array it is called an $m \times n$ matrix. This notion is inexact. A rectangular array is a way of picturing a matrix, but a matrix is really a function, just as a sequence is a function. With the notation $\bar{m} = \{1, \ldots, m\}$ the above matrix is a function assigning a number to every pair of integers $\langle i, j\rangle \in \bar{m} \times \bar{n}$. It is thus an element of the set $\mathbb{R}^{\bar{m} \times \bar{n}}$. The addition of two $m \times n$ matrices is performed in the obvious place by place way, and is just the addition of two functions in $\mathbb{R}^{\bar{m} \times \bar{n}}$; similarly for scalar multiplication. The set of all $m \times n$ matrices is thus the vector space $\mathbb{R}^{\bar{m} \times \bar{n}}$, a Cartesian space with a rather fancy finite index set. We shall use the customary index notation t_{ij} for the value $t(i,j)$ of the function $\underline{t}$ at $< i,j >$, and we shall also write $\{t_{ij}\}$ for $\underline{t}$ just as we do for sequences and other indexed collections.

The additional properties of matrices stem from the fact that each $m \times n$ matrix $\{t_{ij}\}$ directly determines a corresponding transformation $T \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$.

THEOREM 32. Let $\underline{t}^j \in \mathbb{R}^m$ be the j^{th} column of the matrix $\{t_{ij}\}$, and let T be the map $\underline{x} \longmapsto \underline{y} = \Sigma_{j=1}^n x_j \underline{t}^j$ from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then T is linear, and the map $\{t_{ij}\} \longmapsto T$ is an isomorphism from the space $\mathbb{R}^{\bar{m} \times \bar{n}}$ of all $m \times n$ matrices to $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$.

Proof: Clearly, T is just the linear combination map from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by the n-tuple $\{\underline{t}^j\}_1^n \in (\mathbb{R}^m)^n$ and is linear by Theorem 12. Then Theorem 17 shows that $\{\underline{t}^j\}_1^n \longmapsto T$ is an isomorphism from $(\mathbb{R}^m)^n$ to $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$. Also, $\{t_{ij}\} \longmapsto \{\underline{t}^j\}_1^n$ is an isomorphism from $\mathbb{R}^{\bar{m} \times \bar{n}}$ to $(\mathbb{R}^m)^n$, the natural isomorphism arising from viewing a matrix as an n-tuple of column m-tuples (duality). The composition of these isomorphisms is the isomorphism of the theorem, q.e.d.

There is another way of obtaining T from $\{t_{ij}\}$ which is illuminating. If we take i^{th} coordinates in the m-tuple equation

$\underline{y} = \Sigma_{j=1}^{n} x_j \underline{t}^j$ we get the equivalent and familiar system of numerical (scalar) equations $y_i = \Sigma_{j=1}^{n} t_{ij} x_j$, $i = 1, \ldots, m$. Now a mapping $\underline{x} \longmapsto \Sigma_{j=1}^{n} t_j x_j$ from $\mathbb{R}^n$ to $\mathbb{R}$ is simply a linear combination mapping of Theorem 12 for the special case $V = \mathbb{R}$. In the above numerical equations, therefore, we have simply used the m rows of the matrix $\{t_{ij}\}$ to define an m-tuple of linear functionals on $\mathbb{R}^n$, which in turn is equivalent to a single m-tuple valued linear mapping $\bar{T}$ (in $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$) by Theorem 4.

The choice of ordered bases for arbitrary finite dimensional spaces V and W allow us to transfer the above theorem to $\mathrm{Hom}(V, W)$.

THEOREM 33. Let $\{\alpha_j\}_1^n$ and $\{\beta_i\}_1^m$ be ordered bases for the vector spaces V and W respectively. For each matrix $\{t_{ij}\} \in \mathbb{R}^{\bar{m} \times \bar{n}}$ let T be the unique element of $\mathrm{Hom}(V, W)$ such that $T(\alpha_j) = \Sigma_{i=1}^{m} t_{ij} \beta_i$ for $j = 1, \ldots, n$. Then the mapping $\{t_{ij}\} \longmapsto T$ is an isomorphism from $\mathbb{R}^{\bar{m} \times \bar{n}}$ to $\mathrm{Hom}(V, W)$.

<u>Proof:</u> We simply compose the isomorphism $\{t_{ij}\} \longmapsto \bar{T}$ of the above theorem with the isomorphism $\bar{T} \longmapsto \psi \circ \bar{T} \circ \varphi^{-1}$ from $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ to $\mathrm{Hom}(V, W)$, where $\varphi : \underline{x} \longmapsto \Sigma_{j=1}^{n} x_j \alpha_j$ and $\psi : \underline{y} \longmapsto \Sigma_{i=1}^{m} y_i \beta_i$ are the two fixed basis isomorphisms. Then $T = \psi \circ \bar{T} \circ \varphi^{-1}$ is the transformation described in the theorem, for $T(\alpha_j) = \psi(\bar{T}(\varphi^{-1}(\alpha_j))) = \psi(\bar{T}(\delta^j)) = \psi(\underline{t}^j) = \Sigma_{i=1}^{m} t_{ij} \beta_j$. The map $\{t_{ij}\} \longmapsto T$ is the composition of two isomorphisms and so is an isomorphism.

It is helpful to keep in mind the following relationships between $\{t_{ij}\}$ and T.

LEMMA 19. In the situation of the above theorem, if $\xi = \Sigma_{j=1}^{n} x_j \alpha_j$ and $\eta = \Sigma_{i=1}^{m} y_i \beta_i$ then $\eta = T(\xi)$ if and only if $y_i = \Sigma_{j=1}^{n} t_{ij} x_j$, $i = 1, \ldots, m$.

The matrix element t_{ij} can be obtained from T by the formula $t_{kj} = \mu_k(T(\alpha_j))$, where μ_k is the k^{th} element of the dual basis of W^*.

<u>Proof:</u> Rewriting the equations $y_i = \Sigma_{j=1}^{n} t_{ij} x_j$ as

$$\underline{y} = \Sigma_{j=1}^{n} x_j \underline{t}^j = \Sigma_{j=1}^{n} x_j \bar{T}(\delta^j) = \bar{T}(\underline{x}),$$

we see that they hold if and only if

$$\eta = \psi(\underline{y}) = \psi \circ \bar{T}(\underline{x}) = \psi \circ \bar{T} \circ \varphi^{-1}(\xi) = T(\xi).$$

Next, $\mu_k(T(\alpha_j)) = \mu_k(\Sigma_{i=1}^{m} t_{ij}\beta_i) = \Sigma_i t_{ij}\mu_k(\beta_i) = \Sigma_i t_{ij}\delta_i^k = t_{kj}$, completing the proof of the lemma.

We call $\{t_{ij}\}$ the matrix for T with respect to the given ordered bases for V and W.

B. The transpose.

DEFINITION. The <u>transpose</u> of the $m \times n$ matrix $\{t_{ij}\}$ is the $n \times m$ matrix $\{t^*_{ji}\}$ defined by $t^*_{ji} = t_{ij}$, for all i, j.

The rows of $\underline{t}^*$ are of course the columns of $\underline{t}$, and vice versa.

THEOREM 34. The matrix of T^* with respect to the dual bases in W^* and V^* is the transpose of the matrix of T.

<u>Proof:</u> Using the last assertion of the Lemmas 18 and 19, and we have

$$t^*_{ji} = \alpha^{**}_j(T^*(\mu_i)) = \alpha^{**}_j(\mu_i \circ T)$$

$$= (\mu_i \circ T)(\alpha_j) = \mu_i(T(\alpha_j)) = t_{ij}.$$

DEFINITION. The <u>row space</u> of the matrix $\{t_{ij}\} \in \mathbb{R}^{\bar{m} \times \bar{n}}$ is the subspace of $\mathbb{R}^n$ spanned by the m row vectors. The <u>column space</u> is, similarly, the span of the n column vectors in $\mathbb{R}^m$.

COROLLARY. The row and column spaces of a matrix have the same dimension.

Proof: If T is the element of $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ defined by $T(\delta^j) = \underline{t}^j$ then the set $\{\underline{t}^j\}_1^n$ of column vectors in the matrix $\{t_{ij}\}$ is the image under T of the standard basis of $\mathbb{R}^n$, and so its span, which we have called the column space of the matrix, is exactly the range of T. In particular, the dimension of the column space is $d(R(T))$.

Since the matrix of T^* is the transpose t^* of the matrix t, we have similarly that $d(R(T^*))$ is the dimension of the column space of t^*. But the column space of t^* is the row space of t, and the assertion of the corollary is thus reduced to the identity $d(R(T)) = d(R(T^*))$ from the corollary to Theorem 30.

DEFINITION. This common dimension is called the rank of the matrix.

Finally we relate matrix multiplication to transposition.

THEOREM 35. If the product st is defined then so is $t^* s^*$ and $t^* s^* = (st)^*$.

Proof: A direct calculation is easy. We have

$$(st)^*_{jk} = (st)_{kj} = \Sigma_{i=1}^{m} s_{ki} t_{ij} = \Sigma_{i=1}^{m} t^*_{ji} s^*_{ik} = (t^* s^*)_{jk} .$$

Thus $(st)^* = t^* s^*$, as asserted.

This identity is clearly the matrix form of the transformation identity $(S \circ T)^* = T^* \circ S^*$, and it can be deduced from the latter identity if desired.

C. Matrix products.

If $\bar{T} \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ and $\bar{S} \in \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^\ell)$ then of course $\bar{S} \circ \bar{T} \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^\ell)$ and it certainly should be possible to calculate its matrix from the matrices s and t of $\bar{S}$ and $\bar{T}$ respectively. To make this computation, we set $\underline{y} = \bar{T}(\underline{x})$ and $\underline{z} = \bar{S}(\underline{y})$ so that

$\underline{z} = (\bar{S} \circ \bar{T})(\underline{x})$. In terms of the matrices t and s we have

$$y_i = \Sigma_{j=1}^{n} t_{ij} x_j \quad \text{and} \quad z_k = \Sigma_{i=1}^{m} s_{ki} y_i$$

so that

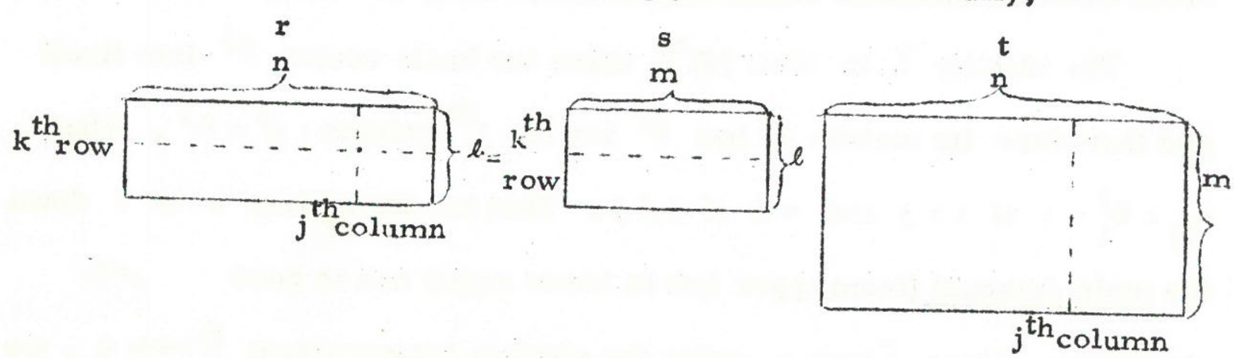
$$z_k = \Sigma_{i=1}^{m} s_{ki} \Sigma_{j=1}^{n} t_{ij} x_j = \Sigma_{j=1}^{n} \left(\Sigma_{i=1}^{m} s_{ki} t_{ij} \right) x_j \; .$$

That is, $z_k = \Sigma_{j=1}^{n} r_{kj} x_j$ for $k = 1, \ldots, \ell$ if we set

$$r_{kj} = \Sigma_{i=1}^{m} s_{ki} t_{ij} \quad \text{for all } k \text{ and } j \; .$$

We thus have found the formula for the matrix r of the map $\bar{R} = \bar{S} \circ \bar{T} : \underline{x} \longrightarrow \underline{z}$. Of course, r is defined to be the product of the matrices s and t and we write $r = s \cdot t$ or $r = st$.

Notice that in order for the product st to be defined the number of columns in the left factor must equal the number of rows in the right factor. We get the element r_{kj} by going across the k^{th} row of s and simultaneously down the j^{th} column of t, multiplying corresponding elements as we go, adding the resulting products. Pictorially,



Since we have defined the product of two matrices as the matrix of the product of the corresponding transformations, i.e., so that the mapping $\bar{T} \longmapsto \{t_{ij}\}$ preserves products : $\bar{S} \circ \bar{T} \longmapsto st$, it follows from the general principle of Lemma 5 that the algebraic laws satisfied by composition of transformations will automatically hold for the product of matrices. For example, we know without making an explicit computation that matrix multiplication is associative. And for square matrices we have the following theorem.

THEOREM 36. The set M_n of square $n \times n$ matrices is an algebra naturally isomorphic to the algebra $\mathrm{Hom}(\mathbb{R}^n)$.

Proof: We already know that $\bar{T} \longmapsto \{t_{ij}\}$ is a natural linear isomorphism from $\mathrm{Hom}(\mathbb{R}^n)$ to M_n (Theorem 32) and we have defined the product of matrices so that the mapping also preserves multiplication. The laws of algebra (for an algebra) therefore follow for M_n from our observation in Theorem 3 that they hold for $\mathrm{Hom}(\mathbb{R}^n)$, q.e.d.

The identity $\bar{I}$ in $\mathrm{Hom}(\mathbb{R}^n)$ takes the basis vector δ^j into itself and therefore its matrix e has δ^j for its j^{th} column : $\underline{e}^j = \delta^j$. Thus $e_{ij} = \delta^j_i = 1$ if $i = j$ and $= 0$ if $i \neq j$. That is, the matrix e is 1 down the main diagonal (from upper left to lower right) and is zero elsewhere. Since $\bar{I} \longmapsto e$ under the algebra isomorphism $\bar{T} \longmapsto t$, we know that e is the identity for matrix multiplication. Of course, we can check this directly : $\Sigma_{j=1}^{n} t_{ij} r_{jk} = t_{ik}$, and similarly on the left.

The symbol "e" is ambiguous in that we have used it to denote the identity in the space $\mathbb{R}^{\bar{n}\times\bar{n}}$ of square $n\times n$ matrices for any n.

COROLLARY. A square $n\times n$ matrix t has a multiplicative inverse if and only if its rank is n.

<u>Proof:</u> By the theorem there exists $s \in M_n$ such that $st = ts = e$ if and only if there exists $\bar{S} \in \mathrm{Hom}(\mathbb{R}^n)$ such that $\bar{S} \circ \bar{T} = \bar{T} \circ \bar{S} = I$. But such an $\bar{S}$ exists if and only if $\bar{T}$ is an isomorphism, and by the corollary to Theorem 25 this is equivalent to the dimension of the range of T being n. But this dimension is the rank of t, and the argument is complete.

THEOREM 36. If $\{\alpha_i\}_1^n$, $\{\beta_j\}_1^m$ and $\{\gamma_k\}_1^\ell$ are ordered bases for the vector spaces U, V and W respectively, and if $T \in \mathrm{Hom}(U,V)$ and $S \in \mathrm{Hom}(V,W)$, then the matrix for $S \circ T$ is the product of the matrices for S and T (with respect to the given bases).

<u>Proof:</u> By definition, the matrix for $S \circ T$ is the matrix of $\overline{S \circ T} = \chi^{-1} \circ (S \circ T) \circ \varphi \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^\ell)$, where $\varphi : \underline{x} \longmapsto \Sigma_1^n x_i \alpha_i$ and $\chi : \underline{z} \longmapsto \Sigma_1^\ell z_k \gamma_k$ are the given basis isomorphisms for U and W. But if ψ is the basis isomorphism for V, we have $\overline{S \circ T} = (\chi^{-1} \circ S \circ \psi) \circ (\psi^{-1} \circ T \circ \varphi) = \bar{S} \circ \bar{T}$, and therefore its matrix is the product of the matrices of

$\bar{S}$ and $\bar{T}$ by the definition of matrix multiplication. The latter are the matrices for S and T with respect to the given bases. Putting these observations together we have the theorem.

D. Cartesian vectors as matrices.

We can view an n-tuple $\underline{x} = < x_1, \ldots, x_n >$ as being alternatively either an $n \times 1$ matrix, in which case we call it a column vector, or a $1 \times n$ matrix, in which case we call it a row vector. Of course these identifications are natural isomorphisms. The point to doing this is, in part, that then the equations $y_i = \Sigma_{j=1}^{n} t_{ij} x_j$ say exactly that the column vector $\underline{y}$ is the matrix product of t and the column vector $\underline{x}$: $\underline{y} = t \cdot \underline{x}$. The linear map $\bar{T} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ becomes left multiplication by the fixed matrix t.

In particular, a linear functional $F \in (\mathbb{R}^n)^*$ becomes left multiplication by its matrix f which is of course $1 \times n$, and therefore is simply the row matrix interpretation of an n-tuple $\underline{f} \in \mathbb{R}^n$. That is, in the natural isomorphism $\underline{a} \longmapsto f_{\underline{a}}$ from $\mathbb{R}^n$ to $(\mathbb{R}^n)^*$, where $f_{\underline{a}}(\underline{x}) = \Sigma_1^n a_i x_i$, the functional $f_{\underline{a}}$ can now be interpreted as left matrix multiplication by the n-tuple $\underline{a}$ viewed as a row vector. The matrix product of the row vector ($1 \times n$ matrix) $\underline{a}$ and the column vector ($n \times 1$ matrix) $\underline{x}$ is a $|x|$ matrix $\underline{a} \cdot \underline{x}$, i.e. a number.

We shall take the column vector as the standard matrix interpretation of an n-tuple $\underline{x}$; then $\underline{x}^*$ is the corresponding row vector.

The number $f_{\underline{a}}(\bar{T}(\underline{x}))$ is thus the 1×1 matrix $\underline{a}^* t \underline{x}$. Since $f_{\underline{a}}(\bar{T}(\underline{x})) = (\bar{T}^*(f_{\underline{a}}))(\underline{x}) = g_{\underline{b}}(\underline{x})$, where $g_{\underline{b}} = \bar{T}^*(f_{\underline{a}}) \in (\mathbb{R}^m)^*$ we have the matrix identity $\underline{a}^* t \underline{x} = \underline{b}^* \underline{x}$ for all $\underline{x}$, and so

$$\underline{b} = t^* \underline{a}$$

$$\underline{b}^* = \underline{a}^* t .$$

$$(t^* a)^* = a^* t$$

That is, if we use the isomorphism of $(\mathbb{R}^n)^*$ with $\mathbb{R}^n$ to represent functionals by n-tuples viewed as row vectors, then the matrix of $\bar{T}^*$ is right multiplication by t. This only repeats something we already know, for the transpose of the above equation is $\underline{b} = t^* \underline{a}$, and since the n-tuples $\underline{b}$ and $\underline{a}$ are the coordinates of the functional $f_{\underline{a}}$ and $g_{\underline{b}}$ with respect to the standard bases in $(\mathbb{R}^n)^*$ and $(\mathbb{R}^m)^*$, this equation simply shows again that the matrix of T^* is t^*.

$\varphi \in \mathrm{Hom}(\mathbb{R}^n, V)$
$x \in \mathbb{R}^n$
$\{\beta_i\}$ is a basis for V
$\xi \in V$

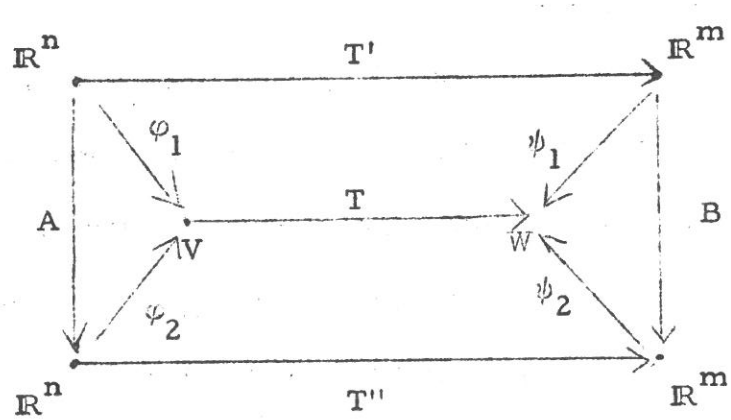
E. Change of basis.

If $\varphi : \underline{x} \longmapsto \xi = \Sigma_1^n x_i \beta_i$ and $\theta : \underline{y} \longmapsto \xi = \Sigma_1^n y_i \beta_i'$ are two basis isomorphisms for V, then $A = \theta^{-1} \circ \varphi$ is the isomorphism in $\mathrm{Hom}(\mathbb{R}^n)$ taking the coordinate n-tuple $\underline{x}$ of a vector ξ with respect to the basis $\{\beta_i\}$ into the coordinate n-tuple $\underline{y}$ of the same vector with respect to the basis $\{\beta_i'\}$. A is called the "change of coordinates" isomorphism.

The change of coordinate map $A = \theta^{-1} \circ \varphi$ should not be confused with the similar looking $T = \varphi \circ \theta^{-1}$. The latter is a mapping on V, and is the element of $\mathrm{Hom}(V)$ which takes each β_i to β_i'.

maps vectors with same coordinates onto one another

We now want to see what happens to the matrix of a transformation $T \in \mathrm{Hom}(V, W)$ when we change bases in its domain and codomain spaces. Suppose then that φ_1 and φ_2 are basis isomorphisms from $\mathbb{R}^n$ to V, that ψ_1 and ψ_2 are basis isomorphisms from $\mathbb{R}^m$ to W, and that t' and t'' are the matrices of T with respect to the first and second bases respectively. That is, t' is the matrix of $T' = (\psi_1)^{-1} \circ T \circ \varphi_1 \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$, and similarly for t''. The mapping $A = \varphi_2^{-1} \circ \varphi_1 \in \mathrm{Hom}(\mathbb{R}^n)$ is the change-of-coordinates transformation in V: if $\underline{x}$ is the coordinate n-tuple of a vector ξ with respect to the first basis (i.e., $\xi = \varphi_1(\underline{x})$) then $A(\underline{x})$ is its coordinate n-tuple with respect to the second basis. Similarly let B be the change of coordinates map $\psi_2^{-1} \circ \psi_1$ for W. The following diagram will help keep the various relationships of these spaces and mappings straight. We say that the diagram



is commutative, meaning that any two paths between two points represent the same map. By selecting various pairs of paths we can read off all the identities holding between the nine maps $T, T', T'', \varphi_1, \varphi_2, A, \psi_1, \psi_2, B$. For example, T'' can be obtained by going backward along A, forward along T' and then forward along B. That is, $T'' = B \circ T' \circ A^{-1}$. Since these "outside maps" are all maps of Cartesian spaces we can then read off the corresponding matrix identity,

$$t'' = bt' a^{-1} ,$$

showing how the matrix for T with respect to the second pair of bases is obtained from its matrix with respect to the first pair.

What we have actually done in reading off the above identity from the diagram is to eliminate certain retraced steps in the longer path which the definition would give us. Thus from the definitions we get

$$\begin{aligned} B \circ T' \circ A^{-1} &= (\psi_2^{-1} \circ \psi_1) \circ (\psi_1^{-1} \circ T \circ \varphi_1) \circ (\varphi_1^{-1} \circ \varphi_2) \\ &= \psi_2^{-1} \circ T \circ \varphi_2 = T'' . \end{aligned}$$

In the above situation the domain and codomain spaces were different and the two basis changes were independent of each other. If $W = V$, so that $T \in \mathrm{Hom}(V)$, then of course there is only one basis change and the formula becomes

$$\underline{t}'' = \underline{a} \cdot \underline{t}' \cdot \underline{a}^{-1} .$$

§ 10. Computations

The computational process by which the reader learned to solve systems of linear equations in secondary school algebra was undoubtedly "elimination by successive substitutions". The first equation is solved for the first unknown and the solution expression is substituted for the first unknown in the remaining equations, thereby eliminating the first unknown from the remaining equations. Next, the second unknown is solved for in the second equation and then eliminated from the remaining equations. In this way the unknowns are eliminated one at a time and a solution is obtained.

It turns out that this same procedure also solves the following additional problems :

(1) to obtain an explicit basis for the linear span of a set of m vectors in $\mathbb{R}^n$, and therefore, in particular,

(2) to find the dimension of such a subspace ;

(3) to compute the determinant of an $m \times m$ matrix ;

(4) to compute the inverse of an invertible $m \times m$ matrix .

In this section we shall briefly study this process and the solutions to these problems.

We start by noticing that the kinds of changes we are going to make on a finite sequence of vectors do not alter its span.

LEMMA 20. Let $\{\alpha_i\}_1^m$ be any m-tuple of vectors in a vector space and let $\{\beta_i\}_1^m$ be obtained from $\{\alpha_i\}_1^m$ by any one of the following elementary operations :

(1) interchanging two vectors ;

(2) multiplying some α_i by a non-zero scalar ;

(3) replacing α_i by $\alpha_i - x\alpha_j$ for some $j \neq i$ and some $x \in \mathbb{R}$.

Then

$$L(\{\beta_i\}_1^m) = L(\{\alpha_i\}_1^m) .$$

<u>Proof:</u> If $\alpha_i' = \alpha_i - x\alpha_j$ then $\alpha_i = \alpha_i' + x\alpha_j$. Thus if $\{\beta_i\}_1^m$ is obtained from $\{\alpha_i\}_1^m$ by one operation of type (3) then $\{\alpha_i\}_1^m$ can be obtained from $\{\beta_i\}_1^m$ by one operation of type (3) . In particular, each sequence is in the linear span of the other and the two linear spans are therefore the same.

Similarly each of the other operations can be undone by one of the same type and the linear spans are unchanged.

When we perform these operations on the sequence of <u>row</u> vectors in a matrix we call them <u>elementary row operations</u>.

We define the <u>order</u> of an n-tuple $\underline{x} = \langle x_1, \ldots, x_n \rangle$ as the index of the first non-zero entry. Thus if $x_i = 0$ for $i < j$ and $x_j \neq 0$ then the order of $\underline{x}$ is j .

Let $\{a_{ij}\}$ be an $m \times n$ matrix, let V be its row space, and let $n_1 < n_2 < \ldots < n_k$ be the integers that occur as orders of non-zero vectors in V . We are going to construct a basis for V consisting of k elements having exactly the above set of orders.

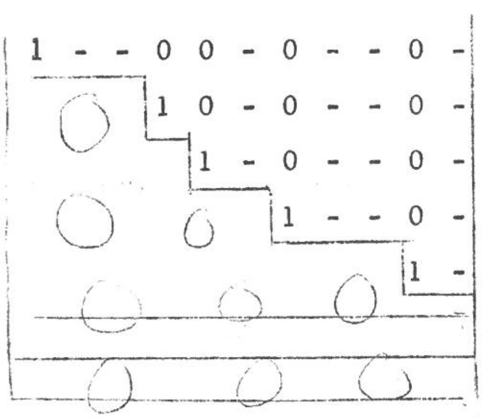
If every non-zero row in $\{a_{ij}\}$ has order $> p$ then each non-zero vector $\underline{x}$ in V has order $> p$, since $\underline{x}$ is a linear combination of these row vectors. Since some vector in V has the minimal order n_1, it follows that some row in $\{a_{ij}\}$ has order n_1. We move such a row to the top by interchanging two rows and then multiply this row by a constant so that its first non-zero entry x_{n_1} is 1. Let $\underline{a}^1, \ldots, \underline{a}^n$ be the row vectors that we now have, so that $\underline{a}^1$ has order n_1 and $a^1_{n_1} = 1$. We next modify each of the remaining rows, replacing $\underline{a}^j$ by $\underline{a}^j - a^j_{n_1} \cdot \underline{a}^1$, so that the new j^{th} row has 0 as its n_1 coordinate. The matrix that we thus obtain has the property that its j^{th} column is the zero m-tuple for each $j < n_1$ and its $n_1{}^{th}$ column is δ^1 in $\mathbb{R}^m$. Its first row has order n_1 and every other row has order $> n_1$. Its row space is still V.

Now let $\underline{x} = \Sigma_1^m c_i \underline{a}^i$ be a vector in V with order n_2. Then $c_1 = 0$, for if $c_1 \neq 0$ then the order of $\underline{x}$ is n_1. Thus $\underline{x}$ is a linear combination of the second to the m^{th} rows, and, just as in the first case, one of these rows must therefore have order n_2.

We now repeat the above process all over again, keying now on this vector. We bring it to the second row, make its n_2 coordinate 1, and subtract multiples of it from all the other rows (including the first) so that the resulting matrix has δ^2 for its $n_2{}^{th}$ column. Next we find a row with order n_3, bring it to the third row and make the $n_3{}^{th}$ column δ^3, etc.

We end up with an $m \times n$ matrix having the same row space V and the following special structure: (1) For $1 \leq j \leq k$ the j^{th} row has order n_j; (2) if $k < m$ the remaining $m-k$ rows are zero (since a non-zero row would have order $> n_k$, a contradiction) ; (3) the n_j^{th} column is δ^j . It follows that any linear combination of the first k rows with coefficients $c_1, \ldots, c_k$ has c_j in the n_j^{th} place , and hence cannot be zero unless all the c_j's are 0 . These k rows thus form a basis for V , solving problems (1) and (2) .

Our final matrix is said to be in row-reduced echelon form . It can be shown to be uniquely determined by the space V and the above requirements relating its rows to the orders of the elements of V . Its rows form the canonical basis of V . We sketch a typical row-reduced echelon matrix below .



This matrix is 8×11 , its orders are 1, 4, 5, 7, 10 and its row space has dimension 5. It is entirely 0 below the broken line. The dashes in the first 5 lines represent arbitrary numbers, but any change in these remaining entries change the spanned space V .

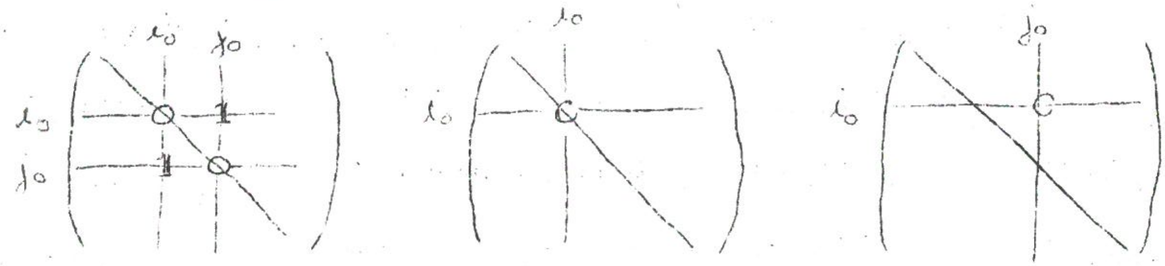
We shall now look for the significance of the row reduction operations from the point of view of general linear theory. In this discussion it will be convenient to use the fact that if an n-tuplet in $\mathbb{R}^n$ is viewed as an $n \times 1$ matrix (i.e., as a column vector) then the system of linear equations $y_i = \Sigma_{j=1}^n a_{ij}x_j$, $i = 1, \ldots, m$ expresses exactly the single matrix equation $\underline{y} = \underline{a} \cdot \underline{x}$. Thus the associated linear transformation $A \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ is now viewed as being simply multiplication by the matrix $\underline{a}$; $\underline{y} = A(\underline{x}) \Longleftrightarrow \underline{y} = \underline{a} \cdot \underline{x}$.

We first notice that each of our elementary row operations on an $m \times n$ matrix $\underline{a}$ is equivalent to premultiplication by a corresponding $m \times m$ elementary matrix $\underline{e}$. Supposing for the moment that this is so, we can find out what the matrix $\underline{e}$ is by using the $m \times m$ identity matrix $\underline{i}_m$. Since $\underline{e} \cdot \underline{a} = (\underline{e} \cdot \underline{i}_m) \cdot \underline{a}$, we see that the result of performing the operation on $\underline{a}$ can also be obtained by premultiplying $\underline{a}$ by the matrix $\underline{e} \cdot \underline{i}_m$. That is, if the elementary operation can be obtained as matrix multiplication by $\underline{e}$, then the multiplier is $\underline{e} \cdot \underline{i}_m$. This argument suggests that we should perform the operation on i_m and then see if premultiplying $\underline{a}$ by the resulting matrix performs the operation on $\underline{a}$.

If the elementary operation is interchanging the i_o^{th} and j_o^{th} rows, then performing it on i_m gives the matrix $\underline{e}$ with $e_{kk} = 1$ for $k \neq i_o$ and $k \neq j_o$, $e_{i_o j_o} = e_{j_o i_o} = 1$ and $e_{k\ell} = 0$ for all other indices. Moreover, examination of the sums defining the elements of the product matrix $\underline{e} \cdot \underline{a}$ will show that premultiplying by this $\underline{e}$ does just interchange the i_o^{th} and j_o^{th} rows of any $m \times n$ matrix $\underline{a}$.

In the same way, multiplying the i_o^{th} row of $\underline{a}$ by c is equivalent to premultiplying by the matrix $\underline{e}$ which is the same as $\underline{i}_m$ except for having $e_{i_o i_o} = c$. Finally, multiplying the j_o^{th} row by c and adding it to the i_o^{th} row is equivalent to premultiplying by the matrix $\underline{e}$ which is the identity $\underline{i}_m$ except that $e_{i_o j_o} = c$ instead of 0 .

These three elementary matrices are indicated schematically below. Each has the value 1 on the main diagonal and 0 off the main diagonal except as indicated.



These elementary matrices $\underline{e}$ are all non-singular (invertible) . The row interchange matrix is its own inverse. The inverse of multiplying the j^{th} row by c is multiplying the same row by 1/c . And the inverse of adding c times the j^{th} row to the i^{th} row is adding -c times the j^{th} row to the i^{th} row.

If $\underline{e}^1, \underline{e}^2, \ldots, \underline{e}^p$ is a sequence of elementary matrices, and if $\underline{b} = \underline{e}^p \cdot \underline{e}^{p-1} \cdot \ldots \cdot \underline{e}^1$ then $\underline{b} \cdot \underline{a}$ is the matrix obtained from $\underline{a}$ by performing the corresponding sequence of elementary row operations on $\underline{a}$. If $\underline{e}^1, \ldots, \underline{e}^p$ is a sequence which row reduces $\underline{a}$ then $\underline{r} = \underline{b} \cdot \underline{a}$ is the resulting row reduced echelon matrix.

Now suppose that $\underline{a}$ is a square $m \times m$ matrix and is non-singular (i.e., invertible) . Thus the dimension of the row space is m , and hence there are m different orders $n_1, \ldots, n_k$. That is, k = m , and since $1 \leqq n_1 < n_2 < \ldots < n_m = m$, we must also have $n_i = i$,

$i = 1, \ldots, m$. Remembering that the n_i^{th} column in $\underline{r}$ is δ^i, we see that now the i^{th} column in $\underline{r}$ is δ^i and therefore that $\underline{r}$ is simply the identity matrix $\underline{i}_m$. Thus $\underline{b} \cdot \underline{a} = \underline{i}_m$ and $\underline{b}$ is the inverse of $\underline{a}$. Finally, since $\underline{b} \cdot \underline{i}_m = \underline{b}$ we see that we get $\underline{b}$ from $\underline{i}_m$ by applying the same row operations (gathered together as premultiplication by $\underline{b}$) that we used to reduce $\underline{a}$ to echelon form. This is probably the best way of computing the inverse of a matrix. To keep track of the operations we can place $\underline{i}_m$ to the right of $\underline{a}$ to form a single $m \times 2m$ matrix $(\underline{a}\ \underline{i}_m)$ and then row reduce it. In echelon form it will then be the $m \times 2m$ matrix $(\underline{i}_m \underline{b})$ and we can read off the inverse $\underline{b}$ of the original matrix $\underline{a}$.

Finally we consider the problem of computing the determinant of a square $m \times m$ matrix. Since this computation is being made before the reader knows what the determinant function really is, he will temporarily have to accept on faith the correctness of the procedure.

We use two elementary operations (one modified) as follows :

(1') interchanging two vectors and simultaneously changing the sign of one of them ;

(2) as before, replacing some row α_i by $\alpha_i - x\alpha_j$ for some $j \neq i$.

When applied to the rows of a square matrix these operations leave the determinant unchanged.

Consider, then, a square $m \times m$ matrix $\{a_{ij}\}$. We interchange the first and p^{th} row to bring a row of minimal order n_1 to the top, and change the sign of the row being moved down (the first row here) . We do not make the leading coefficient of the new first row 1 ; this elementary operation is not being used now. We do subtract multiples of the first row from the remaining rows so as to make all of their entries in the n_1^{th} column 0 . The n_1^{th} column is now $c_1\delta^1$, where c_1 is the leading coefficient in the first row. And the new matrix has the same determinant as the original matrix.

We continue as before subject to the above modifications. We change the sign of a row moved downward in an interchange, we do not make leading coefficients 1 , and we do clear out the n_j^{th} column so that it becomes $c_j\delta^{n_j}$, where c_j is the leading coefficient of the j^{th} row ($1 \leqq j \leqq k$) . As before the remaining m-k rows are 0 (if $k < m$) . Let us call this matrix semi-reduced. Notice that from it we can find the corresponding reduced echelon matrix by k application of (2) ; we simply multiply the j^{th} row by $1/c_j$ for $j = 1,\ldots,k$. If $\underline{s}$ is the semi-reduced matrix which we obtained from $\underline{a}$ using (1') and (3) , then its determinant and therefore the determinant also of $\underline{a}$, is the product of the entries on the main diagonal, $\Pi_{i=1}^{m} s_{ii}$. Recapitulating we compute the determinant of a square matrix $\underline{a}$ by using the operations (1') and (3) to change $\underline{a}$ to a semi-reduced matrix $\underline{s}$, and then take the product of the numbers on the main diagonal of $\underline{s}$.

If the original matrix $\{a_{ij}\}$ is non-singular, so that $k = m$ and $n_i = i$ ($i = 1,\ldots,m$) , then the j^{th} column in the semi-reduced matrix is $c_j \delta^j$ so that $s_{jj} = e_j$ and the determinant is the product $\pi^m_{i=1} c_i$ of the leading coefficients. This is non-zero. On the other hand, if $\{a_{ij}\}$ is singular, so that $k = d(V) < m$, then the m^{th} row in the semi-reduced matrix is 0 and, in particular, $s_{mm} = 0$. The determinant of a singular matrix is therefore 0 . We have thus found that a matrix is non-singular (invertible) if and only if its determinant is non-zero.

Chapter II

NORMS

§ 1. Review in $\mathbb{R}$.

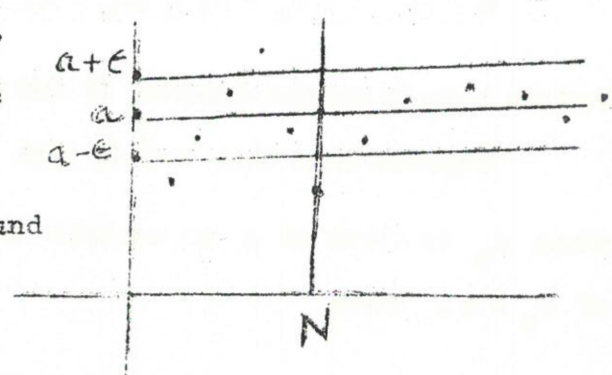
Every student of the calculus is presumed to be familiar with the properties of the real number system and the theory of limits. But more is needed at this point. It is absolutely essential that by now the student understand the ϵ-definitions and be able to work with them. To be on the safe side we shall review some of this material; the confident reader can skip it.

The definition of sequential convergence contains three quantifiers, as follows:

<u>Definition.</u> $x_n \longrightarrow a$ as $n \longrightarrow \infty \Longleftrightarrow (\forall \epsilon > 0)(\exists N)(\forall n)(n > N \Longrightarrow |x_n - a| < \epsilon)$.

The arrow '$\longrightarrow$' is a verbal symbol which is read "tends to" or "converges to". The three prefixing quantifiers makes the definition sound artificial and unidiomatic when read as ordinary prose, but the reader will remember from our introductory discussion of quantification that this artificiality is absolutely necessary in order that the meaning of the sentence be clear and unambiguous. <u>Any change in the order $(\forall \epsilon)(\exists N)(\forall n)$ changes the meaning of the statement.</u>

The meaning of the inner universal quantification $(\forall n)(n > N \Longrightarrow |x_n - a| < \epsilon)$ is intuitive and easily pictured graphically. For all n beyond N the numbers x_n are closer to a than ϵ. The definition begins by saying that such an N can be found for each ϵ. N will of course vary with



ϵ; if ϵ is made smaller we will generally have to go further out in the sequence, i.e., take N larger, before all succeeding terms become ϵ close to a.

This definition is used in various ways. In the simplest situations we are given one or more convergent sequences, say $x_n \longrightarrow a$ and $y_n \longrightarrow b$ (in informal notation), and we want to prove that some other sequence is convergent, say $z_n \longrightarrow c$.

In such cases we always try to find an inequality expressing the quantity which we wish to make small, $|z_n - c|$, in terms of the quantities which we know can be made small, $|x_n - a|$ and $|y_n - b|$. For example, suppose that $z_n = x_n + y_n$. Since x_n is close to a and y_n is close to b, z_n is clearly close to $a + b$, but how close? Setting $c = a + b$ we have $z_n - c = (x_n - a) + (y_n - b)$, and so

$$|z_n - c| \leqq |x_n - a| + |y_n - b|$$

From this it is clear that in order to make $|z_n - c| < \epsilon$ it is sufficient to make each of $|x_n - a|$ and $|y_n - b| < \epsilon/2$. Therefore, given any positive ϵ, we can take N_1 so that $n > N_1 \Rightarrow |x_n - a| < \epsilon/2$ and N_2 so that $n > N_2 \Rightarrow |y_n - b| < \epsilon$, and then take N as the larger of these two numbers, so that if $n > N$ then both inequalities hold. Thus

$$n > N \Rightarrow |z_n - c| \leqq |x_n + a| + |y_n - b| + \epsilon/2 + \epsilon/2 = \epsilon$$

and we have found the desired N for the z_n sequence.

Suppose next that $a \neq 0$, and $z_n = \frac{1}{x_n}$. Clearly $\frac{1}{x_n}$ is close to $\frac{1}{a}$ when x_n is close to a so we take $c = \frac{1}{a}$ and try to express $z_n - c$ in terms of $x_n - a$. Thus

$$z_n - c = \frac{1}{x_n} - \frac{1}{a} = \frac{a - x_n}{x_n a} \text{ and } |z_n - c| \leq \frac{|a - x_n|}{|x_n a|} .$$

The trouble here is that the denominator is variable, and if it should happen to be very small it might cancel the smallness of $|x_n - a|$ and not give a small $|z_n - c|$. But the answer to this problem is easy. Since x_n is close to a and a is not zero, x_n cannot be close to 0. For instance, if x_n is closer to a than $|a|/2$ then x_n must be farther from 0 than $|a|/2$.

We therefore choose N_1 so that $n > N_1 \Longrightarrow |x_n - a| < |a|/2$, from which it follows that $|x_n| > |a|/2$. Then $|z_n - c| < 2|x_n - a|/|a|^2$, and now, given any ϵ, we take N_2 so that $n > N_2 \Longrightarrow |x_n - a| < \epsilon |a|^2/2$. Again taking N as the larger of N_1 and N_2, so that both inequalities will hold simultaneously when $n > N$, we have

$$n > N \Longrightarrow |z_n - c| < 2|x_n - a|/|a|^2 < 2\epsilon |a|^2/2|a|^2 = \epsilon,$$

and again we have found our N for the z_n sequence.

We have tried to show above how one would think about this situation. The actual proof that would be written down would only show the choice of N. Thus:

Lemma 1. If $x_n \longrightarrow a$ and $y_n \longrightarrow b$ then $x_n + y_n \longrightarrow a + b$.

Proof. Given $\epsilon (> 0)$, choose N_1 so that $n > N_1 \Longrightarrow |x_n - a| < \epsilon/2$ (by the assumed convergence of $\{x_n\}$ to a), and, similarly, choose N_2 so that $n > N_2 \Longrightarrow |y_n - b| < \epsilon/2$. Take N as the larger of N_1 and N_2. Then $n > N \Longrightarrow |(x_n + y_n) - (a + b)| \leqq |x_n - a| + |y_n - b| < \epsilon/2 + \epsilon/2 = \epsilon$. Thus we have proved that $(\forall \epsilon^{>0})(\exists N)(\forall n)(n > N) \Longrightarrow |(x_n + y_n) - (x+y)| < \epsilon)$, and we are done.

An even more formal procedure would be to write down the hypotheses explicitly in terms of other ϵ's. Thus, we are given that

$$(\forall \epsilon_1)(\exists N_1)(\forall n)(n > N_1) \Longrightarrow |x_n - a| < \epsilon_1)$$

$$(\forall \epsilon_2)(\exists N_2)(\forall n)(n > N_2) \Longrightarrow |y_n - b| < \epsilon_2)$$

and we want to show that

$$(\forall \epsilon)(\exists N)(\forall n)(n > N \Longrightarrow |x_n + y_n) - (a + b)| < \epsilon.$$

Since $|(x_n + y_n) - (a + b)| \leqq |x_n - a| + |y_n - b|$, we take $\epsilon_1 = \epsilon_2 = \epsilon/2$ and $N = \max(N_1, N_2)$. That is, we start with ϵ, then define ϵ_1 and ϵ_2 suitably in terms of ϵ, and then N suitably in terms of the N_1 and N_2 given for ϵ_1 and ϵ_2 by the hypothesis.

The given proof of the lemma is perfectly adequate for this course.

Besides ϵ techniques in limit theory, it is necessary to understand and be able to use the equivalent fundamental properties of the real number system called the least upper bound property and the compactness of closed intervals. We shall not discuss the equivalence of these properties here; we shall simply assume them both.

The semi-infinite interval $(-\infty, a]$ is of course the subset $\{x \in \mathbb{R} : x \leqq a\}$

(B) If A is a non-empty subset of $\mathbb{R}$ such that $A \subset (-\infty, a]$ for some a then there exists a uniquely determined smallest number b such that $A \subset (-\infty, b]$.

A number a such that $A \subset (-\infty, a]$ is called an upper bound to A; clearly a is an upper bound to $A \Longleftrightarrow (\forall x \in A)(x \leqq a)$. A set having an upper bound is said to bounded above. The principle (B) says that a non-empty set A which is bounded above has a least upper bound.

If we reverse the order relation by multiplying everything by -1 then we have an alternate formulation of (B) which asserts that a non-empty set which is bounded below has a greatest lower bound.

It is true, but not particularly easy to prove, that (B) is also equivalent to (C) below.

(C) Every bounded sequence of real numbers has a convergent subsequence.

A subsequence of a sequence $\{x_n\}$ is a new sequence formed by selecting an infinite number, but generally not all, of the terms x_n, and counting them off in the order of the selected indices.

Thus, if n_1 is the first selected n, n_2 the next, and so on, we obtain the subsequence $\{x_{n_1}, x_{n_2}, \ldots, x_{n_m}, \ldots\}$ or $\{x_{n_m}\}_m$. Strictly speaking, this counting off of the selected set of indices n is a mapping $m \longmapsto n_m$ of $\mathbb{Z}^+$ into $\mathbb{Z}^+$ which preserve order: $n_{m+1} > n_m$ for all m. And the subsequence $m \longmapsto x_{n_m}$, is the composition of the sequence $n \longmapsto x_n$ with the selector mapping.

In order to avoid subscripts on subscripts we may use the notation $n(m)$ instead of n_m. In either case we are being conventionally sloppy: we are using the same symbol 'n' as an integer-valued variable, when we write x_n, and for the selector function, when we write $n(m)$ or n_m. This is one of the standard notational ambiguities to which we submit in elementary calculus because the cure is considered worse than the disease. We could say: let f be a sequence, i.e., a function on $\mathbb{Z}^+$ into $\mathbb{R}$. Then a subsequence of f is a composition $f \circ g$ where g is a function on $\mathbb{Z}^+$ into $\mathbb{Z}^+$ such that $g(m+1) > g(m)$ for all m.

In any case, we have the following more explicit statement of (C).

(C') If $\{x_n\}$ is any sequence into an interval $[a, b] \subset \mathbb{R}$, then there exists a subsequence $\{x_{n(m)}\}_m$ and a number ℓ such that $x_{n(m)} \longrightarrow \ell$ as $m \longrightarrow \infty$.

§2. Norms

In the limit theory of $\mathbb{R}$, as reviewed briefly above, the absolute value function is used prominently in expressions like '$|x - y|$' to designate the distance between two numbers, here between x and y. The definition of the convergence of x_n to a is simply a careful statement of what it means to say that the distance $|x_n - a|$ tends to zero.

The limit theory of vector spaces is studied in terms of functions called <u>norms</u>, which serve as multidimensional analogues of the absolute value function on $\mathbb{R}$. An example of a norm on $\mathbb{R}^n$ is the function $p(\underline{x}) = \max \{|x_i| : i \in \overline{n}\}$ where, of course, $\underline{x} = \langle x_1, x_2, \ldots, x_n \rangle$.

Definition. A <u>norm</u> is a real-valued function p on a vector space V such that

(n1) $p(\alpha) > 0$ if $\alpha \neq 0$ (positivity)

(n2) $p(x\alpha) = |x| \, p(\alpha)$ for all $\alpha \in V$, $x \in \mathbb{R}$. (homogeneity)

(n3) $p(\alpha + \beta) \leq p(\alpha) + p(\beta)$ for all $\alpha, \beta \in V$. (triangle inequality)

A <u>normed linear space</u> (nls), or <u>normed vector space</u>, is a vector space V together with a norm p on V. A normed linear space is thus really a <u>pair</u> $\langle V, p \rangle$, but generally we speak simply of the nls V, a definite norm on V then being understood.

It has been customary to designate the norm of α by $\|\alpha\|$, presumably to suggest the analogy with absolute value. The triangle inequality (n3) then becomes: $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|$, which is almost identical in form with the basic absolute value inequality: $|x + y| \leq |x| + |y|$. Similarly (n2) becomes

$\|x\alpha\| = |x|\,\|\alpha\|$, like $|xy| = |x|\,|y|$ in $\mathbb{R}$. Furthermore, $\|\alpha - \beta\|$ is similarly interpreted as the distance between α and β. Therefore, if we set $\alpha = \xi - \eta$ and $\beta = \eta - \zeta$, (n3) becomes the usual triangle inequality of geometry:

$$\|\xi - \zeta\| \leqq \|\xi - \eta\| + \|\eta - \zeta\|.$$

We shall use both the double bar notation and the "p" notation for norms, each being on occasion superior to the other.

Other common norms on $\mathbb{R}^n$ are $\|\underline{x}\|_1 = \sum_1^n |x_i|$ and $\|\underline{x}\|_2 = \left(\sum_1^n x_i^2\right)^{1/2}$

The norm already mentioned is designated $\|\underline{x}\|_\infty$. Thus $\|\underline{x}\|_\infty = \max\{|x_i|\}_1^n$.

Similar norms on the infinite dimensional vector space $\mathcal{C}^0([a, b])$ are $\|f\|_1 = \int_a^b |f(t)|\,dt$, $\|f\|_2 = \left(\int_a^b |f(t)|^2 dt\right)^{1/2}$ and $\|f\|_\infty = \max\{|f(t)| : 0 \leqq t \leqq 1\}$.

The reader should work through the proofs of (n1) - (n3) for $\|\ \|_1$ and $\|\ \|_\infty$. The proofs for the 2-norm $\|\ \|_2$ are trickier and can be postponed for a while.

Uniform norms. The two norms $\|\ \|_\infty$ considered above are special cases of a very general phenomenon. Let A be an arbitrary non-empty set and let $\mathcal{B}(A, \mathbb{R})$ be the set of all bounded functions $f : A \longrightarrow \mathbb{R}$. That is, $f \in \mathcal{B}(A, \mathbb{R}) \Longleftrightarrow f \in \mathbb{R}^A$ and range f is a bounded subset of $\mathbb{R}$. Then $\mathcal{B}(A, \mathbb{R})$ is a vector space V, since if b and c are bounds for $|f|$ and $|g|$ respectively then $|xf + yg|$ is bounded by $|x|b + |y|c$. The uniform norm on V is defined by

$$\|f\|_\infty = \text{lub}\,\{|f(p)| : p \in A\}$$

That is, $\|f\|_\infty$ is defined as the smallest bound to $|f|$. Of course, it has to be checked that $\| \ \|_\infty$ is a norm. For any $p \in A$, $|f(p) + g(p)| \leqq |f(p)| + |g(p)| \leqq \|f\|_\infty + \|g\|_\infty$. Thus $\|f\|_\infty + \|g\|_\infty$ is a bound to $|f + g|$ and is therefore $\geqq$ the smallest such bound, which is $\|f + g\|_\infty$. This gives the triangle inequality. Next we notice that for any non-empty set A of real numbers and any $x \geqq 0$, lub $(xA) = x$ lub A. It follows that $\|xf\|_\infty = \text{lub}(\text{range } |xf|) = \text{lub}(|x| \text{ range } |f|) = |x| \text{ lub}(\text{range } |f|) = |x| \|f\|_\infty$. Finally, $\|f\|_\infty \geqq 0$ and $\|f\|_\infty = 0 \Longrightarrow f$ is the zero function.

As usual we can replace $\mathbb{R}$ by any normed linear space W and define the uniform norm on $\mathcal{B}(A, W)$ by $\|f\|_\infty = \text{lub}\{\|f(p)\| : p \in A\}$.

If $f \in \mathcal{C}([0, 1])$ then we know that the continuous function $|f|$ assumes the least upper bound of its range as a value, so that then $\|f\|_\infty$ is the maximum value of $|f|$. In general, however, the definition must be given in terms of lub.

<u>Sequential convergenc.</u>. Let V be any normal linear space. If $\{\alpha_n\} \subset V$ and $\alpha \in V$ we are clearly going to want the sentence '$\alpha_n \longrightarrow \alpha$ as $n \longrightarrow \infty$' to mean that the distance between α_n and α tends to zero. But this distance is $\|\alpha_n - \alpha\|$. Thus:

Definition. $\alpha_n \longrightarrow \alpha$ as $n \longrightarrow \infty \Longleftrightarrow \|\alpha_n - \alpha\| \longrightarrow 0$ as $n \longrightarrow \infty$.

If we substitute for the defining sentence <u>its</u> definition, we have:

$$\alpha_n \longrightarrow \alpha \text{ as } n \longrightarrow \infty \Longleftrightarrow (\forall \epsilon^{>0})(\exists N)(\forall n)(n > N \Longrightarrow \|\alpha_n - \alpha\| < \epsilon).$$

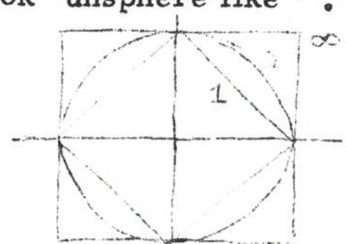
We shall generally work with the latter characterization, and thus involve ourselves directly with ϵ calculations. However, the first definition can sometimes be used to transfer convergence questions from V back to $\mathbb{R}$.

Theorem 1. If $\alpha_n \longrightarrow \alpha$ and $\beta_n \longrightarrow \beta$ then $\alpha_n + \beta_n \longrightarrow \alpha + \beta$. If $\alpha_n \longrightarrow \alpha$ in V and $x_n \longrightarrow x$ in $\mathbb{R}$ then $x_n\alpha_n \longrightarrow x\alpha$ in V.

Proof. If the direct ϵ, N definition is used for norm convergence, the proofs of these two statements are obtained, word for word, from the corresponding proofs for real sequences by replacing absolute value signs by the norm symbol.

Remembering that $\|\alpha-\xi\|$ is interpreted as the distance from α to ξ it is natural to define the <u>open sphere</u> (or <u>ball) of radius</u> r <u>about the center</u> α as $\{\xi: \|\alpha-\xi\| < r\}$. We designate this sphere $S_r(\alpha)$. Translation through β ought to carry $S_r(\alpha)$ into $S_r(\alpha+\beta)$, and we can check that it does, for $T_\beta[S_r(\alpha)] = \{\xi+\beta : \xi \in S_r(\alpha)\} = \{\eta : \eta - \beta \in S_r(\alpha)\} = \{\eta : \eta \in S_r(\alpha+\beta)\}$. Also, multiplication by the scalar c ought to take a sphere of radius r into one of radius cr, and the reader can check that indeed $c\, S_r(\alpha) = S_{cr}(c\alpha)$.

Although $S_r(\alpha)$ behaves like a sphere the actual set being defined is different for different norms, and some of them "look" "unsphere-like". The unit spheres about the origin in $\mathbb{R}^2$ for the three norms $\| \;\|_1$, $\| \;\|_2$ and $\| \;\|_\infty$ are sketched at the right.



A subset A of a nls V is <u>bounded</u> if it lies in some sphere, say $S_r(\alpha)$. Then it also lies in a sphere about the origin, namely $S_{r+\|\alpha\|}(0)$. This is simply the fact that if $\|\xi-\alpha\| < r$ then $\|\xi\| < r + \|\alpha\|$, which we get from the triangle inequality upon rewriting $\|\xi\|$ as $\|(\xi-\alpha)+\alpha\|$.

The radius of the largest sphere about a vector β which does not touch a set A is naturally called the <u>distance from</u> β to A. It is clearly $\text{glb}\{\|\xi-\beta\| : \xi \in A\}$.

§3. Continuity.

Let V and W be any two normed linear spaces. We shall designate both norms by $\| \ \|$. This ambiguous usage does not cause confusion. It is like the ambiguous use of "0" for the zero elements of all the vector spaces under consideration.

Let A be a subset of V and let f be a mapping of A into W.

<u>Definition.</u> The function f is <u>continuous</u> at a point $\alpha \in A \Leftrightarrow$

$$(\forall \epsilon)(\exists \delta)(\forall \xi \in A)(\|\xi - \alpha\| < \delta \Rightarrow \|f(\xi) - f(\alpha)\| < \epsilon).$$

Here, again, the general norm symbol instead of the absolute value sign is all that distinguishes this vector space definition from the standard real variable definition.

<u>Definition.</u> The function f is continuous $\Longleftrightarrow$ f is continuous at each point of its domain.

The following theorem characterizes continuity in terms of sequential convergence, and helps greatly to use the notion of continuity in a flexible way.

<u>Theorem 2.</u> The function f is continuous at $\alpha \Longleftrightarrow$ for every sequence $\{\xi_n\}$ in A, if $\xi_n \longrightarrow \alpha$ then $f(\xi_n) \longrightarrow f(\alpha)$.

Proof. We first prove that continuity implies the sequential condition. Our hypothesis, then, is that given ϵ there exists a δ such that $\|\xi - \alpha\| < \delta \Rightarrow \|f(\xi) - f(\alpha)\| < \epsilon$. If $\xi_n \longrightarrow \alpha$ then there exists N such that $n > N \Rightarrow \|\xi_n - \alpha\| < \delta$. Combining these implications we have that $n > N \Rightarrow \|f(\xi_n) - f(\alpha)\| < \epsilon$. Thus $f(\xi_n) \longrightarrow f(\alpha)$.

Now suppose that f is not continuous at α, i.e., that $(\exists\epsilon)(\forall\delta)(\exists\xi)(\|\xi-\alpha\| < \delta$ and $\|f(\xi) - f(\alpha)\| \geqq \epsilon)$. Take $\delta = 1/n$ and let ξ_n be a corresponding ξ, so that $\|\xi_n - \alpha\| < 1/n$ and $\|f(\xi_n) - f(\alpha)\| \geqq \epsilon$ for all n. The first inequality shows that $\xi_n \longrightarrow \alpha$ and the second that $f(\xi_n) \not\longrightarrow f(\alpha)$. Thus, if f is not continuous at α then the sequential condition is not satisfied. This completes the proof of the theorem.

The above type of argument is used very frequently and amounts almost to an automatic proof procedure in the relevant situations. We want to prove that $(\forall x)(\exists y)(\forall z)P(x, y, z)$. Arguing by contradiction, we suppose this false, so that $(\exists x)(\forall y)(\exists z) \sim P(x, y, z)$. Then, instead of trying to use all numbers y, we let y run through some sequence converging to zero, such as $\{1/n\}$, and choose one corresponding z, z_n, for each such y. We end up with $\sim P(x, 1/n, z_n)$ for the given x and all n, and we finish off by arguing sequentially.

For a linear map $T : V \longrightarrow W$ continuity has a simpler characterization as boundedness.

Definition. A linear mapping $T : V \longrightarrow W$ is bounded $\Longleftrightarrow$ $(\exists C^{>0})(\forall\xi^{\in V})(\|T(\xi)\| \leqq C\|\xi\|)$.

Any such C is called a bound of T.

It should immediately be pointed out that this is not the same notion of boundedness we discussed earlier. There we called a real-valued function bounded if its range was a bounded subset of $\mathbb{R}$. The analogue here would be to call a vector-valued function bounded it its range is norm bounded. But a non-zero linear transformation cannot be bounded in this sense because $\|T(x\alpha)\| = |x| \, \|T(\alpha)\|$. The above definition amounts to the boundedness in the earlier sense of the quotient $T(\alpha) / \|\alpha\|$ (on $V - \{0\}$).

Theorem 3. If T is a linear mapping of a nls V into a nls W then the following conditions are equivalent:

(1) T is continuous at one point;

(2) T is continuous;

(3) T is bounded.

Proof. (1) $\Longrightarrow$ (3). Suppose T is continuous at α_0. Then, taking $\epsilon = 1$, there exists δ such that $\|\alpha - \alpha_0\| < \delta \Longrightarrow \|T(\alpha) - T(\alpha_0)\| < 1$. Setting $\xi = \alpha - \alpha_0$ and using the additivity of T, we have: $\|\xi\| < \delta \Longrightarrow \|T(\xi)\| < 1$. Now for any non-zero η, $\xi = \delta\eta / 2\|\eta\|$ has norm $\delta/2$. Therefore $\|T(\xi)\| < 1$. But $\|T(\xi)\| = \delta\|T(\eta)\|/2\|\eta\|$, giving $\|T(\eta)\| < 2\|\eta\|/\delta$. Thus T is bounded by $C = 2/\delta$.

(3) $\Longrightarrow$ (2). Suppose $\|T(\xi)\| \leqq C\|\xi\|$ for all ξ. Then for any α_0 and any ϵ we can take $\delta = \epsilon/C$ and have $\|\alpha - \alpha_0\| < \delta \Longrightarrow \|T(\alpha) - T(\alpha_0)\| = \|T(\alpha - \alpha_0)\| \leqq C\|\alpha - \alpha_0\| < C\delta = \epsilon$.

(2) $\Longrightarrow$ (1). Trivial.

Lipschitz functions. Let V and W be normed linear spaces, A a subset of V and f a function on A into W. Then f is said to be a Lipschitz function $\Longleftrightarrow$ $(\exists C)(\forall \alpha, \beta \in A)(\|f(\alpha) - f(\beta)\| \leqq C\|\alpha - \beta\|)$.

Every bounded linear T is a Lipschitz mapping, and we prove in the lemma below that the norm function is also a Lipschitz function on V into $\mathbb{R}$.

Lemma 2. For all $\alpha, \beta \in V$, $|\|\alpha\| - \|\beta\|| \leqq \|\alpha - \beta\|$.

Proof. We have $\|\alpha\| = \|(\alpha - \beta) + \beta\| \leqq \|\alpha - \beta\| + \|\beta\|$ so that

$\|\alpha\| - \|\beta\| \leqq \|\alpha - \beta\|$. Similarly, $\|\beta\| - \|\alpha\| \leqq \|\beta - \alpha\| = \|\alpha - \beta\|$. This pair of inequalities is equivalent to the lemma.

Other Lipschitz mappings will appear when we study mappings with continuous differentials. Roughly speaking, the Lipschitz property lies between continuity and continuous differentiability, and it is frequently the condition that we actually apply under the hypothesis of continuous differentiability.

§4. The space of bounded linear transformations

If V and W are normed linear spaces then Hom(V, W) is redefined to be the set of all bounded linear maps $T : V \longrightarrow W$. The results of Thm. 2, Ch all remain true, but require some further arguing.

Theorem 4. Hom(V, W) is itself a normed linear space if $\|T\|$ is defined as the smallest bound for T.

Proof. By definition $\|T\|$ is the smallest number C such that $\|T(\alpha)\| \leqq C\|\alpha\|$ for all $\alpha \in V$. We can then check the triangle inequality and homogeneity in much the same way we did for uniform norms in §2. For example, $\|(S+T)(\alpha)\| \leqq \|S(\alpha)\| + \|T(\alpha)\| \leqq \|S\| \cdot \|\alpha\| + \|T\| \cdot \|\alpha\| = (\|S\| + \|T\|)\|\alpha\|$ Thus $\|S\| + \|T\|$ is a bound for $S + T$ and is therefore $\geqq$ the smallest such bound, $\|S + T\|$.

Other characterizations of $\|T\|$ are useful. It is the smallest C which is $\geqq \|T(\alpha)\|/\|\alpha\|$ for all $\alpha \neq 0$ in V and hence,

$$\|T\| = \text{lub}\left\{\|T(\alpha)\|/\|\alpha\| : \alpha \in V \text{ and } \alpha \neq 0\right\}$$

Also $\|T(\alpha)\|/\|\alpha\| = \|T(\alpha/\|\alpha\|)\|$ by homogeneity, and $\left\|\alpha/\|\alpha\|\right\| = 1$. Therefore

$$\|T\| = \text{lub}\left\{\|T(\alpha)\| : \alpha \in V \text{ and } \|\alpha\| = 1\right\}.$$

Finally, we can define $\|T\|$ as a uniform norm. For this, it is convenient to let p be the norm on V and then notice that $T/p \in \mathcal{B}(V - \{0\}, W)$ and

$$\|T\| = \|T/p\|_{\infty}$$

The proof that $\|T\|$ is a norm is now unnecessary by virtue of §2.

Theorem 5. If U, V, W are normed linear spaces, and if $T \in \mathrm{Hom}(U, V)$ and $S \in \mathrm{Hom}(V, W)$, then $S \circ T \in \mathrm{Hom}(U, W)$ and $\|S \circ T\| \leqq \|S\| \, \|T\|$. It follows that composition on the right by a fixed T is a bounded linear transformation of $\mathrm{Hom}(V, W)$ into $\mathrm{Hom}(U, W)$, and similarly for convolution on the left by a fixed S.

Proof. $\|(S \circ T)(\alpha)\| = \|S(T(\alpha))\| \leqq \|S\| \, \|T(\alpha)\| \leqq \|S\| (\|T\| \, \|\alpha\|) = (\|S\| \cdot \|T\|)(\|\alpha\|)$. Thus $S \circ T$ is bounded by $\|S\| \cdot \|T\|$ and everything else follows at once.

As before, the conjugate space V^* is $\mathrm{Hom}(V, \mathbb{R})$, now the space of all bounded linear functionals.

§5. Equivalent norms.

Definition. Two norms p and q on the same vector space V are equivalent $\Leftrightarrow$ there exist, constants a and b such that $p \leqq aq$ and $q \leqq bp$.

Then $(1/b)q \leqq p \leqq aq$ and $(1/a)p \leqq q \leqq bp$, so that two norms are equivalent $\Leftrightarrow$ either can be bracketted by two multiples of the other. The above definition simply says that the identity map on V, considered as a map from the normal linear space $\langle V, p \rangle$ to the normed linear space $\langle V, q \rangle$ is bounded in both directions. Theorems 2 and 3 then imply :

Theorem 6. Two norms on V are equivalent if and only if they define exactly the same convergent sequences in V.

If V is infinite dimensional two norms will in general not be equivalent. For example, if $V = \mathcal{C}^0([0, 1])$ and $f_n(t) = t^n$ then $\|f_n\|_1 = 1/n+1$ and $\|f_n\|_\infty = 1$, so that $f_n \longrightarrow 0$ in the 1-norm but not in the uniform norm. The theorem above then implies that $\| \ \|_1$ and $\| \ \|_\infty$ are not equivalent. This is why the very notion of a normed linear space depends on the assumption of a given norm.

§6. Norms on finite dimensional spaces.

We shall devote the present section to proving that if V is finite dimensional then all norms on V are equivalent. We shall also see that then every linear transformation is continuous, and that $\alpha_n \longrightarrow \alpha \iff \ell(\alpha_n) \longrightarrow \ell(\alpha)$ in $\mathbb{R}$ for every $\ell \in V^*$. It follows that the study of limits and continuity over a finite dimensional vector space is really independent of norm considerations. However, norms are exceedingly handy even here, and the fact that all norms are equivalent means that in a given situation one can use whatever norm seems most suitable.

If $\theta : V \longrightarrow W$ is an isomorphism and q is a norm on W it is easy to see that $p = q \circ \theta$ is a norm on V. If $\theta : V \longrightarrow \mathbb{R}^n$ is a coordinate isomorphism and q is the 1-norm on $\mathbb{R}^n$ we shall call p the 1-norm on V associated with the basis defining θ.

Lemma 3. A sequence $\{\underline{x}^i\}$ in $\mathbb{R}^n$ converges to $\underline{x}$ in the 1-norm $\iff$ each coordinate sequence $\{x^i_j\}_i$ converges to x_j in $\mathbb{R}$, $j = 1, \ldots, n$.

Proof. Since $\|\underline{x}^i - \underline{x}\|_1 = \sum_{j=1}^n |x^i_j - x_j|$ the proof follows from Lemma 1.

Theorem 7. If $p = \|\ \|_1$ with respect to some basis in V then every p-bounded sequence in V has a p-convergent subsequence.

Proof. We may as well suppose that V is $\mathbb{R}^n$. The proof is by induction on n. If $n = 1$, V is $\mathbb{R}$ and the theorem is just the fundamental principle (C) of §1. Suppose now that the theorem is true for dimensions $< n$, and let $\{\underline{x}^i\}$ be a bounded sequence in $\mathbb{R}^n$. Thinking of $\mathbb{R}^n$ as $\mathbb{R}^{n-1} \times \mathbb{R}$, we have $\underline{x}^i = \langle \underline{y}^i, z_i \rangle$ where y^i is the (n-1)-tuplet composed of the first n-1 coordinates of $\underline{x}^i$ and z_i is the n^{th} coordinate of $\underline{x}^i$. By the inductive hypothesis, a subsequence $\{\underline{y}^{i(j)}\}_j$ is convergent, say to $\underline{y}_0$. By the principle (C) for $\mathbb{R}$ the bounded sequence $\{z_{i(j)}\}$ has a subsequence $\{z_{i(j(k))}\}_k$ which converges, say to z_0. Then $y^{i(j(k))} \xrightarrow[k]{} \underline{y}_0$ and so $\underline{x}^{i(j(k))} \xrightarrow[k]{} \underline{x}_0 = \langle y_0, z_0 \rangle$ by the above lemma.

Theorem 8. On a finite dimensional vector space any two norms are equivalent.

Proof. It is sufficient to prove that an arbitrary norm $\| \ \|$ is equivalent to the 1-norm $\| \ \|_1$ associated with a basis $\{\beta_i\}_1^n$. Setting $b = \max\{\|\beta_i\| : 1 \leq i \leq n\}$ we have $\|\xi\| = \|\sum_1^n \epsilon_i(\xi)\beta_i\| \leq \sum_1^n |\epsilon_i(\xi)| \|\beta_i\| \leq b\|\xi\|_1$.

Conversely, we assert that there exists $a > 0$ such that $\|\xi\|_1 \leq a\|\xi\|$. Otherwise $(\forall a^{>0})(\exists \xi)(\|\xi\|_1 > a\|\xi\|)$, and if we take $a = m$, with corresponding ξ_m, we have $\|\xi_m\|_1 > m\|\xi_m\|$ for all m. Setting $\zeta_m = \xi_m/\|\xi_m\|_1$, we have $\|\zeta_m\|_1 = 1$ and $\|\zeta_m\| < 1/m$. Since $\{\zeta_m\}$ is thus bounded in the 1-norm it has by the above theorem a subsequence $\{\zeta_{m(i)}\}$ which converges in the 1-norm, say to ζ, and since the 1-norm is continuous with respect to its own convergence, we have $\|\zeta\|_1 = \lim_{i\to\infty}\|\zeta_{m(i)}\|_1 = 1$. But from $\|\zeta_{m(i)}\| < 1/m(i) \leq 1/i$ we have $\zeta_{m(i)} \longrightarrow 0$, and from $\|\zeta_{m(i)} - \zeta\| \leq b\|\zeta_{m(i)} - \zeta\|_1$ we have $\zeta_{m(i)} \longrightarrow \zeta$, both with respect to the other norm. Therefore $\zeta = 0$, contradicting $\|\zeta\|_1 = 1$.

Theorem 9. If V and W are finite dimensional then every linear $T: V \longrightarrow W$ is bounded. Thus Hom(V, W) is the same no matter how V and W are regarded.

Proof. Choose bases in V and W and let the matrix of T be $\{t_{ij}\}$.

Let ξ in V have coordinates $\{x_j\}_1^n$, let η in W have coordinates $\{y_i\}_1^m$, and suppose that $\eta = T(\xi)$. Then $\|\eta\|_1 = \sum_1^m |y_i| = \sum_{i=1}^m \left| \sum_{j=1}^n t_{ij}x_j \right| \leqq \sum_{i,j} |t_{ij}| \, \|\xi\|_\infty$. Thus T is bounded by the 1-norm of its matrix when the uniform norm is used in V and the 1-norm in W.

Theorem 10. If V is finite dimensional then $\xi_n \longrightarrow \xi$ in V $\Longleftrightarrow$ $(\forall \ell \in V^*)(\ell(\xi_n) \longrightarrow \ell(\xi)$ in $\mathbb{R})$.

Proof. $\Longleftarrow$. If $\{\beta_i\}_1^n$ is a basis then $\epsilon_i(\xi_n) \longrightarrow \epsilon_i(\xi)$ for each i and $\|\xi_n - \xi\|_1 \longrightarrow 0$ by the lemma.

$\Longrightarrow$. If $\xi_n \longrightarrow \xi$ and $\ell \in V^*$ then $\ell(\xi_n) \longrightarrow \ell(\xi)$ by the continuity of ℓ.

Remark. If V is an arbitrary normed linear space, so that $V^* = \mathrm{Hom}(V, \mathbb{R})$ is the set of bounded linear functionals, then we say that $\xi_n \longrightarrow \xi$ weakly $\Leftrightarrow$ $\ell(\xi_n) \longrightarrow \ell(\xi)$ for each $\ell \in V^*$. The above theorem can therefore be rephrased to say that in a finite dimensional space weak convergence and norm convergence are equivalent notions.

§7. Product norms.

We now ask what kind of norm we might want on the Cartesian product $V = V_1 \times V_2$ of two normed linear spaces. A reasonable requirement is that a sequence should converge in V if and only if its two component sequences in V_1 and V_2 both converge. Now if p_i is the given norm on V_i, $i = 1, 2$, then p, defined by $p(\langle \alpha_1, \alpha_2 \rangle) = p_1(\alpha_1) + p_2(\alpha_2)$ clearly is a norm on V and has the required property. Then by §5 any other suitable norm on V must be equivalent to the above sum norm. Any such norm will be called a <u>product norm.</u> The reader can check that $q(\langle \alpha_1, \alpha_2 \rangle) = \max \{p_1(\alpha_1), p_2(\alpha_2)\}$ is another product norm, and so is $\| \langle \alpha_1, \alpha_2 \rangle \| = (p_1(\alpha_1)^2 + p_2(\alpha_2)^2)^{1/2}$.

Each of these three norms can be defined as well for n factor spaces as for two, and we gather the facts for this general case into a theorem.

<u>Theorem</u> 11. If $\{\langle V_i, p_i \rangle\}_1^n$ is a finite set of normed linear spaces then $\| \; \|_1$, $\| \; \|_2$ and $\| \; \|_\infty$ defined on $V = \prod_{i=1}^n V_i$ by $\|\alpha\|_1 = \sum_1^n p_i(\alpha_i)$, $\|\alpha\|_2 = (\sum_1^n p_i(\alpha_i)^2)^{1/2}$ and $\|\alpha\|_\infty = \max\{p_i(\alpha_i) : i = 1, \ldots, n\}$ are equivalent norms on V and each is a product norm in the sense that a sequence converges in V if and only if its n component sequences all converge in their corresponding factor spaces.

It looks above as though all we are doing is taking <u>any</u> norm $\| \; \|$ on $\mathbb{R}^n$ and then defining a norm $||| \; |||$ on the product space V by $|||\alpha||| = \| \langle p_1(\alpha_1), \ldots, p_n(\alpha_n) \rangle \|$.

This is almost correct. The interested reader will discover, however, that $\| \; \|$ on $\mathbb{R}^n$ must have the property that $\underline{x} \leq \underline{y} \Longrightarrow \|\underline{x}\| \leq \|\underline{y}\|$ for the

?↗

$x_i \le y_i \; \forall i$

triangle inequality to follow for $|||\ \ |||$ in V. If we call such a norm on $\mathbb{R}^n$ an <u>increasing</u> norm, then it is true that:

If $\|\ \ \|$ is any increasing norm on $\mathbb{R}^n$ then $|||\alpha||| = \| \langle p_1(\alpha_1), \ldots, p_n(\alpha_n) \rangle$ is a product norm on $V = \prod_1^n V_i$. However, we shall use only the 1, 2, ∞ product norms in this course.

§8. Metric spaces; open and closed sets

In the preceding sections we have occasionally treated questions of convergence and continuity in situations where the domain was an arbitrary subset A of a nls V. The vector operations by themselves were never made use of in these discussions. Indeed, they could not be used because A is not a vector space. What was used was the combination $\|\alpha - \beta\|$, interpreted as the distance from α to β, together with the triangle inequality for this distance function. If we distil out of those contexts what was essential to the convergence and continuity arguments we end up with a space A and a function $\rho : A \times A \longrightarrow \mathbb{R}$, $\rho(x, y)$ being the distance from x to y, such that:

(1) $\rho(x, y) > 0$ if $x \neq y$, and $\rho(x, x) = 0$;

(2) $\rho(x, y) = \rho(y, x)$ for all $x, y \in A$;

(3) $\rho(x, z) \leqq \rho(x, y) + \rho(y, z)$ for all $x, y, z \in A$.

Any set A together with such a function ρ on $A \times A$ to $\mathbb{R}$ is called a metric space, the function ρ being the metric. It is obvious that any subset B of a metric space A is itself a metric space under $\rho \restriction B \times B$.

Moreover, metric spaces do very often arise as subsets of normed linear spaces, the metric being the restriction of the norm metric $\rho(\alpha, \beta) = \|\alpha - \beta\|$. But they come from other sources, too. And even in the normed linear space context, metrics other than the norm metric are used.

For example, S might be the two dimensional surface of an ordinary sphere in $\mathbb{R}^3$, say $S = \{\underline{x} : \sum_1^3 x_i^2 = 1\}$, and $\rho(\underline{x}, \underline{y})$ might be the great circle distance from $\underline{x}$ to $\underline{y}$. Or, more generally, S might be any smooth 2-dimensional surface in $\mathbb{R}^3$ and $\rho(\underline{x}, \underline{y})$ might be the length of the shortest curve connecting **x** to y in S.

For the rest of this chapter we shall adopt the metric space context for our arguments. We do this so that the student may become familiar with this more general but very intuitive notion. It is not accidental that the change over tak. place now. Up until this point a large part of our concern has been with linear matters, whereas the rest of the chapter is largely independent of linearity.

Although probably unnecessary, we begin by reproducing the basic definitions.

Definition. $x_n \longrightarrow x$ as $n \longrightarrow \infty \iff (\forall \epsilon^{>0})(\exists N)(\forall n)(n > N \implies \rho(x_n, x) < \epsilon)$.

Definition. If A and B are metric spaces, then $f : A \longrightarrow B$ is continuous at $a \in A \iff (\forall \epsilon^{>0})(\exists \delta^{>0})(\forall x^{\in A})(\rho(x, a) < \delta \implies \rho(f(x), f(a)) < \epsilon)$. The proof of the sequential characterization of continuity is correct as it stands (when $\|\alpha - \beta\|$ is replaced by $\rho(\alpha, \beta)$).

The (open) sphere of radius r about α, $S_r(\alpha)$, is simply the set of points whose distance from α is less than r:

$$S_r(\alpha) = \{\xi : \rho(\xi, \alpha) < r\}.$$

A subset $A \subset V$ is open $\iff (\forall \alpha \in A)(\exists r^{>0})(S_r(\alpha) \subset A)$. Thus A is open $\iff$ every point of A is the center of some sphere included in A.

Lemma 4. Every sphere is open; in fact, if $\beta \in S_r(\alpha)$ and $\delta = r - \rho(\alpha, \beta)$ then $S_\delta(\beta) \subset S_r(\alpha)$.

Proof. This amounts to the triangle inequality.

For, $\xi \in S_\delta(\beta) \implies \rho(\xi, \beta) < \delta \implies \rho(\xi, \alpha) \leq \rho(\xi, \beta) + \rho(\beta, \alpha) < \delta + \rho(\alpha, \beta) = r \implies \xi \in S_r(\alpha)$.

<u>Theorem 12</u>. The family $\mathcal{T}$ of all open sets has the following properties:

(1) the union of any collection of open sets is open; that is, $\mathcal{a} \subset \mathcal{T} \Rightarrow \cup \mathcal{a} \in \mathcal{T}$;

(2) the intersection of two open sets if open; that is, $A, B \in \mathcal{T} \Rightarrow A \cap B \in \mathcal{T}$;

(3) $\emptyset, V \in \mathcal{T}$.

Proof. These properties follow immediately from the definition. Thus, any point α in $\cup a$ lies in some $A \in \mathcal{Q}$ and therefore, since A is open, some sphere about α is a subset of A and hence of the larger set $\cup a$.

Corollary: A set is open $\Longleftrightarrow$ it is a union of spheres.

Proof. This follows from the definition of open set, the lemma above and (1) of the theorem.

The union of <u>all</u> the open subsets of an arbitrary set A is by (1) an open subset of A and therefore is the <u>largest</u> open subset of A. It is called the <u>interior</u> of A and is designated A^*. Clearly, $\alpha \in A^* \Longleftrightarrow (\exists r^{>0})(S_r(\alpha) \subset A)$. Also A is open $\Longleftrightarrow A = A^*$.

The definition of sequential convergence becomes perhaps more intuitive when it is rephrased in terms of spheres and the notion of something being true for almost all n. We say that $P(n)$ is true <u>for almost all</u> $n \Longleftrightarrow P(n)$ is true for all but a finite number of integers, or, equivalently, $(\exists N)(\forall n^{>N})P(n)$. Then the definition of convergence can be restated as follows:

$\alpha_n \longrightarrow \alpha$ as $n \longrightarrow \infty \Longleftrightarrow$ every sphere about α contains almost all the α_n.

<u>Definition.</u> A set A is <u>closed</u> $\Longleftrightarrow$ A' is open.

The theorem above and de Morgan's law then yield the following complementary set of properties for closed sets.

<u>Theorem 13</u> (1) The intersection of any family of closed sets is closed.

(2) The union of two closed sets is closed.

(3) $\emptyset$ and V are closed.

Continuing our "complementary" development, we define the <u>closure</u>, $\bar{A}$, of an arbitrary set A as the intersection of all closed sets including A, and have from (1) above that $\bar{A}$ <u>is the smallest closed set including</u> A and from de Morgan the important identity:

$$(\bar{A})' = (A')^*$$

This identity yields a direct characterization of closure:

<u>Lemma 5</u>. $\beta \in \bar{A} \Longleftrightarrow$ every sphere about β intersects A .

Proof. $\sim(\beta \in \bar{A}) \Longleftrightarrow \beta \in (A')^* \Longleftrightarrow (\exists r^{>0})(S_r(\beta) \subset A')$. Negating the extreme members of this equivalence gives the lemma.

Finally, this lemma in turn leads to a sequential characterization.

<u>Lemma 6</u>. $\beta \in \bar{A} \Longleftrightarrow$ there exists a sequence in A converging to β.

Proof. If $\{\alpha_n\} \subset A$ and $\alpha_n \longrightarrow \beta$ then every sphere about β contains almost all the α_n and hence, in particular, intersects A. Thus $\beta \in \bar{A}$ by the above lemma.

Conversely, if $\beta \in \bar{A}$ then every sphere about β intersects A. In particular, we can choose for each n one point $\alpha_n \in A \cap S_{1/n}(\beta)$ and so construct a sequence $\{\alpha_n\}$ in A which converge to β.

<u>Definition.</u> The <u>boundary</u>, ∂A, of an arbitrary set A is the difference between its closure and its interior. Thus

$$\partial A = \overline{A} - A^*$$

Since $A-B = A \cap B'$, we have that $\partial A = \overline{A} \cap (\overline{A'})$ and therefore that $\partial A = \partial(A')$. Also,

$$\beta \in \partial A \iff \text{every sphere about } \beta \text{ intersects both } A \text{ and } A'.$$

Example. A sphere $S_r(\alpha)$ is an open set. Its closure is the closed sphere about α of radius $r = \{\xi : \rho(\xi, \alpha) \leq r\}$. This is most easily seen by using the sequential characterization of a point in the closure. The boundary $\partial S_r(\alpha)$ is then the spherical surface of radius r about $\alpha = \{\xi : \rho(\xi, \alpha) = r\}$. If some but not all of the points of this surface are added to the open sphere we obtain a set that is neither open nor closed. The student should expect of a random set he may encounter that it will be neither open nor closed.

Continuous functions furnish an important source of closed sets by the following lemma .

Lemma 7. If V, W are metric spaces, if $f : A \longrightarrow W$ is a continuous function and if A is a closed subset of V, then $f^{-1}[B]$ is closed whenever B is closed.

Proof. If $\{\alpha_n\} \subset f^{-1}[B]$ and $\alpha_n \longrightarrow \alpha$, then $\alpha \in A$ since A is closed, $f(\alpha_n) \longrightarrow f(\alpha)$ since f is continuous, and $\{f(\alpha_n)\} \subset B \Longrightarrow f(\alpha) \in B$ since B is closed. Therefore $\alpha \in f^{-1}[B]$ and $f^{-1}[B]$ has been proved closed.

If $\text{dom}(f) = V$ in the above lemma then we can take complements and get the following corollary.

Corollary. If $f : V \longrightarrow W$ is continuous then $f^{-1}(A)$ is open in V whenever A is open in W.

This can also be argued directly from the ϵ, δ definition of continuity. The converse holds as well. As an example of the use of this lemma consider for a fixed $\alpha \in V$ the continuous function $f : V \longrightarrow \mathbb{R}$ defined by $f(\xi) = \rho(\xi, \alpha)$. The sets $(-r, r)$ $[0, r]$, $\{r\}$ are respectively open, closed and closed subsets of $\mathbb{R}$. Therefore their inverse images under f, the sphere $S_r(\alpha)$, the closed sphere $\{\xi : \rho(\xi, \alpha) \leqq r\}$ and the spherical surface $\{\xi : \rho(\xi, \alpha) = r\}$ are open, closed and closed in V. In particular, the triangle inequality argument demonstrating directly that $S_r(\alpha)$ is open is now seen to be unnecessary by virtue of the earlier triangle inequality argument which demonstrated the continuity of the norm function and which remains unchanged for a general metric function.

It is not true that continuous functions take closed sets into closed sets in the forward direction. For example, the sequence $\{2n\pi + 1/n\}$ is a closed subset of $\mathbb{R}$ (strictly speaking, has a closed range), but its image under the sine function is the sequence $\{\sin(1/n)\}$ which is not a closed set.

If A is not assumed closed in Lemma 7 then the conclusion is that the set $C = f^{-1}[B]$ is the intersection of a closed set with A; in fact, $C = \overline{C} \cap A$. For $C \subset \overline{C}$ and $C \subset A \Rightarrow C \subset \overline{C} \cap A$, and conversely, if $\alpha \in \overline{C} \cap A$ then $\exists \{\alpha_n\} \subset C$ with $\alpha_n \longrightarrow \alpha$ and we are now back to the beginning of the above proof. A subset of A that is the intersection of A wth a closed subset of V is said to be relatively closed in A. It's complement in A is relatively open in A and ts the intersection of A with an open subset of V.

§9. Topology

If V is an arbitrary set and $\mathcal{T}$ is any family of subsets of V satisfying the conditions (1) - (3) in Theorem 12 then $\mathcal{T}$ is called a topology on V. That theorem thus asserts that the open subsets of a normed linear space V form a topology on V. The subsequent definitions of interior, closed set and closure were purely topological in the sense that they depended only on having the topology $\mathcal{T}$, as was Theorem 13, and the identity $(\bar{A})' = (A')^*$. The study of the consequences of the existence of a topology is called general topology.

On the other hand, the definitions of sphere, sequential convergence and continuity given earlier were metric definitions, and therefore part of metric space theory. In metric spaces, then, we have not only the topology but also our ϵ definitions of sequential convergence, continuity and spheres, and the spherical characterizations of closure and interior.

The reader may be surprised to be told now that although continuity and convergence were defined metrically they also have purely topological characterizations and are therefore topological ideas. This change-over is easy to see if one keeps in mind that in a metric space an open set is nothing but a union of spheres. We have:

(a) $\alpha_n \longrightarrow \alpha$ as $n \longrightarrow \infty$ $\Longleftrightarrow$ every open set containing α contains almost all the α_n.

(b) f is continuous at α $\Longleftrightarrow$ for every open set A containing $f(\alpha)$ there exists an open set B containing α such that $f[B] \subset A$.

Furthermore, these local conditions involving behavior around a single point α are more fluently rendered in terms of the notion of neighborhood. A

Bourbaki

set A is a neighborhood of a point $\beta \Longleftrightarrow \beta \in A^*$. Then we have:

(a') $\alpha_n \longrightarrow \alpha$ as $n \longrightarrow \infty \Longleftrightarrow$ every neighborhood of α contains almost all α_n.

(b') If dom f is a neighborhood of α then f is continuous α $\Longleftrightarrow$ for every neighborhood N of $f(\alpha)$. $f^{-1}[N]$ is a neighborhood of α.

Finally there are elegant topological characterizations of global continuity. Suppose that S_1 and S_2 are topological spaces. Then $f : S_1 \rightarrow S_2$ is continuous (every where) $\Longleftrightarrow$ $f^{-1}[A]$ if open whenever A is open $\Longleftrightarrow$ $f^{-1}[B]$ is closed whenever B is closed. These conditions are not surprising in view of Lemma 7 .

§10. Sequential compactness

Definition. A metric space A is sequentially compact $\Longleftrightarrow$ every sequence in A has a subsequence which converges to a point of A.

Lemma 8. A sequentially compact set A is closed and bounded.

Proof. Suppose that $\{\alpha_n\} \subset A$ and that $\alpha_n \longrightarrow \beta$. By compactness there exists a subsequence $\{\alpha_{n(i)}\}_i$ converging to a point $\alpha \in A$. But $\alpha_n \longrightarrow \beta \Longrightarrow \alpha_{n(i)} \longrightarrow \beta$. Therefore $\beta = \alpha$ and $\beta \in A$. Thus A is closed.

Boundedness here will mean lying in some sphere about a given point β. If A is not bounded there exists $\alpha_n \in A$ such that $\rho(\alpha_n, \beta) > n$. By compactness a subsequence $\{\alpha_{n(i)}\}_i$ converges to a point $\alpha \in A$ and $\rho(\alpha_{n(i)}, \beta) \longrightarrow \rho(\alpha, \beta)$. This clearly contradicts $\rho(\alpha_{n(i)}, \beta) > n(i) \geqq i$.

Theorem 14 If V is a finite dimensional nls then every closed and bounded subset of V is sequentially compact.

Proof. This has essentially been proved already in §6. There we showed that if A is bounded then every sequence in A has a convergent subsequence. If A is also closed then the limit of this subsequence is in A, and we are done.

Sequential compactness in infinite dimensional spaces is a much rarer phenomenon, but is very important when it does occur. We shall study one such occurence in connection with Sturm-Liouville theory in Chapter IV.

Continuous functions carry compact sets into compact sets. The proof of the following result will be left as an exercise.

Theorem 15. If f is continuous and A is a sequentially compact subset of its domain, then f[A] is sequentially compact.

A non-empty compact set $A \subset \mathbb{R}$ contains maximum and minimum elements. This is because lub A is the limit of a sequence in A, and hence belongs to A itself since A is closed. Combining this fact with the above theorem we obtain the following well-known corollary.

Corollary. If f is a continuous, real-valued function and dom(f) is sequentially compact, then f is bounded and assumes maximum and minimum values.

The proof in §6 actually involved this circle of ideas. We first showed that bounded closed sets were compact in the 1-norm, and in particular, the surface S of the unit 1-norm sphere is compact. The inequality $\| \| \leq b \| \|_1$ implies that the second norm is continuous with respect to the 1-norm and therefore assumes a minimum value m on S. Since $\| \|$ cannot be zero on S we have $m > 0$, and therefore $\| \|_1 \leq (1/m) \| \|$ on S. By homogeneity the inequality holds everywhere.

The following very useful result is related to the above theorem.

Theorem 16. If f is continuous and one-to-one and if dom(f) is sequentially compact then f^{-1} is continuous.

Proof. We have to show that if $\beta_n \longrightarrow \beta$ in the range of f, and if $\alpha_n = f^{-1}(\beta_n)$ and $\alpha = f^{-1}(\beta)$, then $\alpha_n \longrightarrow \alpha$. We do so by showing that every subsequence $\{\alpha_{n(i)}\}_i$ has itself a subsequence converging to α (§). But, since dom(f) is compact, there is a subsequence $\{\alpha_{n(i(j))}\}_j$ converging to some γ, and the continuity of f implies that $f(\gamma) = \lim_{j \to \infty} f(\alpha_{n(i(j))}) = \lim_{j \to \infty} \beta_{n(i(j))} = \beta$. Therefore $\gamma = f^{-1}(\beta) = \alpha$, which is what we had to prove.

§11. Compactness and uniformity

A. The work 'uniform' is frequently used as a qualifying adjective in mathematics. Roughly speaking, it concerns a "point" property $P(\alpha)$ which may or may not hold at each point α in a domain A, and whose definition involves an existential quantifier. A typical form for $P(\alpha)$ is $(\forall c)(\exists d)\,Q(\alpha, c, d)$. The property holds on A if it holds for all $\alpha \in A$, i.e., if

$$(\forall \alpha^{\in A})(\forall c)(\exists d)\, Q(\alpha, c, d)\,.$$

Here d will in general depend on both α and c; if either α or c is changed the corresponding d may have to be changed. The property holds uniformly on A, or uniformly in α, if a value d can be found that is independent of α, though still depending on c. Thus the property holds uniformly in $\alpha \iff$

$$(\forall c)(\exists d)(\forall \alpha^{\in A})\, Q(\alpha, c, d)\,;$$

the uniformity of the property is expressed in the reversal of the order of the quantifiers $(\forall \alpha^{\in A})$ and $(\exists d)$.

According to the definition of continuity, a function $f : A \longrightarrow W$ is continuous on its domain A $\iff$

$$(\forall \alpha^{\in A})(\forall \epsilon^{>0})(\exists \delta^{>0})(\forall \beta^{\in A})(\rho(\alpha, \beta) < \delta \Rightarrow \rho(f(\alpha), f(\beta)) < \epsilon)$$

Then f is uniformly continuous on A $\iff$

$$(\forall \epsilon^{>0})(\exists \delta^{>0})(\forall \alpha, \beta^{\in A})(\rho(\alpha, \beta) < \delta \Rightarrow \rho(f(\alpha), f(\beta)) < \epsilon)$$

Now δ is independent of the point at which continuity is being asserted, though still dependent on ϵ , of course.

A sequence of functions $\{f_n\} \subset W^A$ converge to $f : A \longrightarrow W$ at a point $\alpha \in A \iff f_n(\alpha) \longrightarrow f(\alpha)$ in $W. \iff (\forall \epsilon^{>0})(\exists N)(\forall n)(n > N \implies \rho(f_n(\alpha), f(\alpha)) < \epsilon)$. The sequence converges pointwise to f if it converges to f at every point $\alpha \in A$, i.e., $\iff$

$$(\forall_\alpha{}^{\in A})(\forall \epsilon^{>0})(\exists N)(\forall n)(n > N \implies \rho(f_n(\alpha), f(\alpha)) < \epsilon.$$

The sequence converges <u>uniformly</u> on $A \iff N$ exists which is independent of α, i.e., $\iff$

$$(\forall \epsilon^{>0})(\exists N)(\forall n)(\forall \alpha)(n > N \implies \rho(f_n(\alpha), f(\alpha)) < \epsilon)$$

When $\rho(\alpha, \beta) = \|\alpha - \beta\|$, saying that $\rho(f_n(\alpha), f(\alpha)) < \epsilon$ for all α implies that $\|f_n - f\|_\infty \leqq \epsilon$. Thus $f_n \longrightarrow f$ uniformly $\iff \|f_n - f\|_\infty \longrightarrow 0$; this is why the norm $\|f\|_\infty$ is called the uniform norm.

Pointwise convergence does not imply uniform convergence. Thus $f_n(x) = x^n$ on $A = (0, 1)$ converges pointwise to the zero function but not uniformly.

Nor does continuity on A imply uniform continuity. The function $f(x) = 1/x$ is continuous on $(0, 1)$ but not uniformly continuous. The function $\sin(1/x)$ is continuous and bounded on $(0, 1)$ but is not uniformly continuous. Compactness changes this situation, however.

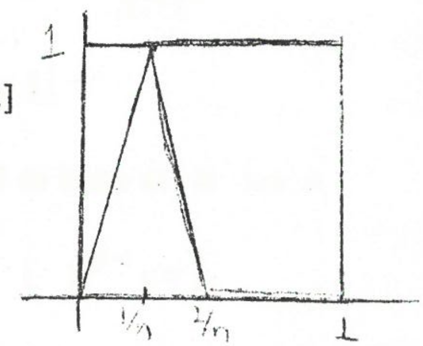
<u>Theorem 17</u>. If f is continuous on A and A is compact then f is uniformly continuous on A.

Proof. This is one of our "automatic" negation proofs. Uniform continuity (UC) is the property

$$(\forall \epsilon^{>0})(\exists \delta^{>0})(\forall \alpha, \beta^{\in A})(\rho(\alpha, \beta) < \delta \implies \rho(f(\alpha), f(\beta)) < \epsilon)$$

Thus $\sim$ UC $\iff (\exists \epsilon)(\forall \delta)(\exists \alpha, \beta)(\rho(\alpha, \beta) < \delta$ and $\rho(f(\alpha), f(\beta)) \geq \epsilon)$. Take $\delta = 1/n$, with corresponding α_n & β_n. Thus, for all n, $\rho(\alpha_n, \beta_n) < 1/n$ and $\rho(f(\alpha_n), f(\beta_n)) \geq \epsilon$, where ϵ is a fixed positive number. Now $\{\alpha_n\}$ has a convergent subsequence, say $\alpha_{n(i)} \longrightarrow \alpha$ as $i \longrightarrow \infty$, by the compactness of A. Since $\rho(\beta_{n(i)}, \alpha_{n(i)}) < 1/i$, we also have $\beta_{n(i)} \longrightarrow \alpha$. By the continuity of f at α, $\rho(f(\alpha_{n(i)}), f(\beta_{n(i)})) \leq \rho(f(\alpha_{n(i)}), f(\alpha)) + \rho(f(\alpha), f(\beta_{n(i)})) \longrightarrow 0$, which contradicts $\rho(f(\alpha_{n(i)}), f(\beta_{n(i)})) \geq \epsilon$. This completes the proof by negation.

The compactness of A does not, however, automatically convert pointwise convergence on A into uniform convergence. The "piecewise linear" functions $f_n : [0, 1] \longrightarrow [0, 1]$ defined by the graph at the right converge pointwise to 0 on the compact domain $[0, 1]$, but the convergence is not uniform.



The distance between two non-empty sets A and B, $\rho(A, B)$, is defined as glb $\{\rho(\alpha, \beta) : \alpha \in A$ and $\beta \in B\}$. If A and B intersect the distance is zero. If A and B are disjoint the distance may still be zero. For example, the interior and exterior of a circle in the plane are disjoint open sets whose distance is zero. The x-axis and (the graph of) the function $f(x) = 1/x$ are disjoint closed sets whose distance apart is 0. However, if one of the closed sets is compact then the distance must be positive.

Theorem 18. If A is compact, B is closed and $\{A, B\}$ is disjoint, then $\rho(A, B) > 0$.

Proof. Automatic contradiction.

This result is again a uniformity condition. Saying that a set A is disjoint from a closed set B is saying that $(\forall \alpha^{\in A})(\exists r^{>0})(S_r(\alpha) \cap B = \emptyset)$. Saying that $\rho(A, B) > 0$ is saying that $(\exists r^{>0})(\forall \alpha^{\in A}) \cdots$.

B. As a last consequence of sequential compactness we shall establish a very powerful property which is taken as the definition of compactness in general topology. First, however, we need some preparatory work. If A is a subset of a metric space W, the spherical neighborhood of A of radius r, $S_r[A]$, is simply the union of all the r-spheres about points of A:

$$S_r[A] = \bigcup \{S_r(\alpha) : \alpha \in A\}$$

$$= \{\xi : (\exists \alpha^{\in A})(\rho(\xi, \alpha) < r\}$$

A set B is said to be totally-bounded $\iff$

$$(\forall r^{>0})(\exists \text{ a finite subset } F \subset B)(B \subset S_r[F]) .$$

Thus for every positive r there exists a finite set $\{\alpha_i\}_1^n \subset B$ such that for every $\xi \in B$ we have $\rho(\xi, \alpha_i) < r$ for some i.

Lemma 9. Every sequentially compact set B is totally-bounded.

Proof. If B is not totally-bounded then

$$(\exists r^{>0})(\forall \text{ finite subset } F \subset B)(B \cap (S_r[F]' \neq \emptyset)$$

We can then define a sequence $\{\alpha_n\}$ inductively by taking α_1 as any point of B,

α_2 as any point of B not in $S_r(\alpha_1)$, and α_n as any point of B not in $S_r[\bigcup_1^{n-1} \alpha_i] = \bigcup_1^{n-1} S_r(\alpha_i)$. Then $\{\alpha_n\}$ is a sequence in B such that $\rho(\alpha_i, \alpha_j) \geqq r$ for all $i \neq j$. But this sequence can have no convergent subsequence. Thus if B is not totally-bounded then B is not sequentially compact, proving the lemma.

Lemma 10. Suppose that A is sequentially compact and that $\mathcal{F}$ is an open covering of A (i.e., $\mathcal{F}$ is a family of open sets and $A \subset \bigcup \mathcal{F}$). Then there exists $r > 0$ such that every sphere of radius r about a point of A lies entirely in some set of the family $\mathcal{F}$. r = Lebesgue number

Proof. Otherwise, $(\forall r^{>0})(\exists \alpha^{\in A})(\forall B^{\in \mathcal{F}})(S_r(\alpha)$ is not a subset of B). Take $r = 1/n$, with corresponding sequence $\{\alpha_n\}$. Thus $S_{1/n}(\alpha_n)$ is not a subset of any $B \in \mathcal{F}$. Since A is sequentially compact $\{\alpha_n\}$ has a convergent subsequence $\alpha_{n(i)} \longrightarrow \alpha$ as $i \longrightarrow \infty$. Since $\mathcal{F}$ covers A, some B in $\mathcal{F}$ contains α, and then $S_\epsilon(\alpha) \subset B$ for some $\epsilon > 0$ since B is open. Taking i large enough so that $1/i < \epsilon/2$ and also $\rho(\alpha_{n(i)}, \alpha) < \epsilon/2$, we have $S_{1/n(i)}(\alpha_{n(i)}) \subset S_\epsilon(\alpha) \subset B$, contradicting the fact that $S_{1/n}(\alpha_n)$ is not a subset of any $B \in \mathcal{F}$. The lemma has thus been proved.

Theorem 12. If $\mathcal{F}$ is an open covering of a sequentially compact set A then some finite subfamily of $\mathcal{F}$ covers A.

Proof. By the lemma immediately above there exists $r > 0$ such that $(\forall \alpha^{\in A})(\exists B^{\in \mathcal{F}})(S_r(\alpha) \subset B)$, and by the first lemma there exist $\alpha_1, \ldots, \alpha_n \in A$ such that $A \subset \bigcup_1^n S_r(\alpha_i)$. Taking corresponding sets $B_i \in \mathcal{F}$ with $S_r(\alpha_i) \subset B_i$, $i = 1, \ldots, n$, we clearly have $A \subset \bigcup_1^n B_i$.

In general topology, a set A such that every open covering of A includes a finite open covering is said to be compact. The above theorem says that in a ~~normed linear~~ metric space every sequentially compact set is compact.

§11a. Further properties of the distance between two sets in a nls.

1. Distance is unchanged by a translation: $\rho(A, B) = \rho(A+\gamma, B+\gamma)$ (because $\|(\alpha+\gamma)-(\beta+\gamma)\| = \|\alpha-\beta\|$).

2. $\rho(kA, kB) = |k|\,\rho(A, B)$ (because $\|k\alpha - k\beta\| = |k|\,\|\alpha-\beta\|$)

3. If N is a subspace then the distance from B to N is unchanged if we translate parallel to N: $\rho(N, B) = \rho(N, B+\eta)$ if $\eta \in N$ (because $N+\eta = N$).

4. If $T \in \mathrm{Hom}(V, W)$ then $\rho(T[A], T[B]) \leqq \|T\|\,\rho(A, B)$ (because $\|T(\alpha) - T(\beta)\| \leqq \|T\|\cdot\|\alpha-\beta\|$).

5. In any metric space $p \in \overline{A} \iff \rho(p, A) = 0$ (see the lemmas on page 2.26) and therefore if A is closed and $p \notin A$ then $\rho(p, A) > 0$.

6. If N is a (proper) closed subspace and $\epsilon > 0$, there exists α such that $\|\alpha\| = 1$ and $\rho(\alpha, N) > 1-\epsilon$.

Proof. Choose any $\beta \notin N$. Then $\rho(\beta, N) > 0$ (by (5)) and there exists $\eta \in N$ such that $\|\beta-\eta\| < \rho(\beta, N)/(1-\epsilon)$ (by the definition of $\rho(\beta, N)$). Set $\alpha = \beta-\eta/\|\beta-\eta\|$. Then $\|\alpha\| = 1$ and $\rho(\alpha, N) = \rho(\beta-\eta, N)/\|\beta-\eta\|$ $= \rho(\beta, N)/\|\beta-\eta\| > \rho(\beta, N)(1-\epsilon)/\rho(\beta, N) = 1-\epsilon$, by (2), (3) and the definition of η.

§12. Completeness.

If $\xi_n \longrightarrow \alpha$ as $n \longrightarrow \infty$ then the terms ξ_n obviously get close to each other as n gets large. On the other hand, if $\{\xi_n\}$ is a sequence whose terms get arbitrarily close to each other as $n \longrightarrow \infty$ then $\{\xi_n\}$ clearly ought to converge to a limit. It may not, however; the desired limit vector may be missing in the space V. If V is such that every sequence which ought to converge actually does converge then we say that V is complete. We now make this notion precise.

Definition. $\{\xi_n\}$ is a Cauchy sequence $\Longleftrightarrow$ $(\forall \epsilon)(\exists N)(\forall m, n)$ $(m > N$ and $n > N \Longrightarrow \rho(\xi_m, \xi_n) < \epsilon)$.

Lemma 11. If $\xi_n \longrightarrow \alpha$ as $n \longrightarrow \infty$ then $\{\xi_n\}$ is Cauchy.

Proof. Given ϵ we choose N such that $n > N \Longrightarrow \rho(\xi_n, \alpha) < \epsilon/2$. Then if m and n are both greater than N we have

$$\rho(\xi_m, \xi_n) \leqq \rho(\xi_m, \alpha) + \rho(\alpha, \xi_n) < \epsilon/2 + \epsilon/2 = \epsilon.$$

Lemma 12. If $\{\xi_n\}$ is Cauchy and if a subsequence is convergent, say $\xi_{n(i)} \longrightarrow \alpha$ as $i \longrightarrow \infty$, then $\xi_n \longrightarrow \alpha$ as $n \longrightarrow \infty$.

Proof. Given ϵ, we take N so that for all m and $n > N$ we have $\rho(\xi_m, \xi_n) < \epsilon$. Now take an arbitrary δ and choose I so that $i > I \Longrightarrow$ both $n(i) > N$ and $\rho(\xi_{n(i)}, \alpha) < \delta$. Then $m > N$ and $i > I \Longrightarrow \rho(\xi_m, \alpha) \leqq \rho(\xi_m, \xi_{n(i)}) + \rho(\xi_{n(i)}, \alpha) < \epsilon + \delta$. Thus $m > N \Longrightarrow \rho(\xi_m, \alpha) < \epsilon + \delta$, for all δ, and so $(\xi_m, \alpha) \leqq \epsilon$. Thus we have shown that $\xi_m \longrightarrow \alpha$ as $m \longrightarrow \infty$.

Theorem 20. If V and W are normed linear spaces, $\{\xi_n\}$ is Cauchy in V and $T \in \text{Hom}(V, W)$ then $\{T(\xi_n)\}$ is Cauchy in W.

Proof. Given ϵ choose N so that $m, n > N \Longrightarrow \|\xi_m - \xi_n\| < \epsilon/\|T\|$. Then $m, n > N \Longrightarrow \|T(\xi_m) - T(\xi_n)\| = \|T(\xi_m - \xi_n)\| \leqq \|T\| \cdot \|\xi_m - \xi_n\| < \epsilon$.

This lemma has a substantial generalization, as follows.

Theorem 21. If A and B are metric spaces, $\{\xi_n\}$ is Cauchy in A and $F : A \longrightarrow B$ is uniformly continuous, then $\{F(\xi_n)\}$ is Cauchy in B.

The proof will be left as an exercise. $m(\epsilon) = N(\delta(\epsilon))$

The student should try to acquire a good intuitive feel for the truth of these lemmas, after which the technical proofs become mere transcriptions of the obvious.

Definition. A metric space A is complete $\Longleftrightarrow$ every Cauchy sequence in A converges to a limit in A. A complete normed linear space is called a Banach space.

We are now going to list some important examples of Banach spaces. In each case a proof is necessary so the list becomes a collection of theorems.

Theorem 22. $\mathbb{R}$ is complete.

Proof. Let $\{x_n\}$ be Cauchy in $\mathbb{R}$. Then $\{x_n\}$ is bounded (why?) and so, by C, has a convergent subsequence. Lemma 12 then implies that $\{x_n\}$ is convergent; q. e. d.

Theorem 23. If p and q are equivalent norms on V and $\langle V, p\rangle$ is complete then so is $\langle V, q\rangle$.

Proof. If $\{\xi_n\}$ is q-Cauchy then it is p-Cauchy, hence p-convergent, hence q-convergent.

Theorem 24. If V_1 and V_2 are Banach spaces then so is $V_1 \times V_2$.

Proof. Use the sum norm on $V_1 \times V_2$. If $\{\langle \xi_n, \eta_n \rangle\}$ is Cauchy then so are each of $\{\xi_n\}$ and $\{\eta_n\}$ (since $\|\xi\| \leqq \|\xi\| + \|\eta\| = \|\langle \xi, \eta \rangle\|$). Then $\xi_n \longrightarrow \alpha$ and $\eta_n \longrightarrow \beta$ for some $\alpha \in V_1$ and $\beta \in V_2$. Thus $\|\xi_n - \alpha\| + \|\eta_n - \beta\| \longrightarrow 0$, i.e., $\|\langle \xi_n, \eta_n \rangle - \langle \alpha, \beta \rangle\| \longrightarrow 0$, i.e., $\langle \xi_n, \eta_n \rangle \longrightarrow \langle \alpha, \beta \rangle$ in $V_1 \times V_2$.

Corollary 1. If $\{V_i\}_1^n$ are Banach spaces then so is $\prod_{i=1}^n V_i$.

Corollary 2. Every finite dimensional space is a Banach space (in any norm).

Proof. $\mathbb{R}^n$ is complete (in the 1-norm, say) by Theorem 22 and Corollary 1 above. We then impose a 1-norm on V by choosing a basis.

Theorem 25. Let W be a Banach space, let A be any set, and let $\mathcal{B}(A, W)$ be the vector space of all bounded mappings of A into W with the uniform norm $\|f\|_\infty = \text{lub}\,\{\|f(a)\| : a \in A\}$. Then $\mathcal{B}(A, W)$ is complete.

Proof. Let $\{f_n\}$ be Cauchy. Given ϵ, let N be such that $m, n > N \Rightarrow \|f_n - f_m\|_\infty < \epsilon$. Choose any $a \in A$. Since $\|f_n(a) - f_m(a)\| \leqq \|f_n - f_m\|_\infty < \epsilon$, it follows that $\{f_n(a)\}$ is Cauchy in W and so convergent. Define $g : A \longrightarrow W$ by $g(a) = \lim f_n(a)$, for each $a \in A$. We have to show that g is bounded and that $f_n \longrightarrow g$.

Now, just as in the proof of Lemma ,

$$\|f_m(a) - f_n(a)\| < \epsilon \quad \text{and} \quad f_n(a) \longrightarrow g(a)$$

together imply that $\|f_m(a) - g(a)\| \leqq \epsilon$. Thus $m > N \Rightarrow \|f_m(a) - g(a)\| \leqq \epsilon$ for all $a \in A$ and hence $m > N \Rightarrow \|f_m - g\|_\infty \leqq \epsilon$. This implies both that $f_m - g \in \mathcal{B}(A, W)$, and so $g = f_m - (f_m - g) \in \mathcal{B}(A, W)$, and also that $f_m \longrightarrow g$ in the uniform norm. q.e.d.

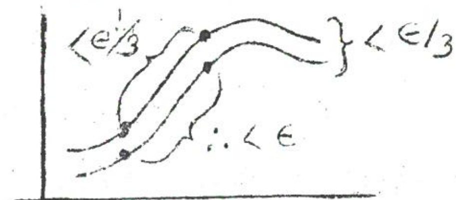
Theorem 26 . A closed subset of a complete metric space is complete.

The proof is left to the reader.

Theorem 27 . In the context of Theorem 25 let A be a metric space, let $\mathcal{C}(A, W)$ be the space of continuous mappings of A into W, and set $\mathcal{BC}(A, W) = \mathcal{B}(A, W) \cap \mathcal{C}(A, W)$.

Then $\mathcal{BC}$ is a closed subspace of $\mathcal{C}$.

Proof. We suppose that $\{f_n\} \subset \mathcal{BC}$ and that $\|f_n - g\|_\infty \longrightarrow 0$ where $g \in \mathcal{B}$. We have to show that g is continuous. This is an application of a much used "up, over and down" argument which can be schematically indicated as at the right.



Given ε we first choose any n such that $\|f_n - g\|_\infty < \epsilon/3$. Consider, now, any $\alpha \in A$. Since f_n is continuous at α, there exists δ such that $\rho(\xi, \alpha) < \delta \Rightarrow \|f_n(\xi) - f_n(\alpha)\| < \epsilon/3$. Then, $\rho(\xi, \alpha) < \delta \Rightarrow \|g(\xi) - g(\alpha)\| \leq \|g(\xi) - f_n(\xi)\| + \|f_n(\xi) - f_n(\alpha)\| + \|f_n(\alpha) - g(\alpha)\| < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon$. Thus g is continuous at α, for every $\alpha \in A$ and so $g \in \mathcal{BC}$, q. e. d.

This important classical result is traditionally stated: the limit of a uniformly convergent sequence of continuous functions is continuous.

Corollary: $\mathcal{BC}(A, W)$ is a Banach space.

Theorem 28 . If A is a compact metric space then A is complete.

Proof. A Cauchy sequence in A has a subsequence converging to a limit in A, and therefore, by Lemma 12 , itself converges to that limit. Thus A is complete.

In the last section we proved that a compact set is also totally bounded. It can be shown, conversely, that a complete, totally-bounded set A is compact, so that these two properties together are equivalent to compactness.

The crucial fact is that if A is totally bounded then every sequence in A has a Cauchy subsequence. If A is also complete this Cauchy subsequence will converge to a point of A. Thus the fact that total boundedness and completeness together are equivalent to compactness follows directly from the next lemma.

Lemma 13 . If A is totally bounded then every sequence in A has a Cauchy subsequence. diagonal method

Proof. Let $\{p_m\}$ be any sequence in A. Since A can be covered by a finite number of spheres of radius 1 at least one sphere in such a covering contains infinitely many of the points $\{p_m\}$. More precisely, there exists an infinite set $M_1 \subset \mathbb{Z}^+$ such that the set $\{p_m : m \in M_1\}$ lies in a single sphere of radius 1. Suppose that $M_1, \ldots, M_n \subset \mathbb{Z}^+$ have been defined so that $M_{i+1} \subset M_i$, $i = 1, \ldots, n-1$, M_n is infinite, and $\{p_m : m \in M_n\}$ is a subset of a sphere of radius $1/n$. Since A can be covered by a finite family of spheres of radius $1/(n+1)$, at least one of the covering spheres contains infinitely many points of the set $\{p_m : m \in M_n\}$. More precisely, there exists an infinite set $M_{n+1} \subset M_n$ such that $\{p_m : m \in M_{n+1}\}$ is a subset of a sphere of radius $1/(n+1)$. We thus define an infinite sequence $\{M_n\}$ of subsets of $\mathbb{Z}^+$ having the above properties.

Now choose $m_1 \in M_1$, $m_2 \in M_2$ so that $m_2 > m_1$, and, in general, $m_{n+1} \in M_{n+1}$ so that $m_{n+1} > m_n$. Then the subsequence $\{p_{m_n}\}_n$ is Cauchy. For, given ϵ, we can choose n so that $1/n < \epsilon/2$. Then $i, j > n \Longrightarrow m_i, m_j \in M_n \Longrightarrow \rho(p_{m_i}, p_{m_j}) < 2(1/n) < \epsilon$. This proves the lemma.

§13. Some applications of completeness.

As we have seen, completeness, as a property of a metric space is weaker than compactness. This means that whenever compactness is present we use it as the primary tool. However, in infinite dimensional normed linear space settings compactness is very rare and it is fortunate indeed that so much can be done with completenss. We prove below two important theorems involving completeness which we shall apply in the next chapter.

Theorem 29. Let U be a subspace of a normed linear space V and let T be a bounded linear mapping of U into a Banach space W. Then T has a uniquely determined extension to a bounded linear transformation S on the closure $\overline{U}$ of U into W. Moreover, $\|S\| = \|T\|$.

Proof. Fix $\alpha \in \overline{U}$ and choose $\{\xi_n\} \subset U$ so that $\xi_n \longrightarrow \alpha$. Then $\{\xi_n\}$ is Cauchy and $\{T(\xi_n)\}$ is Cauchy (by the lemmas of the preceding section), so that $\{T(\xi_n)\}$ converges to some $\beta \in W$. If $\{\eta_n\}$ is any other sequence in U converging to α then $\xi_n - \eta_n \longrightarrow 0$, $T(\xi_n) - T(\eta_n) = T(\xi_n - \eta_n) \longrightarrow 0$ and so $T(\eta_n) \longrightarrow \beta$ also. Thus β is independent of the sequence chosen and, clearly, β must be the value $S(\alpha)$ at α of any continuous extension S of T. If $\alpha \in U$, then $\beta = \lim T(\alpha_n) = T(\alpha)$ by the continuity of T.

We thus have S uniquely defined on $\overline{U}$ by the requirement that it be a continuous extension of T, and $S \restriction U = T$.

It remains to be shown that S is linear and bounded by $\|T\|$. For any $\alpha, \beta \in \overline{U}$ we choose $\{\xi_n\}, \{\eta_n\} \subset U$ so that $\xi_n \longrightarrow \alpha$ and $\eta_n \longrightarrow \beta$ and then have $S(x\alpha + y\beta) = \lim T(x\xi_n + y\eta_n) = x \lim T(\xi_n) + y \lim T(\eta_n) = xS(\alpha) + yS(\beta)$. Thus S is linear. Finally $\|S(\alpha)\| = \lim \|T(\xi_n)\| \leq \|T\| \lim \|\xi_n\| = \|T\| \cdot \|\alpha\|$. Thus $\|T\|$ is a bound for S, and, since S includes

T, $\|S\| = \|T\|$. q. e. d.

The above theorem has many applications, but we shall use it only once, to obtain the Riemann integral $\int_a^b f(t)\,dt$ of a continuous function f mapping a closed interval [a, b] into a Banach space W as an extension of the trivial integral for step functions.

We prove next the very simple and elegant fixed point theorem, for contraction mappings, which will be the basis for our proofs in the next chapter of the implicit function theorem and the fundamental existence and uniqueness theorem for ordinary differential equations.

Definition. A mapping $F : A \longrightarrow A$ is a contraction if it is a Lipschitz mapping with constant < 1; that is, if there exists C with $0 < C < 1$ such that $\rho(F(\alpha), F(\beta)) \leqq C\rho(\alpha, \beta)$ for all $\alpha, \beta \in A$. A fixed point of F is of course a point α such that $F(\alpha) = \alpha$.

Theorem 30. Let A be a non-empty complete metric space and let $F : A \longrightarrow A$ be a contraction. Then F has a uniquely determined fixed point in A.

Proof. Choose any $\alpha_0 \in A$. Define the sequence $\{\alpha_n\}_0^\infty$ inductively by setting $\alpha_1 = T(\alpha_0)$, $\alpha_2 = T(\alpha_1) = T^2(\alpha_0)$ and $\alpha_n = T(\alpha_{n-1}) = T^n(\alpha_0)$. Set $a = \rho(\alpha_1, \alpha_0)$. Then $\rho(\alpha_2, \alpha_1) = \rho(T(\alpha_1), T(\alpha_0)) \leqq C\rho(\alpha_1, \alpha_0) = Ca$, and, by induction, $\rho(\alpha_{n+1}, \alpha_n) = \rho(T(\alpha_n), T(\alpha_{n-1})) \leqq C\rho(\alpha_n, \alpha_{n-1}) \leqq C \cdot C^{n-1}a = C^n a$. It follows that $\{\alpha_n\}$ is Cauchy, for if $m > n$ then $\rho(\alpha_m, \alpha_n) \leqq \sum_n^{m-1} \rho(\alpha_{i+1}, \alpha_i) \leqq a\sum_n^{m-1} C^i < aC^n/(1-C)$ which $\longrightarrow 0$ as $n \longrightarrow \infty$.

Since A is complete, $\{\alpha_n\}$ converges to some $\alpha \in A$, and then it follows that $T(\alpha) = \lim T(\alpha_n) = \lim \alpha_{n+1} = \alpha$, so that α is a fixed point. If β is any fixed point then $\rho(\beta, \alpha) = \rho(T(\beta), T(\alpha)) \leqq C\rho(\beta, \alpha)$, which implies that $\rho(\beta, \alpha) = 0$ and thus that α is the only fixed point. This completes the proof of the theorem.

The rest of this chapter is only used in III.12 (p. III.32-33)
III.13 (p. III 34-35)

If we tried the above construction for a contraction mapping $K : A \longrightarrow X$, where A is an open subset of the complete space X, then the limit point which turns out to be the fixed point might not be in the domain of K at all. However if A is an open sphere, and if K doesn't move the center very far, then we can recapture the earlier situation.

<u>Corollary 1.</u> Let X be a complete metric space and let $K : S_r(p_0) \longrightarrow X$ be a contraction, with contracting constant C. <1 Suppose that $\rho(K(p_0), p_0) < (1-C)r$. i.e., $\exists \epsilon > 0$ Then K has a unique fixed point.

<u>Proof.</u> We simply check that if ϵ is small enough then K maps the closed sphere $\overline{S_{r-\epsilon}(p_0)}$ into itself, so that the theorem applies.

Take $\epsilon \leqq (1-C)r - \rho(K(p_0), p_0)$. >0 Then $\rho(p, p_0) \leqq r-\epsilon \Rightarrow$ $\rho(K(p), p_0) \leqq \rho(K(p), K(p_0)) + \rho(K(p_0), p_0) \leqq C(r-\epsilon)+(1-C)r - \epsilon \leqq r-\epsilon$. Thus $K[A] \subset A$ if $A = \overline{S_{r-\epsilon}(p_0)}$. The uniqueness of the fixed point is shown directly exactly as in the theorem.

Now let A and B be metric spaces with A complete, and let K be a continuous mapping of $A \times B$ into A such that $p \longrightarrow K(p, q)$ is a contraction on A for each q in B. Then, of course, for each $q \in B$ a unique fixed point $p \in A$ is determined, so that a function $f : B \longrightarrow A$ is uniquely defined by $f(q) = K(f(q), q)$. In order to make the notation a little simpler we shall suppose that A and B are subsets of normed linear spaces.

<u>Corollary 2.</u> If the contraction is <u>uniform</u> with respect to q, that is if, if there exists $C < 1$ such that $\| K(p_1, q) - K(p_2, q) \| \leqq C \| p_1 - p_2 \|$ for all $q \in B$ and all $p_1, p_2 \in A$, then the solution function f above is continuous. (if K is continuous)

$\bar{X} = [0,2]$ $\quad p_0 = 1 \quad r = 1$ $\qquad K = \begin{cases} 0 & \text{if } x \in (0,\delta) \\ (1-\epsilon)(x-\delta) & \text{otherwise} \end{cases}$

$|p_0 - K(p_0)| = 1-(1-\epsilon)(1-\delta) = \delta + \epsilon - \epsilon\delta < \epsilon = (1-C)r$ if $\epsilon < 1$; however, only $\bar{K}$ has fixed pt (0). $\qquad C = (1-\epsilon) < 1$

Cont. in 2nd variable only is used?

Proof. Given ϵ and q_0, choose δ so that $\|q-q_0\| < \delta \Rightarrow \|K(p_0, q_0) - K(p_0, q)\| < (1-C)\epsilon$, where $p_0 = f(q_0)$. Similarly setting $p = f(q)$, we have $\|q-q_0\| < \delta \Rightarrow \|p-p_0\| = \|K(p,q) - K(p_0, q_0)\| \leqq \|K(p,q) - K(p_0,q)\| + \|K(p_0,q) - K(p_0,q_0)\| \leqq C\|p-p_0\| + (1-C)\epsilon$ so that $\|p-p_0\| \leqq (1-C)\epsilon/(1-C) = \epsilon$. That is, $\|q-q_0\| < \delta \Rightarrow \|f(q) - f(q_0)\| < \epsilon$, q. e. d.

We shall also find later use for the following theorem.

Theorem 31. Let V and W be Banach spaces. Suppose that $T \in \mathrm{Hom}(V, W)$ has a bounded inverse and set $m = 1/\|T^{-1}\|$. Then S has a bounded inverse whenever $\|T-S\| < m$. Moreover, $\|T-S\| \leqq \epsilon < m \Rightarrow \|T^{-1} - S^{-1}\| \leqq \epsilon/m(m-\epsilon)$. In particular the invertible maps form an open subset $\mathcal{J}$ of $\mathrm{Hom}(V, W)$ and the mapping $T \longrightarrow T^{-1}$ is continuous on $\mathcal{J}$.

Proof. If $\alpha \in V$ then $\alpha = T^{-1}(T(\alpha))$ and $\|\alpha\| \leqq \|T^{-1}\| \cdot \|T(\alpha)\|$. Thus $\|T(\alpha)\| \geqq m\|\alpha\|$ for every $\alpha \in V$; that is, T is bounded below by m. Suppose now that $\|T-S\| \leqq \epsilon < m$. Then $\|S(\alpha)\| = \|T(\alpha) + (S-T)(\alpha)\| \geqq \|T(\alpha)\| - \|(T-S)(\alpha)\| \geqq m\|\alpha\| - \epsilon\|\alpha\| = (m-\epsilon)\|\alpha\|$. Thus, if $\|T-S\| \leqq \epsilon < m$, then S is bounded below by $m-\epsilon$, and, in particular, is injective. It also follows that R = range S is a complete subspace of W (§2, 13) and therefore closed. We assert that in fact R = W. Otherwise $\exists \beta \in W$ such that $\|\beta\| = 1$ and $\rho(\beta, R) > \epsilon/m$ (§11a). Then $\alpha = T^{-1}(\beta)$ has norm $\leqq \|T^{-1}\| = 1/m$ and

$\|\beta - S(\alpha)\| = \|(T-S)(\alpha)\| \leqq \|(T-S)\| \cdot \|\alpha\| \leqq \epsilon(1/m) = \epsilon/m$, a contradiction. Therefore $R = W$ and we have proved that S has a bounded inverse whenever $\|T-S\| < m$.

Finally, if $\alpha = T^{-1}(\beta)$ and $\alpha' = S^{-1}(\beta)$, then $T(\alpha - \alpha') = \beta - T(\alpha') = (S-T)(\alpha')$ and so $(\alpha - \alpha') = (T^{-1} \circ (S-T) \circ S^{-1})(\beta)$. Therefore $\|T^{-1} - S^{-1})(\beta)\| = \|\alpha - \alpha'\| \leqq 1/m \cdot \epsilon \cdot 1/m{-}\epsilon \|\beta\|$, so that $\|S^{-1} - T^{-1}\| \leqq \epsilon/m(m-\epsilon)$. This completes the proof of the theorem.

Chapter 3. The differential calculus

§1. Vector functions of a real variable ; integration.

We shall begin by taking a brief look at the integral and differential calculus on a one dimensional domain, which we shall always take to be a closed interval $[a, b] \subset \mathbb{R}$. The integral of such a function will be thought of as (b-a) times the vector which is the "average value" or "average vector" of the vectors making up the range of the function. The derivative of a function at a point will be interpreted as the tangent vector to the curve that is the range of the function.

To begin with we shall integrate only certain elementary functions called step functions. A finite subset A of [a, b] which contains the two end points a and b will be called a partition of [a, b]. Thus A is (the range of) some finite sequence $\{t_i\}_0^n$ where $a = t_0 < t_1 < \cdots < t_n = b$, and A subdivides [a, b] into a sequence of smaller intervals. To be definite, we shall take the open intervals (t_{i-1}, t_i), $i = 1, \ldots, n$ as the intervals of the subdivision. If A and B are partitions and $A \subset B$ we shall say that B is a refinement of A. Then each interval (s_{j-1}, s_j) of the B subdivision is included in an interval (t_{i-1}, t_i) of the A subdivision ; t_{i-1} is the largest element of A which is $\leq s_{j-1}$, and t_i the smallest $\geq s_j$. A step function is simply a map $f : [a, b] \longrightarrow W$ which is constant on the intervals of some subdivision $A = \{t_i\}_0^n$. That is, there exists a sequence of vectors $\{\alpha_i\}_1^n$ such that $f(\xi) = \alpha_i$ when $\xi \in (t_{i-1}, t_i)$. The values of f at the subdividing points may be among these values or different.

For each step function f we define $\int_a^b f(t)\,dt$ as $\sum_{i=1}^{n} \alpha_i \Delta t_i$, where $f = \alpha_i$ on (t_{i-1}, t_i) and $\Delta t_i = t_i - t_{i-1}$. If f were real-valued this would be simply the sum of the areas of the rectangles making up the region between the graph of f and the t-axis. Now, f may be described as a step function in terms of many different subdivisions. For example, if f is constant on the intervals of A and if we obtain B from A by adding one new point s, then f is constant on the (smaller) intervals of B. We have to be sure that the value of the integral of f doesn't change when we change the describing subdivision. In the case just mentioned this is easy to see. The one new point s lies in some interval (t_{i-1}, t_i), defined by the partition A. The contribution of this interval to the A sum is $\alpha_i(t_i - t_{i-1})$, while in the B sum it splits into $\alpha_i(t_i - s) + \alpha_i(s - t_{i-1})$ But this is the same vector. The remaining summands are the same in the two sums, and the integral is therefore unchanged. In general, suppose that f is a step function with respect to A and also with respect to C. Set $B = A \cup C$, the "common refinement" of A and C. We can pass from A to B in a sequence of steps at each of which we add one new point. As we have seen, the integral remains unchanged at each of these steps and so it is the same for A as for B. It is similarly the same for C and B, and so for A and C. We have thus shown that $\int_a^b f$ is independent of the subdivision used to define f.

Now fix [a, b] and W and let $\mathcal{E}$ be the set of all step functions on [a, b] into W. Then $\mathcal{E}$ is a vector space. For, if f and g in $\mathcal{E}$ are step functions relative to partitions A and B then both functions are constant on the intervals of $C = A \cup B$ and so therefore is $xf + yg$. Moreover, if $C = \{t_i\}_0^n$, and if on (t_{i-1}, t_i) $f = \alpha_i$ and $g = \beta_i$, so that $xf + yg = x\alpha_i + y\beta_i$ there, then the equation

$$\sum_{i=1}^{n} (x\alpha_i + y\beta_i)\Delta t_i = x\left(\sum_1^n \alpha_i \Delta t_i\right) + y\left(\sum_1^n \beta_i \Delta t_i\right)$$

is just $\int_a^b (xf + yg) = x\int_a^b f + y\int_a^b g$, and the map $f \longrightarrow \int_a^b f$ is thus linear from $\mathcal{E}$ into W. Finally

$$\left\| \int_a^b f \right\| = \left\| \sum_1^n \alpha_i \Delta t_i \right\| \leqq \sum_1^n \|\alpha_i\| \Delta t_i \leqq \|f\|_\infty (b - a)$$

where $\|f\|_\infty = \text{lub}\,\{\|f(t)\| : t \in [a, b]\} = \max\,\{\|\alpha_i\| : 1 \leqq i \leqq n\}$. That is, if we use on $\mathcal{E}$ the uniform norm defined from the norm of W, then the linear mapping $f \longmapsto \int_a^b f$ is bounded by $(b - a)$. If W is <u>complete</u> this transformation therefore has a unique bounded linear extension to the closure $\overline{\mathcal{E}}$ of $\mathcal{E}$ in $\mathcal{C}([a, b], W)$ by Theorem 29 of Chapter II. But we can show that $\overline{\mathcal{E}}$ includes the space $\mathcal{C}([a, b], W)$ of all <u>continuous</u> functions on $[a, b]$ into W, and the integral of a continuous function is thus uniquely defined.

<u>Lemma</u> 1. $\mathcal{C}([a, b], W) \subset \overline{\mathcal{E}}$.

Proof. A continuous function f on $[a, b]$ is uniformly continuous (II, Thm. 27.) That is, given $\epsilon > 0$, there exists $\delta > 0$ such that $|s - t| < \delta \Rightarrow \|f(s) - f(t)\| < \epsilon$. Now take any partition $A = \{t_i\}_0^n$ on $[a, b]$ such that $\Delta t_i = t_i - t_{i-1} < \delta$ for all i, and take α_i as any value of f on (t_{i-1}, t_i). Then $\|f(t) - \alpha_i\| < \epsilon$ on $[t_{i-1}, t_i]$. Thus if g is the step function with value α_i on $(t_{i-1}, t_i]$ and $g(a) = \alpha_1$ then $\|f - g\|_\infty \leqq \epsilon$.

Our main theorem is a recapitulation.

Theorem 1. If W is a Banach space and $V = \mathcal{C}([a, b], W)$, under the uniform norm, then there exists $J \in \mathrm{Hom}(V, W)$ uniquely determined by setting $J(f) = \lim \int_a^b f_n$, where $\{f_n\}$ is any sequence in $\mathcal{E}$ converging to f, and $\int_a^b f_n$ is the elementary integral on $\mathcal{E}$. Moreover, $\|J\| \leqq (b - a)$.

If f is elementary on [a, b] into W and $c \in [a, b]$ then of course f is elementary on each of [a, c] and [c, b]. If c is added to a subdivision A used in defining f and if the sum defining $\int_a^b f$ with respect to $B = A \cup \{c\}$ is broken into two sums at c, we clearly have $\int_a^b f = \int_a^c f + \int_c^b f$. This same identity then follows for any continuous function f on [a, b] since $\int_a^b f = \lim \int_a^b f_n =$

$$\lim \left(\int_a^c f_n + \int_c^b f_n \right) = \lim \int_a^c f_n + \lim \int_c^b f_n = \int_a^c f + \int_c^b f .$$

§2. Vector functions of a real variable ; differentiation

In the last chapter we were principally discussing continuity and had no compelling need for the general notion of limit. The difference quotients of the calculus, on the other hand, are functions not defined at the point near which their behavior is crucial, and their discussion requires the notion of the limit of $f(\xi)$ as ξ approaches, but stays distinct from, α.

Definition. $f(\xi) \longrightarrow \beta$ as $\xi \longrightarrow \alpha \iff$

$$(\forall \epsilon^{>0})(\exists \delta^{>0})(\forall \xi)(\xi \in \text{dom } f \text{ and } 0 < \|\xi - \alpha\| < \delta \Rightarrow \|f(\xi) - \beta\| < \epsilon).$$

One word of warning. It might happen that for some δ, $S_\delta(\alpha) \cap$ dom f is empty or contains just the one point α. In the first case $\alpha \notin \overline{(\text{dom } f)}$, and in the second α is an "isolated point" of dom f. In either case for this δ the hypothesis of the implication in the definition is never true and the implication is therefore "vacuously" true for <u>any</u> β. On the other hand, if every sphere about α intersects dom f in points other than α (in which case we say that α is a <u>limit point</u> of dom f) then we have the usual proof that

$$f(\xi) \longrightarrow \beta \text{ and } f(\xi) \longrightarrow \beta' \Rightarrow \beta = \beta',$$

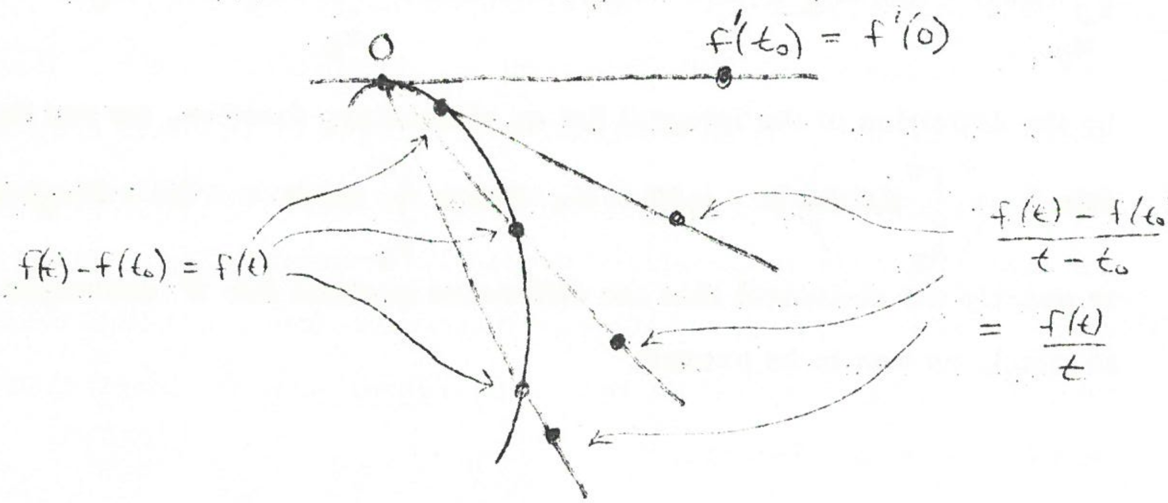
so that now β is <u>the</u> limit of f as $\xi \longrightarrow \alpha$, and we can write $\beta = \lim_{\xi \to \alpha} f(\xi)$.

A <u>continuous</u> f on [a, b] into W will be called a <u>parametrized arc</u> in W. We now want to consider the derivative of such a function (arc) at $t_0 \in [a, b]$. The old definition works perfectly:

$$f'(t_0) = \lim_{t \to t_0} \frac{f(t) - f(t_0)}{t - t_0}$$

where the difference quotient on the right above has domain $[a, b] - \{t_0\}$.

We shall also call $f'(t_0)$ the <u>tangent vector</u> to the parametrized arc f at t_0. This terminology fits our geometric intuition as the following sketch suggests. For simplicity we have set $t_o = 0$ and $f(t_o) = 0$.



There is no need to go into any detail about the derivative here because our later discussion of the differential will include everything, but it should be clear that we could establish the standard laws of the differential calculus by the same old arguments. In particular, $f \mapsto f'(t_0)$ is a linear map from the set of arcs having a tangent at t_0 into W.

The fundamental theorem of the calculus is still with us:

Theorem 2. If $f \in \mathcal{C}([a, b], W)$ and $F : [a, b] \longrightarrow W$ is defined by $F(x) = \int_a^x f(t)\,dt$, then F' exists on (a, b) and $F'(x) = f(x)$.

Proof: By the continuity of f at x_0, for every ϵ there exists δ such that $\|f(x_0) - f(x)\| < \epsilon$ whenever $|x - x_0| < \delta$. But then $\left\| \int_{x_0}^x (f(x_0) - f(t))\,dt \right\| \leqq \epsilon\,|x - x_0|$, and since $\int_{x_0}^x f(x_0)\,dt = f(x_0)(x - x_0)$ by the definition of the integral for an elementary function, we see that $\|f(x_0) - \left(\int_{x_0}^x f(t)\,dt\right)/(x - x_0)\| < \epsilon$. Since $\int_{x_0}^x f(t)\,dt = F(x) - F(x_0)$ this is exactly the statement that the difference quotient for F converges to $f(x_0)$, as was to be proved.

§3. The mean value theorem.

We agree with Dieudonné [ ] in taking the mean value theorem to be an inequality.

<u>Theorem 3</u>. Let f be a continuous function on a closed interval $[a, b]$ into a normed linear space, and suppose that $f'(t)$ exists and $\|f'(t)\| \leqq m$ for every $t \in (a, b)$. Then $\|f(b) - f(a)\| \leqq m(b - a)$.

not necessary for usual mean value! Such an m always exists since [a, b] is compact

Proof. Fix $\epsilon > 0$ and $c > a$ and let A be the set of points $x \in [c, b]$ such that $\|f(x) - f(c)\| \leqq (m + \epsilon)(x - c)$. A is non-empty since $c \in A$. Set $\ell = \text{lub } A$. Then $\exists\, x_n \in A$ with $x_n \longrightarrow \ell$, and since $\|f(x_n) - f(c)\| \leqq (m + \epsilon)(x_n - c)$ for each n, we have $\|f(\ell) - f(c)\| \leqq (m + \epsilon)(\ell - c)$ by the continuity of f at ℓ. Thus $\ell \in A$. We claim that $\ell = b$. For if $\ell < b$ then $f'(\ell)$ exists and $\|f'(\ell)\| \leqq m$. Therefore there exists δ such that $\left\| \frac{f(x) - f(\ell)}{x - \ell} \right\| < m + \epsilon$ when $|x - \ell| \leqq \delta$. It follows that $\|f(\ell + \delta) - f(c)\| \leqq \|f(\ell + \delta) - f(\ell)\| + \|f(\ell) - f(c)\| \leqq (m + \epsilon)\delta + (m + \epsilon)(\ell - c) = (m + \epsilon)(\ell + \delta - c)$, so that $\ell + \delta \in A$, a contradiction. Therefore $\ell = b$. We thus have $\|f(b) - f(c)\| \leqq (m + \epsilon)(b - c)$ for every $c \in (a, b)$. Using the continuity of f at a we get $\|f(b) - f(a)\| \leqq (m + \epsilon)(b - a)$, and finally, since ϵ is arbitrary, $\|f(b) - f(a)\| \leqq m(b - a)$.

* § 4. Weak methods.

If V is finite dimensional the above theory can all be thrown back to the standard calculus of real-valued functions of a real variable by applying functionals from V^* and using the natural isomorphism of V^{**} with V. Thus, if $f \in \mathcal{C}([a, b], V)$ and $\lambda \in V^*$, then $\lambda \circ f \in \mathcal{C}([a,b], \mathbb{R})$ and so the integral $\int_a^b \lambda \circ f$ exists from standard calculus. If we vary λ we can check that the map $\lambda \longmapsto \int_a^b \lambda \circ f$ is linear, hence $\in (V^{**})$ and therefore is given by a uniquely determined vector $\alpha \in V$ (by duality, see I, Thm. 28). That is, there exists a unique $\alpha \in V$ such that $\lambda(\alpha) = \int_a^b \lambda \circ f$ for every $\lambda \in V^*$, and we define this α to be $\int_a^b f$. Thus integration is defined so as to commute with the application of linear functionals: $\int_a^b f$ is that vector such that

$$\lambda\left(\int_a^b f\right) = \int_a^b \lambda(f(t))\, dt \quad \text{for all } \lambda \in V^* .$$

Similarly, if all the real-valued functions $\{\lambda \circ f : \lambda \in V^*\}$ are differentiable at x_0 then the mapping $\lambda \longmapsto (\lambda \circ f)'(x_0)$ is linear by the linearity of the derivative in the standard calculus: $((c_1\lambda_1 + c_2\lambda_2) \circ f)' = (c_1(\lambda_1 \circ f) + c_2(\lambda_2 \circ f))' = c_1(\lambda_1 \circ f)' + c_2(\lambda_2 \circ f)'$. Therefore there is again a unique $\alpha \in V$ such that $(\lambda \circ f)'(x_0) = \lambda(\alpha)$ for all $\lambda \in V^*$, and if we define this α to be the derivative $f'(x_0)$ we have again defined an operation of the calculus by commutativity with linear functionals:

$$(\lambda \circ f')(x_0) = (\lambda \circ f)'(x_0)$$

Now the fundamental theorem of the calculus appears as follows. If $F(x) = \int_a^x f$ then $(\lambda \circ F)(x) = \int_a^x \lambda \circ f$ by the weak definition of the integral. The fundamental theorem of the standard calculus then says that $(\lambda \circ F)'$ exists and $(\lambda \circ F)'(x) = (\lambda \circ f)(x) = \lambda(f(x))$. By the weak definition of the derivative we then have that F' exists and $F'(x) = f(x)$.

The one conclusion that we don't get so easily by weak methods is the norm inequality $\left\| \int_a^b f \right\| \leqq (b - a) \|f\|_\infty$. This requires a theorem about norms on finite dimensional spaces that we shall not prove in this course.

Theorem 4. $\|\alpha^{**}\| = \|\alpha\|$ for each $\alpha \in V$. What is being asserted is that $\operatorname{lub}_{\lambda \neq 0} |\alpha^{**}(\lambda)| / \|\lambda\| = \|\alpha\|$. Since $\alpha^{**}(\lambda) = \lambda(\alpha)$ and since $|\lambda(\alpha)| \leqq \|\lambda\| \cdot \|\alpha\|$ by the definition of $\|\lambda\|$, we see that $\operatorname{lub} |\alpha^{**}(\lambda)| / \|\lambda\| \leqq \|\alpha\|$. Our problem is therefore to find $\lambda \in V^*$ with $\|\lambda\| = 1$ and $|\lambda(\alpha)| = \|\alpha\|$. If we multiply through by a suitable constant (replacing α by $c\alpha$, where $c = 1/\|\alpha\|$) we can suppose $\|\alpha\| = 1$. Then α is on the unit spherical surface, and the problem is to find a functional $\lambda \in V^*$ such that the affine subspace (hyperplane) where $\lambda = 1$ touches the unit sphere at α (so that $\lambda(\alpha) = 1$) and otherwise lies outside the unit sphere (so that $|\lambda(\xi)| \leqq 1$ when $\|\xi\| = 1$, and hence $\|\lambda\| \leqq 1$). It is clear geometrically that such "tangent planes" must exist but we shall drop the matter there.

If we assume this theorem, then, since

Use min. dist. to closed convex subset thm.

$$|\lambda(\int_a^b f)| = |\int_a^b \lambda(f(t))\,dt| \leqq (b-a) \max\{|\lambda(f(t))| : t \in [a, b]\} \leqq$$

$$(b-a)\,\|\lambda\| \max\{\|f(t)\|\} \quad (\text{from } |\lambda(\alpha)| \leqq \|\lambda\|\cdot\|\alpha\|) \; = (b-a)\,\|\lambda\|\cdot\|f\|_{\infty},$$

we get

$$\|\int_a^b f\| = \|(\int_a^b f)^{**}\| = \text{lub } |\lambda(\int_a^b f)| / \|\lambda\| \leqq (b-a)\|f\|_{\infty},$$

the extreme members of which form the desired inequality.

§5. Infinitesimals

Definition. A subset $A \subset V$ is a neighborhood of a point $\alpha \Longleftrightarrow A$ includes some open sphere about α. A deleted neighborhood of α is a neibhborhood of α minus the point α itself.

We define special sets of functions $\mathcal{J}$, $\mathcal{O}$, $\mathfrak{o}$ as follows. As usual, V and W are any two normed linear spaces.

$f \in \mathcal{J}(V, W) \Longleftrightarrow$ dom f is a neighborhood of 0 in V, range $f \subset W$, $f(0) = 0$ and f is continuous at 0. These functions are the infinitesimals.

$f \in \mathcal{O}(V, W) \Longleftrightarrow f \in \mathcal{J}(V, W)$ and there exist positive constants r and c such that $\|f(\xi)\| \leqq c\|\xi\|$ on $S_r(0)$. Not only diff. f's.: $x \sin \frac{1}{x} = f(x)$ $\Rightarrow f \in \mathcal{O}$ but is not diff.

$f \in \mathfrak{o}(V, W) \Longleftrightarrow f \in \mathcal{J}(V, W)$ and $\|f(\xi)\|/\|\xi\| \longrightarrow 0$ as $\xi \longrightarrow 0$.

$df(0) = 0$

A simple set of functions on $\mathbb{R}$ into $\mathbb{R}$ makes the qualitative difference between these classes apparent. The function $f : f(x) = |x|^{1/2}$ is in $\mathcal{J}(\mathbb{R}, \mathbb{R})$ but not in $\mathcal{O}$ (or $\mathfrak{o}$), $g : g(x) = x$ is in $\mathcal{O}$ and therefore in $\mathcal{J}$ but not in $\mathfrak{o}$, and $h : h(x) = x^2$ is in all three classes.

Ex. of $f \in \mathcal{J} - \mathcal{O}$??

Our previous notion of the sum of two functions does not apply to a pair of functions $f, g \in \mathcal{J}(V, W)$ because their domains may be different. However $f+g$ is defined on the intersection dom $f \cap$ dom g, which is still a neighborhood of 0. Moreover, addition remains commutative and associative when extended in this way. It is clear that then $\mathcal{J}(V, W)$ is almost a vector space. The only trouble occurs in connection with the equation $f + (-f) = 0$, where the domain of the function on the left is dom f, whereas we naturally take 0 to be the zero function on the whole of V.

*The way out of this difficulty is to identify two functions f and g in $\mathcal{D}$ if they are the same on some sphere about 0. Strictly speaking, we define f and g to be <u>equivalent</u> $(f \sim g) \iff$ there exists $r > 0$ such that $f = g$ on $S_r(0)$. We then check (in our minds) that this <u>is</u> an equivalence relation, and that we now do have a vector space. Its elements are called <u>germs of functions at</u> 0. A germ is thus, strictly speaking, an equivalence class of functions, but in practice one tends to think of germs in terms of their representing functions, only keeping in mind that two functions are the same as germs when they agree on a neighborhood of 0.

The algebraic properties of the three classes $\mathcal{D}$, $\mathcal{O}$, and $\mathfrak{o}$ are crucial for the differential calculus. We gather them together in the following theorem.

<u>Theorem 5</u>. (1) $\mathfrak{o}(V, W) \subset \mathcal{O}(V, W) \subset \mathcal{D}(V, W)$, and each of the three classes is closed under addition and multiplication by scalars.

(2) $f \in \mathcal{O}(V_1, V_2)$ and $g \in \mathcal{O}(V_2, V_3) \implies g \circ f \in \mathcal{O}(V_1, V_3)$, where $\operatorname{dom} g \circ f = f^{-1}[\operatorname{dom} g]$.

(3) If either f or g above is in $\mathfrak{o}$ then so is $g \circ f$.

(4) $f \in \mathcal{O}(V, W)$ and $g \in \mathcal{D}(V, \mathbb{R}) \implies fg \in \mathfrak{o}(V, W)$, and similarly if $f \in \mathcal{D}$ and $g \in \mathcal{O}$.

(5) In (4) if either f or g is in $\mathfrak{o}$ and the other is merely bounded

on a neighborhood of 0, then $fg \in \mathfrak{o}(V, W)$.

(6) $\mathrm{Hom}(V, W) \subset \mathcal{O}(V, W)$

(7) $\mathrm{Hom}(V, W) \cap \mathfrak{o}(V, W) = \{0\}$.

Proof. Set $\mathcal{O}_\epsilon(V, W) = \{f \in \mathcal{J}(V, W) : \|f(\xi)\| \leqq \epsilon\|\xi\|$ on some sphere about $0\}$. Then $f \in \mathcal{O} \Longleftrightarrow (\exists\, \epsilon)(f \in \mathcal{O}_\epsilon)$ and $f \in \mathfrak{o} \Longleftrightarrow (\forall\, \epsilon^{>0})(f \in \mathcal{O}_\epsilon)$. Obviously $\mathfrak{o} \subset \mathcal{O}$.

(1) If $\|f(\xi)\| \leqq \epsilon_1\|\xi\|$ on $S_{r_1}(0)$ and $\|g(\xi)\| \leqq \epsilon_2\|\xi\|$ on $S_{r_2}(0)$ then $\|f(\xi) + g(\xi)\| \leqq (\epsilon_1 + \epsilon_2)\|\xi\|$ on $S_r(0)$, where $r = \min\{r_1, r_2\}$. Thus $\mathcal{O}$ is closed under addition. The closure of $\mathfrak{o}$ under addition follows similarly, or simply from the limit of a sum being the sum of the limits.

(2) If $\|f(\xi)\| \leqq \epsilon_1\|\xi\|$ when $\|\xi\| \leqq r_1$ and $\|g(\eta)\| \leqq \epsilon_2\|\eta\|$ when $\|\eta\| \leqq r_2$ then $\|g(f(\xi))\| \leqq \epsilon_2\|f(\xi)\| \leqq \epsilon_2\epsilon_1\|\xi\|$ when $\|\xi\| \leqq r = \min\{r_1, r_2/\epsilon_1\}$.

(3) Now suppose that $f \in \mathfrak{o}$ in (2). Then, given ϵ we can take $\epsilon_1 = \epsilon/\epsilon_2$ and have $\|g(f(\xi))\| \leqq \epsilon\|\xi\|$ when $\|\xi\| \leqq r$. Thus $g \circ f \in$. The argument when $g \in \mathfrak{o}$ and $f \in \mathcal{O}$ is essentially the same.

(4) Given $\|f(\xi)\| \leqq c\|\xi\|$ on $S_r(0)$ and given ϵ, we choose $\delta : |g(\xi)| \leqq \epsilon/c$ on $S_\delta(0)$ and have $\|f(\xi)g(\xi)\| \leqq \epsilon\|\xi\|$ when $\|\xi\| \leqq \min(\delta, r)$. The other result follows similarly (by commutativity) as also does (5).

(6) A bounded linear transformation is in $\mathcal{O}$ by definition.

(7) Suppose that $f \in \mathrm{Hom}(V, W) \cap \mathfrak{o}(V, W)$. Take any $\alpha \neq 0$. Given ϵ choose r so that $\|f(\xi)\| \leqq \epsilon\|\xi\|$ on $S_r(0)$. Then write α as $\alpha = x\xi$, where $\|\xi\| < r$. (Find ξ and x). Then $\|f(\alpha)\| = \|f(x\xi)\| =$

$x > \frac{\|\alpha\|}{r}$; $\xi = \frac{\alpha}{x}$

$|x| \cdot \| f(\xi) \| \leqq |x| \cdot \epsilon \cdot \| \xi \| = \epsilon \| \alpha \|$. Thus $\| f(\alpha) \| \leqq \epsilon \| \alpha \|$ for every positive ϵ and so $f(\alpha) = 0$. Thus $f = 0$, proving (7).

Remark: The additivity of f was not used in this argument - only its homogeneity. It follows therefore that there is no homogeneous function in $\mathscr{o}$ except 0.

Sometimes when more than one variable is present it is necessary to indicate with respect to which variable a function is in $\mathcal{O}$ or $\mathscr{o}$. We then write "$f(\xi) = \mathcal{O}(\xi)$" for "$f \in \mathcal{O}$", "$\mathcal{O}(\xi)$" being used to designate an arbitrary element of $\mathcal{O}$.

§6. The differential

Definition. $\Delta F(\alpha, \xi) = \Delta F_\alpha(\xi) = F(\alpha + \xi) - F(\alpha)$. Δ is linear

Definition. Let A be a neighborhood of α in V. A mapping $F : A \longrightarrow W$ is differentiable at $\alpha \iff \exists\, T \in \mathrm{Hom}(V, W)$ such that $\Delta F_\alpha(\xi) = T(\xi) + \mathcal{O}(\xi)$.

The $\mathcal{O}o$ theorem implies then that T is uniquely determined, for if also $\Delta F_\alpha = S + \mathcal{O}$ then $T\text{-}S \in \mathcal{O}$ and so $T\text{-}S = 0$ by (7) of the theorem. This uniquely determined T is called the differential of F at α and is designated dF_α. Thus

$$\Delta F_\alpha = dF_\alpha + \mathcal{O}, \text{ where } dF_\alpha \in \mathrm{Hom}(V, W),$$

and dF_α is the unique (bounded) linear approximation to ΔF_α. We gather together in the next theorem certain elementary facts which follow immediately from the definition.

If $f \in V^*$ and $\beta \in W$ then the mapping $\xi \longmapsto f(\xi)\beta$ clearly belongs to $\mathrm{Hom}(V, W)$. It is called a dyad ; we naturally designate it '$f\beta$'. cf. vec. anal. dyad $\bar{x}(\bar{y}\cdot\bar{z}) = \bar{x}\bar{y}\cdot\bar{z}$ (Inner prod. sp. is self-adjoint)

It will be convenient to use the notation $\mathcal{D}_\alpha(V, W)$ for the set of all mappings on neighborhoods of $\alpha \in V$ into W which are differentiable at α.

Theorem 6. (1) $F \in \mathcal{D}_\alpha(V, W) \implies \Delta F_\alpha \in \mathcal{O}(V, W)$.

(2) $F, G \in \mathcal{D}_\alpha(V, W) \implies F + G \in \mathcal{D}_\alpha(V, W)$ and $d(F + G)_\alpha = dF_\alpha + dG_\alpha$.

(3) $F \in \mathcal{D}_\alpha(V, \mathbb{R})$ and $G \in \mathcal{D}_\alpha(V, W) \implies FG \in \mathcal{D}_\alpha(V, W)$ and

$$d(FG)_\alpha = F(\alpha) dG_\alpha + dF_\alpha G(\alpha),$$

the second term being a dyad.

(4) If F is a constant function on V then F is differentiable and $dF_\alpha = 0$.

(5) If $F \in \mathrm{Hom}(V, W)$ then F is differentiable at every $\alpha \in V$ and $dF_\alpha = F$.

Proof. (1) $\Delta F_\alpha = dF_\alpha + o = \mathcal{O} + o = \mathcal{O}$ by (1) and (6) of the $\mathcal{O}o$ theorem.

(2) It is clear that $\Delta(F + G)_\alpha = \Delta F_\alpha + \Delta G_\alpha$.

Therefore $\Delta(F + G)_\alpha = (dF_\alpha + o) + (dG_\alpha + o) = (dF_\alpha + dG_\alpha) + o$, by (1) of the $\mathcal{O}o$ theorem. Since $dF_\alpha + dG_\alpha \in \mathrm{Hom}(V, W)$, we have (2).

(3) $\Delta(FG)_\alpha(\xi) = F(\alpha + \xi) G(\alpha + \xi) - F(\alpha) G(\alpha)$

$= \Delta F_\alpha(\xi) G(\alpha) + F(\alpha) \Delta G_\alpha(\xi) + \Delta F_\alpha(\xi) \Delta G_\alpha(\xi)$. Thus

$\Delta(FG)_\alpha = (dF_\alpha + o) G(\alpha) + F(\alpha)(dG_\alpha + o) + \mathcal{O}\mathcal{O} = dF_\alpha G(\alpha) + F(\alpha) dG_\alpha + o$ by the $\mathcal{O}o$ theorem.

(4) $\Delta F_\alpha = 0 \implies dF_\alpha = 0$

(5) $\Delta F_\alpha(\xi) = F(\alpha + \xi) - F(\alpha) = F(\xi)$.

Thus $\Delta F_\alpha = F \in \mathrm{Hom}(V, W)$, q. e. d.

Somewhat more complicated is the chain rule.

Theorem 7 . $F \in \mathcal{D}_\alpha(V_1, V_2)$ and $G \in \mathcal{D}_{F(\alpha)}(V_2, V_3)$ $\Rightarrow G \circ F \in \mathcal{D}_\alpha(V_1, V_3)$ and

$$d(G \circ F)_\alpha = dG_{F(\alpha)} \circ dF_\alpha$$

Proof. $\Delta(G \circ F)_\alpha(\xi) = G(F(\alpha + \xi)) - G(F(\alpha)) = G(F(\alpha) + \Delta F_\alpha(\xi)) - G(F(\alpha)) = dG_{F(\alpha)}(\Delta F_\alpha(\xi)) + o(\Delta F_\alpha(\xi)) = dG_{F(\alpha)}(dF_\alpha(\xi)) + dG_{F(\alpha)}(o(\xi)) + o \circ \mathcal{O} = (dG_{F(\alpha)} \circ dF_\alpha)(\xi) + \mathcal{O} \circ o + o \circ \mathcal{O}$.

Thus $\Delta(G \circ F)_\alpha = dG_{F(\alpha)} \circ dF_\alpha + o$ by (3) of the $\mathcal{O}o$ theorem. Since $dG_{F(\alpha)} \circ dF_\alpha \in \mathrm{Hom}(V_1, V_3)$, this proves the theorem.

§7. Directional derivatives

If the parametrized arc $\gamma : [a, b] \longrightarrow V$ is differentiable at $x \in (a, b)$, then $d\gamma_x(h) = h d\gamma_x(1) = h\alpha$, where $\alpha = d\gamma_x(1)$. Since $\Delta\gamma_x - d\gamma_x \in \mathfrak{o}$, this gives $\dfrac{\| \Delta\gamma_x(h) - h\alpha \|}{|h|} \longrightarrow 0$ and so $\dfrac{\Delta\gamma_x(h)}{h} \longrightarrow \alpha$ as $h \longrightarrow 0$. Thus α is the derivative $\gamma'(x)$ in the ordinary sense. By reversing the above steps we see that the existence of $\gamma'(x)$ implies the differentiability of γ at x. Thus:

$d\gamma_x(1) = \gamma'(x)$

<u>Theorem 3.</u> A parametrized arc $\gamma : [a, b] \longrightarrow V$ is differentiable at $x \in (a, b) \iff \gamma'(x)$ exists in the ordinary sense, and $d\gamma_x(h) = h\gamma'(x)$.

Now let A be a neighborhood of α in V and suppose given $F : A \longrightarrow W$. We can then consider the behavior of F along the various straight lines through α. For any non-zero $\xi \in V$ the straight line through α in the direction ξ has the parametric representation $t \longrightarrow \alpha + t\xi$. Composing with F, we have the parametrized arc $\gamma : \gamma(t) = F(\alpha + t\xi)$. Its tangent vector (derivative) at the origin $t = 0$, if it exists, is called the <u>derivative of</u> F <u>in the direction</u> ξ <u>at</u> α, or the derivative of F <u>with respect to</u> ξ at α, and is designated $D_\xi F(\alpha)$. As noted above it is a derivative in the ordinary sense:

$$D_\xi F(\alpha) = \lim_{t \longrightarrow 0} \frac{F(\alpha + t\xi) - F(\alpha)}{t}$$

We notice also that if $V = \mathbb{R}^n$ then the derivative of F in the direction of the j^{th} standard basis vector δ^j is just the partial derivative $\partial F / \partial x_j$:

NB.: D_ξ depends linearly on ξ — see Cor.
This depends on mag. as well as direction of ξ.

$$D_{\delta^j}F(\underline{a}) = \lim_{t \to 0} \frac{F(a_1, a_2, \ldots, a_j + t, \ldots, a_n) - F(a_1, a_2, \ldots, a_n)}{t} = \frac{\partial F}{\partial x_j}(\underline{a}).$$

The theorem above has the following corollary.

Corollary. If $F \in \mathcal{S}_\alpha(V, W)$ then $D_\xi F(\alpha)$ exists for every ξ and $D_\xi F(\alpha) = dF_\alpha(\xi)$.

Proof. The parametrized line $\lambda : \lambda(t) = \alpha + t\xi$ is the sum of the constant function α and the linear function $t \longmapsto t\xi$ and is thus differentiable at every t, with $d\lambda_t(h) = h\xi$. The composite $\gamma = F \circ \lambda$ is thus differentiable and $d\gamma_0 = d(F \circ \lambda)_0 = dF_\alpha \circ d\lambda_0$, giving $\gamma'(0) = d(F \circ \lambda)_0(1) = (dF_\alpha \circ d\lambda_0)(1) = dF_\alpha(\lambda'(0)) = dF_\alpha(\xi)$.

This computation can also easily be made directly as in the theorem.

It is natural to wonder whether the existence of all the directional derivatives of F at α imply the differentiability of F there. The answer is "no" and the following theorem supplies most of it.

Theorem 9. Let F be any homogeneous but not linear mapping of V into W. Then $D_\xi F(0)$ exists and equals $F(\xi)$ for all ξ, but F is not differentiable at 0.

Proof. The homogeneity of F is the condition $F(x\xi) = xF(\xi)$ for all $x \in \mathbb{R}$, $\xi \in V$. Then the composition $\gamma : \gamma(t) = F(t\xi)$ of F with any parametrized straight lines through 0 is linear and $D_\xi F(0) = \gamma'(0) = \gamma(1) = F(\xi)$. On the other hand, if F is differentiable at 0 then $dF_0(\xi) = D_\xi F(0) = F(\xi)$ implying that $F = dF_0$ is linear, a contradiction.

We have left only to find a non-linear homogeneous function. Taking the simplest possible situation, where $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$, set $F(x, y) = \dfrac{x^3}{x^2 + y^2}$

if $\langle x, y\rangle \neq \langle 0, 0\rangle$ and $F(0, 0) = 0$. Then $F(tx, ty) = tF(x, y)$ so that F is homogeneous, but F is not linear.

We shall see later (§) that if the directional derivative $D_\xi F(\alpha)$ exists and is continuous as a function of α in some sphere around α_0, for each ξ in some finite spanning set of ξ's (V is therefore finite dimensional) then dF_α exists and is continuous in α.

§8. The differentiability of an m-tuple of functions.

We know that an m-tuple of functions on a common domain, $F^i : A \longrightarrow W_i$ $i = 1, \ldots, m$, is equivalent to a single m-tuple valued function $F : A \longrightarrow W = \prod_{i=1}^m W_i$, $F(\alpha)$ being the m-tuple $\{F^i(\alpha)\}_1^m$ for each $\alpha \in A$. We now check the obviously necessary result that F is differentiable at $\alpha \in A$ if and only if each F^i is differentiable at α. By virtue of the inductive definition $W = (\prod_{i=1}^{m-1} W_i) \times W_m$, it is sufficient to consider only an ordered pair of functions, although the reader will notice that the proof could be made about as easily directly for the general case.

Theorem 10. Given $F : A \longrightarrow W_1$, $G : A \longrightarrow W_2$ and $\langle F, G \rangle = H : A \longrightarrow W_1 \times W_2$ then H is differentiable at $\alpha \in A$ if and only if F and G are both differentiable at α, and $dH = \langle dF, dG \rangle$.

Proof. Let θ_1 be the injection map $\xi \longrightarrow \langle \xi, 0 \rangle$ of W_1 into $W_1 \times W_2$ and, similarly, let $\theta_2(\eta) = \langle 0, \eta \rangle$. Then, strictly speaking, $H = \theta_1 \circ F + \theta_2 \circ G$ (the equation $H = \langle F, G \rangle$ not being really accurate). Since θ_1 and θ_2 are linear and hence differentiable, with $d(\theta_i)_\alpha = \theta_i$, we see that if F and G are differentiable at α then so is H, and that $dH_\alpha = \theta_1 \circ dF_\alpha + \theta_2 \circ dG_\alpha$. Less exactly, this is the statement $dH_\alpha = \langle dF_\alpha, dG_\alpha \rangle$.

Conversely let π_1 be the projection map $\langle \xi, \eta \rangle \longrightarrow \xi$ of $W_1 \times W_2$ onto W_1. We know that π_1 is linear and obviously $F = \pi_1 \circ H$, so that if H is differentiable at α then so is F, with $dF_\alpha = \pi_1 \circ dH_\alpha$. Similarly, $dG_\alpha = \pi_2 \circ dH_\alpha$.

<u>Corollary 1.</u> Given $F^i : A \longrightarrow W_i$, $i = 1, \ldots, m$ and $\langle F^1, \ldots, F^m \rangle = F : A \longrightarrow W = \prod_1^m W_i$ then F is differentiable at $\alpha \in A$ if and only if each F^i is differentiable at α, and then $dF_\alpha = \langle dF^1_\alpha, \ldots, dF^m_\alpha \rangle$.

The reader will notice that again the really correct statements are $F = \sum_1^m \theta_i \circ F^i$ and $dF_\alpha = \sum_1^m \theta_i \circ dF^i_\alpha$, where θ_j is the injection of W_j into $\prod_1^n W_i$.

§9. The differentiability of a function of n-tuples.

The question at the domain end is much more complicated and its answer constitutes one of our fundamental theorems. We regard a function $F(\xi_1, \ldots, \xi_n)$ of n variables as a function of the single n-tuple variable $\underline{\xi} = \langle \xi_1, \ldots, \xi_n \rangle$ and we are therefore concerned with a function $F : A \longrightarrow W$ where A is an open subset of $V = \prod_1^n V_i$. If F is differentiable at $\underline{\alpha} = \langle \alpha_1, \ldots, \alpha_n \rangle$ we define the partial differentials $dF_\alpha^1, \ldots, dF_\alpha^n$ as the restrictions of dF_α to the n coordinate spaces $V_1, \ldots, V_n$. Strictly speaking we must again use the injection θ_j of V_j into $V = \prod_1^n V_i$, and then set $dF_\alpha^i = dF_\alpha \circ \theta_i$. Thus if $\xi = \langle \xi_1, \ldots, \xi_n \rangle$ then $\xi = \sum_1^n \theta_i(\xi_i)$ and $dF_\alpha(\xi) = \sum_1^n dF_\alpha^i(\xi_i)$.

Suppose now that we compose F with a differentiable n-tuplet of functions $G = \langle g^1, \ldots, g^n \rangle$ defined on an open set containing a point γ such that $\alpha = G(\gamma)$, and ask for the differential of the composite function $H = F \circ G$ at γ. Actually, of course, $G = \sum_1^n \theta_i \circ g^i$, so that $d(F \circ G)_\gamma = dF_\alpha \circ dG_\gamma = \sum_1^n dF_\alpha \circ \theta_i \circ dg_\gamma^i = \sum_1^n dF_\alpha^i \circ dg_\gamma^i$, giving a kind of chain rule in terms of partial differentials:

$$d(F \circ G)_\gamma = \sum_1^n dF_\alpha^i \circ dg_\gamma^i .$$

Since $\Delta F_\alpha(\langle 0, \ldots, 0, \xi_j, 0, \ldots, 0 \rangle) - dF_\alpha^j(\xi_j)$

$= \Delta F_\alpha(\langle 0, \ldots, \xi_j, 0, \ldots, 0 \rangle) - dF_\alpha(\langle 0, \ldots, \xi_j, 0, \ldots, 0 \rangle) = \mathcal{O}(\xi_j)$

we see that dF^i_α is simply the differential of the function of the one vector variable ξ_j obtained from F by holding all other variables fixed at α_i $(i \neq j)$. If $V_j = \mathbb{R}$ this function of one vector variable $\xi_j = t$ is now a parametrized arc and its differential is equivalent to a directional derivative as in §3.7. If <u>every</u> $V_j = \mathbb{R}$, so that $V = \mathbb{R}^n$, these directional derivatives are (vector-valued) partial derivatives, as we saw earlier, so that

$\frac{\partial F}{\partial x_i}(\underline{a}) = D_{\delta^i} F(\underline{a})$

$$dF^j_{\underline{a}}(h_j) = h_j \frac{\partial F}{\partial X_j}(\underline{a}) \quad \text{and}$$

$$dF_{\underline{a}}(\underline{h}) = \sum_1^n dF^i_{\underline{a}}(h_i) = \sum_1^n h_i \frac{\partial F}{\partial X_i}(\underline{a}) .$$

This final equation is clearly our old expansion

$$T(\underline{h}) = T\left(\sum h_i \delta^i\right) = \sum h_i T(\delta^i)$$

which simply expresses again the fact that $\frac{\partial F}{\partial X_i}(\underline{a}) = dF_{\underline{a}}(\delta^i) = \underline{D}F_{\delta^i}(\underline{a})$. Remember, though, that $\partial F/\partial X_i(\underline{a})$ is a vector in W and not a number. To get numerical partial derivatives we must have W also a Cartesian space ; see the next section.

We already know that the existence of the n partial differentials $\{dF^i_\alpha\}_1^n$ does not guarantee the existence of dF_α (see § 3.7). However if the dF^i_α exist for all $\alpha \in A$ <u>and are continuous functions of</u> α then it does follow that dF_α exists for each $\alpha \in A$ and that $\alpha \longmapsto dF_\alpha$ is continuous. The proof of this important fact is our next concern. Again it will suffice to consider only the case $n = 2$.

<u>Theorem 11</u>. Let A be an open subset of $V = V_1 \times V_2$ and suppose

that $F : A \longrightarrow W$ has continuous partial differentials $dF^1_{\langle\alpha,\beta\rangle}$ and $dF^2_{\langle\alpha,\beta\rangle}$ on A. Then $dF_{\langle\alpha,\beta\rangle}$ exists and is continuous on A, and $dF_{\langle\alpha,\beta\rangle}(\langle\xi, \eta\rangle) = dF^1_{\langle\alpha,\beta\rangle}(\xi) + dF^2_{\langle\alpha,\beta\rangle}(\eta)$.

Proof. Since $dF^1_{\langle\alpha,\beta\rangle}$ is linear on V_1 and $dF^2_{\langle\alpha,\beta\rangle}$ is linear on V_2, the transformation T on $V_1 \times V_2$ defined loosely by $T(\langle\xi, \eta\rangle) = dF^1_{\langle\alpha,\beta\rangle}(\xi) + dF^2_{\langle\alpha,\beta\rangle}(\eta)$ is linear. Strictly speaking $T = dF^1 \circ \pi_1 + dF^2 \circ \pi_2$ where π_1 is the projection $\langle\xi,\eta\rangle \longrightarrow \xi$ and similarly for π_2. In any event, we are claiming that $dF_{\langle\alpha,\beta\rangle}$ is this T, so our problem is to show that $\Delta F_{\langle\alpha,\beta\rangle}(\xi, \eta) - T(\langle\xi, \eta\rangle) = \mathcal{O}(\langle\xi, \eta\rangle)$, $\langle\alpha, \beta\rangle$ being held fixed.

We shall use the sum norm (the sum of the given norms on V_1 and V_2) for V. Given ϵ, we choose δ so that $\| dF^i_{\langle\alpha',\beta'\rangle} - dF^i_{\langle\alpha,\beta\rangle}\| < \epsilon$ for every $\langle\alpha', \beta'\rangle$ in the δ-sphere about $\langle\alpha, \beta\rangle$ and for $i = 1, 2$. We then express $\Delta F_{\langle\alpha,\beta\rangle}(\xi, \eta)$ as $\Delta^1 + \Delta^2$, where $\Delta^1 = F(\alpha + \xi, \beta + \eta) - F(\alpha, \beta + \eta)$ and $\Delta^2 = F(\alpha, \beta + \eta) - F(\alpha, \beta)$. We now estimate Δ^i in terms of dF^i when $\|\langle\xi, \eta\rangle\| < \delta$. Consider first the function $f(t) = F(\alpha + t\xi, \beta + \eta) - t\,dF^1_{\langle\alpha,\beta\rangle}(\xi)$ for $t \in [0, 1]$. We know that f' exists and $f'(t) = dF^1_{\langle\alpha + t\xi,\beta+\eta\rangle}(\xi) - dF^1_{\langle\alpha,\beta\rangle}(\xi)$. Since $\|\langle\xi, \eta\rangle\| < \delta$ we have $\|f'(t)\| < \epsilon\|\xi\|$. Therefore $\|f(1) - f(0)\| \le \epsilon\|\xi\|$, by the mean value theorem. This is exactly the inequality $\|\Delta^1 - dF^1_{\langle\alpha,\beta\rangle}(\xi)\| \le \epsilon\|\xi\|$. Similarly $\|\Delta^2 - dF^2_{\langle\alpha,\beta\rangle}(\eta)\| \le \epsilon\|\eta\|$. Therefore $\|\Delta F_{\langle\alpha,\beta\rangle}(\xi, \eta) - T(\langle\xi, \eta\rangle)\| \le \epsilon\|\langle\xi, \eta\rangle\|$, and this is just what was to be found.

Corollary 1. If A is an open subset of $\prod_1^n V_i$ and $F : A \longrightarrow W$ is such that for each $i = 1, \ldots, n$ the partial differential dF_α^i exists for all $\alpha \in A$ and is continuous as a function of $\alpha = \langle \alpha_1, \ldots, \alpha_n \rangle$, then dF_α exists and is continuous on A. If $\xi = \langle \xi_1, \ldots, \xi_n \rangle$ then $dF_\alpha(\xi) = \sum_1^n dF_\alpha^i(\xi_i)$.

Proof. The existence and continuity of dF_α^1 and dF_α^2 imply by the theorem that $dF_\alpha^1(\xi_1) + dF_\alpha^2(\xi_2)$ is the differential of F considered as a function of the first two variables when the others are held fixed. Being the sum of continuous functions it is itself continuous in α, and we can now apply the theorem again to add dF_α^3 to this sum partial differential, to get $\sum_1^3 dF_\alpha^i(\xi_i)$ to be the differential of F on $V_1 \times V_2 \times V_3$. And so on (which is colloquial for induction).

Corollary 2. If A is an open subset of $\mathbb{R}^n$ and $F : A \longrightarrow W$ has all of its first partial derivatives $\partial F/\partial x_i$ existing and continuous then F is continuously differentiable on A and $dF_{\underline{a}}(\underline{h}) = \sum_1^n (\partial F/\partial x_i(\underline{a}))h_i$.

§10. The Jacobian matrix

We are now concerned with a mapping $F : A \longrightarrow \mathbb{R}^m$ where A is an open subset of $\mathbb{R}^n$. If F is differentiable at $\underline{a} \in A$ then $dF_{\underline{a}} \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m)$, and $dF_{\underline{a}}$ is therefore given by an $\underline{m} \times \underline{n}$ matrix, say $\{t_{ij}\}$. We want to evaluate the entries t_{ij}.

F is an m-tuplet valued function and therefore equivalent (by duality) to an m-tuplet of real-valued functions. As usual, we assert this equivalence loosely as an equality: $F = \langle f^1, \ldots, f^m \rangle$. According to §3.8, we then have $dF_{\underline{a}} = \langle df^1_{\underline{a}}, \ldots, df^m_{\underline{a}} \rangle$, which simply displays the linear mapping $dF_{\underline{a}}$ into $\mathbb{R}^m$ as an m-tuplet of linear functionals.

From our general matrix theory we know that the j^{th} column of the matrix $\{t_{ij}\}$ is the m-tuplet $T(\delta_j)$; here it is the m-tuplet $dF_{\underline{a}}(\delta_j) = DF_{\delta_j}(\underline{a}) = \frac{\partial F}{\partial x_j}(\underline{a})$. By the above paragraph, this m-tuplet is $\langle df^1_{\underline{a}}(\delta_j), \ldots, df^m_{\underline{a}}(\delta_j) \rangle = \langle \frac{\partial f^1}{\partial x_j}(\underline{a}), \ldots, \frac{\partial f^m}{\partial x_j}(\underline{a}) \rangle$. That is,

$$t_{ij} = \frac{\partial f^i}{\partial x_j}(\underline{a}).$$

If we use the notation $y_i = f^i(\underline{x})$, we can also write $t_{ij} = \frac{\partial y_i}{\partial x_j}(\underline{a})$. This matrix of first partial derivatives evaluated at $\underline{a}$ is called the <u>Jacobian matrix</u> of the transformation $F = \langle f^1, \ldots, f^m \rangle$ at $\underline{a}$.

If we also have a differentiable map $\underline{z} = G(\underline{y}) = \langle g^1(\underline{y}), \ldots, g^\ell(\underline{y}) \rangle$ of an open set $B \subset \mathbb{R}^m$ into $\mathbb{R}^\ell$, and if B contains $\underline{b} = F(\underline{a})$, then $dG_{\underline{b}}$ has, similarly, the matrix $\frac{\partial g^k}{\partial y_i}(\underline{b}) = \frac{\partial z_k}{\partial y_i}(\underline{b})$. Then the chain rule

$$d(G \circ F)_{\underline{a}} = dG_{\underline{b}} \circ dF_{\underline{a}}$$

becomes

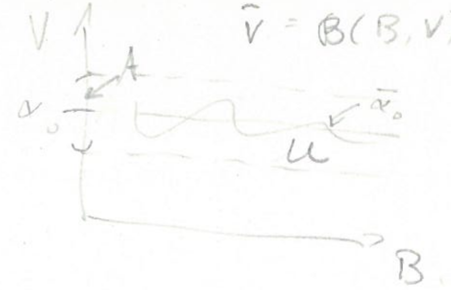
$$\frac{\partial z_k}{\partial x_j}(\underline{a}) = \sum_{i=1}^{m} \frac{\partial z_k}{\partial y_i}(\underline{b}) \frac{\partial y_i}{\partial x_j}(\underline{a})$$

or simply

$$\frac{\partial z_k}{\partial x_j} = \sum_{i=1}^{m} \frac{\partial z_k}{\partial y_i} \frac{\partial y_i}{\partial x_j}.$$

This is the usual form of the chain rule in the calculus. We see that it is merely the expression as matrix multiplication of the composition of linear maps.

Suppose now that A is an open subset of a finite dimensional vector space V and that $H : A \longrightarrow W$ is differentiable at $\alpha \in A$. Suppose that W is also finite dimensional and that $\Theta : V \longrightarrow \mathbb{R}^n$ and $\Phi : W \longrightarrow \mathbb{R}^m$ are any coordinate isomorphisms. If $\overline{A} = \Theta[A]$ then $\overline{A}$ is an open subset of $\mathbb{R}^n$ and $\overline{H} = \Phi \circ H \circ \Theta^{-1}$ is a mapping of $\overline{A}$ into $\mathbb{R}^m$ which is differentiable at $\underline{a} = \Theta(\alpha)$, with $d\overline{H}_{\underline{a}} = \Phi \circ dH_{\alpha} \circ \Theta^{-1}$. Then $d\overline{H}_{\underline{a}}$ is given by its Jacobian matrix $\{\partial h^i / \partial x_j(\underline{a})\}$ which we now call the <u>Jacobian matrix of</u> H <u>with respect to the chosen bases in V and</u> W. Change of bases in V and W changes the Jacobian matrix according to the rule in §I. 9.



* §11. The differentiability of composition

Given $g : A \longrightarrow W$ and an arbitrary set B, let $G : A^B \longrightarrow W^B$ be composition by g. That is, $h = G(f) \iff h = g \circ f$. The question we wish to consider here is the differentiability of G. It will be sufficient for our purposes to take A to be an open sphere $S_r(\alpha_0)$ in a normed linear space V. Then let U be the open sphere $S_r(\bar{\alpha}_0)$ in $\bar{V} = \mathcal{B}(B, V)$ under the uniform norm. Here $\bar{\alpha}_0$ is the constant function on B with value α_0, and $f \in U \iff f \in A^B$ and $\|f - \bar{\alpha}_0\|_\infty < r$.

Theorem 12. Let V and W be (not nec. finite dim.) normed linear spaces, let A be an open sphere $S_r(\alpha_0) \subset V$, let B be an arbitrary set and let U be the open sphere $S_r(\bar{\alpha}_0) \subset \bar{V} = \mathcal{B}(B, V)$. Let $g : A \longrightarrow W$ be uniformly continuously differentiable. That is, dg_α exists for all $\alpha \in A$ and $\alpha \longmapsto dg_\alpha$ is a uniformly continuous mapping of A into $\mathrm{Hom}(V, W)$. Suppose also that $\|dg_\alpha\|$ is bounded on A, say by b. (It has to, by $\|dg_{\alpha_0}\| + \frac{2r\epsilon}{\delta}$) Then the mapping $G : U \longrightarrow \bar{W} = \mathcal{B}(B, W)$ defined by $G(f) = g \circ f$ is continuously differentiable. For any $f \in V$, $dG_f : \bar{V} \longrightarrow \bar{W}$ is defined by $(dG_f(h))(t) = dg_{f(t)}(h(t))$ for all $t \in B$, and $\|dG_f\| \leqq b$.

Proof. We first notice that $\alpha \in A \Rightarrow \|g(\alpha) - g(\alpha_0)\| < rb$, by the mean value theorem, so that g is bounded on A and $g \circ f$ is bounded for all $f \in U$.

Given ϵ, choose δ so that $\alpha, \alpha' \in A$ and $\|\alpha - \alpha'\| < \delta \Rightarrow \|dg_\alpha - dg_{\alpha'}\| < \epsilon$. This is the hypothesis of uniform continuity. It then follows exactly as in the proof of the theorem in §3.9 that $\|\xi\| < \delta \Rightarrow$

$$\|\Delta g_\alpha(\xi) - dg_\alpha(\xi)\| \leqq \epsilon \|\xi\|,$$

provided α and $\alpha + \xi$ are both in A. We now take any $f \in U$, and set $\delta' = r - \|f - \bar{\alpha}_0\|_\infty$. Then for any $h \in \bar{V}$, $\|h\|_\infty < \delta' \Rightarrow f + h \in U$, and the above displayed inequality implies that

$$\|\Delta g_{f(t)}(h(t)) - dg_{f(t)}(h(t))\| \leqq \epsilon \|h(t)\|$$

for all $t \in B$. If we hold f fixed and define $T : \bar{V} \longrightarrow \bar{W}$ by $(T(h))(t) = dg_{f(t)}(h(t))$, then this inequality says exactly that

$$\|\Delta G_f(h) - T(h)\|_\infty \leqq \epsilon \|h\|_\infty .$$

Since this holds for every h in $\bar{V}$ such that $\|h\|_\infty < \delta'$, we have $\Delta G_f = T + \mathcal{O}$, and it remains only to show that $T \in \mathrm{Hom}(\bar{V}, \bar{W})$.

The linearity of T is easy to check. Thus, $(T(h_1 + h_2))(t) = dg_{f(t)}((h_1 + h_2)(t)) = dg_{f(t)}(h_1(t) + h_2(t)) = dg_{f(t)}(h_1(t)) + dg_{f(t)}(h_2(t)) = (T(h_1))(t) + (T(h_2))(t)$. Thus $T(h_1 + h_2) = T(h_1) + T(h_2)$, and homogeneity follows similarly. Also, $\|T(h)\|_\infty = \mathrm{lub}\,\{\|T(h(t))\| : t \in B\} \leqq \mathrm{lub}\,\{\|dg_{f(t)}\| \cdot \|h(t)\| : t \in B\} \leqq b\|h\|_\infty$. Therefore $T \in \mathrm{Hom}(\bar{V}, \bar{W})$ and $\|T\| \leqq b$.

Finally, if $\|f_1 - f_2\|_\infty < \delta$, so that $\|dg_{f_1(t)} - dg_{f_2(t)}\| < \epsilon$ for all t, then we see just as above that $\|dG_{f_1}(h) - dG_{f_2}(h)\|_\infty \leqq \epsilon \|h\|_\infty$. That is, $\|f_1 - f_2\| < \delta \Rightarrow \|dG_{f_1} - dG_{f_2}\| \leqq \epsilon$ and the mapping $f \longrightarrow dG_f$ is continuous. This finishes the proof of the theorem.

§12. The fixed point as a differentiable function

We now return to the situation studied in the second corollary in §2.13 under the additional hypothesis that $K(\alpha, \beta)$ is continuously differentiable, and show that the solution function F defined by $F(\beta) = K(F(\beta), \beta)$, must then also be continuously differentiable.

Theorem 13. Suppose that $S = S_r(\alpha_0) \subset V$, when V is a Banach space, that B is an open subset of a normed linear space W, that $K : S \times B \longrightarrow V$ is continuously differentiable, and that there exists a constant $C < 1$ such that $\| dK^1_{\langle \alpha, \beta \rangle} \| \leqq C$ on $S \times B$ and $\| \alpha_0 - K(\alpha_0, \beta) \| < (1-C)r$ on B. Then for each $\beta \in B$ there exists a unique $\alpha \in S$ such that $\alpha = K(\alpha, \beta)$ and the function $F : B \longrightarrow S$ thus defined is continuously differentiable.

Proof. The inequality $\| dK^1 \| \leqq C$ on $S \times B$ implies that $\| K(\alpha_1, \beta) - K(\alpha_2, \beta) \| \leqq C \| \alpha_1 - \alpha_2 \|$ for any $\alpha_1, \alpha_2 \in S$ and any $\beta \in B$ by the mean value theorem. Thus for each fixed $\beta \in B$, $K(\alpha, \beta)$ satisfies the hypotheses of the first corollary of the fixed point theorem (§2.13) and therefore there exists for each $\beta \in B$ a unique $\alpha \in S$ such that $\alpha = K(\alpha, \beta)$. Let $F : B \longrightarrow S$ be the function thus defined.

We now fix $\beta \in B$ and choose δ so that $\overline{S_\delta(\beta)} \subset B$, and so that $\| dK^2_{\langle F(\beta'), \beta' \rangle} \|$ is bounded, say by b, for β' in this sphere. Setting $\alpha = F(\beta)$ and $\alpha' = F(\beta')$ we have, first, that

$$\| \alpha' - \alpha \| = \| K(\alpha', \beta') - K(\alpha, \beta) \| \leqq \| K(\alpha', \beta') - K(\alpha, \beta') \|$$
$$+ \| K(\alpha, \beta') - K(\alpha, \beta) \| \leqq C \| \alpha' - \alpha \| + b \| \beta' - \beta \| ,$$

so that $\|\alpha' - \alpha\| \leqq (b/(1-C)) \|\beta' - \beta\|$, or

$$\|\xi\| \leqq (b/(1-C)) \|\eta\|.$$

where $\xi = \alpha' - \alpha$ and $\eta = \beta' - \beta$. Thus $\xi = \mathcal{O}(\eta)$. Also,

$\xi = \alpha' - \alpha = K(\alpha', \beta') - K(\alpha, \beta) = K(\alpha + \xi, \beta + \eta) - K(\alpha, \beta)$

$$= \Delta K_{\langle\alpha,\beta\rangle}(\xi, \eta) = dK_{\langle\alpha,\beta\rangle}(\xi, \eta) + \mathfrak{o}(\langle\xi, \eta\rangle)$$

$$= dK^1_{\langle\alpha,\beta\rangle}(\xi) + dK^2_{\langle\alpha,\beta\rangle}(\eta) + \mathfrak{o}(\langle\mathcal{O}(\eta), \eta\rangle)$$

Since $\|dK^1_{\langle\alpha,\beta\rangle}\| \leqq C < 1$, it follows that $T = I - dK^1_{\langle\alpha,\beta\rangle}$ is invertible, by §2.13. We can therefore solve the above equation for ξ, getting $\xi = T^{-1} \circ dK^2_{\langle\alpha,\beta\rangle}(\eta) + \mathfrak{o}(\eta)$. Since $\xi = \alpha' - \alpha = F(\beta') - F(\beta) = F(\beta + \eta) - F(\beta) = \Delta F_\beta(\eta)$, this says that F is differentiable at β. The differentiability of F on B implies its continuity, and since the partial differentials $dK^1_{\langle\alpha,\beta\rangle}$ and $dK^2_{\langle\alpha,\beta\rangle}$ are continuous functions of $\langle\alpha, \beta\rangle$ by hypothesis, and since $S \longmapsto (I - S)^{-1}$ is a continuous mapping of the open unit sphere $\|S\| < 1$ in $\mathrm{Hom}(V, V)$ into $\mathrm{Hom}(V, V)$ by the lemma in §2.13, we have, altogether, that

$$dF_\beta = (I - dK^1_{\langle F(\beta),\beta\rangle})^{-1} \circ dK^2_{\langle F(\beta),\beta\rangle}$$

is a continuous function of β on B. This completes the proof of the theorem.

§13. The implicit function theorem

We state first an important special case. It says that if the linear transformation dH_{α_0} is invertible then H itself is invertible near α_0.

Theorem 14. Let A be an open set of a Banach space V and let $H : A \longrightarrow W$ be a continuously differentiable map of A into a normed linear space W. Suppose that ~~at the point~~ for some $\alpha_0 \in A$, dH_{α_0} has a bounded inverse, and set $\beta_0 = H(\alpha_0)$. Then there exist $r, \rho > 0$ such that for each $\beta \in S_\rho(\beta_0)$ there exists a unique $\alpha \in S_r(\alpha_0)$ such that $\beta = H(\alpha)$. The inverse function $F : S_\rho \longrightarrow S_r$ thus uniquely defined is continuously differentiable, and its differential at $\beta = H(\alpha)$ is given by $dF_\beta = (dH_\alpha)^{-1}$.

This theorem is a corollary of the implicit function theorem which we now state and prove.

Theorem 15. Let V and W be Banach spaces and let $A \times B$ be an open subset of $V \times W$. Let $G : A \times B \longrightarrow W$ be continuously differentiable, and suppose given a point $\langle \alpha_0, \beta_0 \rangle \in A \times B$ such that $G(\alpha_0, \beta_0) = 0$ and such that $dG^1_{\langle \alpha_0, \beta_0 \rangle}$ has a bounded inverse. Then there exist r and ρ greater than zero such that for each $\beta \in S_\rho(\beta_0)$ there exists a unique $\alpha \in S_r(\alpha_0)$ satisfying $G(\alpha, \beta) = 0$. The function $F : S_\rho(\beta_0) \longrightarrow S_r(\alpha_0)$ thus uniquely defined by the equation $G(F(\beta), \beta) = 0$ is continuously differentiable, and dF can be calculated from the equation

$$dG^1_{\langle F(\beta), \beta \rangle} \circ dF_\beta + dG^2_{\langle F(\beta), \beta \rangle} = 0$$

obtained by differentiating the equation $G(F(\beta), \beta) = 0$.

$K: J \times W \to V$

Proof. Set $T = dG^1_{\langle \alpha_0, \beta_0 \rangle}$ and $K(\alpha, \beta) = \alpha - T^{-1}(G(\alpha, \beta))$. Then $dK^1_{\langle \alpha_0, \beta_0 \rangle} = 0$ and $K(\alpha_0, \beta_0) = \alpha_0$. Therefore, for any $C \in (0, 1)$ we can choose r such that $\| dK^1_{\langle \alpha, \beta \rangle} \| \leqq C$ when $\| \langle \alpha, \beta \rangle - \langle \alpha_0, \beta_0 \rangle \| < r$ and $\delta \leqq r$ so that $\| K(\alpha_0, \beta) - \alpha_0 \| < (1 - C)r$ when $\| \beta - \beta_0 \| < \delta$. Then §3.13, together with the fact that $G(\alpha, \beta) = 0 \iff K(\alpha, \beta) = \alpha$, gives the theorem.

The inverse mapping theorem follows from the implicit function theorem upon setting $G(\alpha, \beta) = H(\alpha) - \beta$ and observing that $dG^1_{\langle \alpha, \beta \rangle} = dH_\alpha$.

§14. Ordinary differential equations ; the basic theorem.

Let A be an open subset of a Banach space W, let I be an open interval in $\mathbb{R}$, and let $F : A \times I \longrightarrow W$ be continuous. We want to study the differential equation

$$d\alpha/dt = F(\alpha, t)$$

If $I_0 = (t_0 - \delta, t_0 + \delta) \subset I$ then a function $f : I_0 \longrightarrow A$ is a solution of this equation on I_0 $\iff$ $f'(t)$ exists for every $t \in I_0$ and

$$f'(t) = F(f(t), t)$$

on I_0. Notice that a solution f has to be continuously differentiable, for if f' exists on I_0 then f is continuous on I_0 and then $f'(t) = F(f(t), t)$ is continuous there by the continuity of F.

We are going to see that if F is uniformly a Lipschitz mapping in its first variable, then there exists a uniquely determined "local" solution through any point $\langle \alpha_0, t_0 \rangle \in A \times I$.

In saying that the solution f goes through $\langle \alpha_0, t_0 \rangle$ we mean, of course, that $\alpha_0 = f(t_0)$. The requirement that the solution f have the value α_0 when $t = t_0$ is also called an <u>initial condition.</u>

<u>Theorem 16.</u> Let A, I and F be as above and suppose that for some constant c, $\| F(\alpha_1, t) - F(\alpha_1, t) \| \leq c \|\alpha_1 - \alpha_2\|$ for all $\alpha_1, \alpha_2 \in A$, $t \in I$. Then for any $\langle \alpha_0, t_0 \rangle \in A \times I$ there exists $r, \delta > 0$ such that there is a unique solution f to $d\alpha/dt = F(\alpha, t)$ satisfying $f(t_0) = \alpha_0$, $\operatorname{dom} f = I_0 = (t_0 - \delta, t_0 + \delta)$ and range $f \subset S_r(\alpha_0)$.

Proof. If f is a solution on I_0 through $\langle \alpha_0, t_0 \rangle$ then

$$f(t) - f(t_0) = \int_{t_0}^{t} f'(s)\,ds = \int_{t_0}^{t} F(f(s), s)\,ds \text{ , so that}$$

$$f(t) = \alpha_0 + \int_{t_0}^{t} F(f(s), s)ds$$

for $t \in I_0$. Conversely, if f satisfies this equation then the fundamental theorem of the calculus implies that $f'(t)$ exists and equals $F(f(t), t)$ on I_0, so that f is a solution which clearly goes through $\langle \alpha_0, t_0 \rangle$. Thus the solution f is a fixed point of the mapping K defined by setting

$$g = K(f) \Longleftrightarrow g(t) = \alpha_0 + \int_{t_0}^{t} F(f(s), s)ds .$$

We must now find the right domain for K to act on as a contraction mapping.

We start by choosing r such that $S_r(\alpha_0) \subset A$. If $A = W$ we can take $r = \infty$. We also choose δ_0 such that $[t_0 - \delta_0, t_0 + \delta_0] \subset I$ and set $b = \max \{ \| F(\alpha_0, t) \| : |t - t_0| \leq \delta_0 \}$. We take some $\delta \leq \delta_0$ and let V be the Banach space $\mathcal{BC}(I_0, W)$ of bounded continuous functions on $I_0 = (t_0 - \delta, t_0 + \delta)$ into W (see §2.12). Our later calculation shows how small we have to take δ.

We can inject W into V as the constant functions ; in particular, we take $\bar{\alpha}_0$ as the constant function on I_0 with value α_0. Then any $f \in S_r(\bar{\alpha}_0) \subset V$ has its range in A and $F(f(t), t)$ is in V, with $\| F(f(t), t) \| \leq \| F(f(t), t) - F(\alpha_0, t) \| + \| F(\alpha_0, t) \| \leq rc + b$ for all

$t \in I_0$. Therefore K as defined above maps $S_r(\bar{\alpha}_0)$ into V, and

(1) $\|K(\bar{\alpha}_0) - \bar{\alpha}_0\|_\infty \leqq \text{lub}\left\{\left\|\int_{t_0}^{t} F(\alpha_0, s)ds\right\| : |t - t_0| < \delta\right\} \leqq \delta b,$

(2) $\|K(f_1) - K(f_2)\|_\infty \leqq \text{lub}\left\{\left\|\int_{t_0}^{t}(F(f_1(s), s) - F(f_2(s), s)\right\|\right\}$

$\leqq \delta\, \text{lub}\left\{\|F(f_1(s), s) - F(f_2(s), s\|\right\} \leqq \delta C\, \text{lub}\left\{\|f_1(s) - f_2(s)\right\}$

$= \delta c \|f_1 - f_2\|_\infty$.

Thus K is a contraction if $\delta c < 1$ and K has a unique fixed point by the first corollary to the fixed point theorem if also $\delta b < (1 - \delta c)r$. We take any δ satisfying these two inequalities and the theorem is proved.

*§15. Differentiable dependence on parameters.

It is exceedingly important in some applications to know how the solution to the system

$$f'(t) = G(f(t), t), \quad f'(t_1) = \alpha_1$$

varies with the initial point $\langle \alpha_1, t_1 \rangle$. In order to state the problem precisely we fix $I_0 = (t_0 - \delta, t_0 + \delta)$ and $U = S_r(\bar{\alpha}_0) \subset V = \mathcal{BC}(I_0, W)$ as in the previous section and require a solution in U passing through $\langle \alpha_1, t_1 \rangle$, where $\langle \alpha_1, t_1 \rangle$ is near $\langle \alpha_0, t_0 \rangle$. Supposing that a unique solution f exists we then have a mapping $\langle \alpha_1, t_1 \rangle \longrightarrow f$, and it is the continuity and differentiability of this map that we wish to study.

The continuity of the solution as a function of the initial point follows already from the last section and the second corollary to the fixed point theorem. But the differentiability of this mapping depends on §3.11.

<u>Theorem 17.</u> Let A be an open subset of a Banach space W, let I be an open interval in $\mathbb{R}$, and let $F : A \times I \longrightarrow W$ be continuously differentiable, with $dF^1_{\langle \alpha, t \rangle}$ <u>uniformly</u> continuous on $A \times I$. Then for any $\langle \alpha_0, t_0 \rangle \in A \times I$ there exist r, δ, and ϵ such that for any $\langle \alpha_1, t_1 \rangle \in S_\epsilon(\langle \alpha_0, t_0 \rangle)$, the differential equation $d\alpha/dt = F(\alpha, t)$ has a unique solution f on $I_0 = (t_0 - \delta, t_0 + \delta)$ into $S_r(\alpha_0)$ satisfying $f(t_1) = \alpha_1$, and $\langle \alpha_1, t_1 \rangle \longrightarrow f$ is a continuously differentiable mapping

of $S_\epsilon(\langle \alpha_0, t_0\rangle)$ into U.

<u>Proof.</u> We choose r so that $\|dF_{\langle\alpha,t\rangle}\|$ is bounded, say by c, on $S_r(\langle\alpha_0, t_0\rangle)$. Then $\|F(\alpha_1, t) - F(\alpha_2, t)\| \leq c\|\alpha_1 - \alpha_2\|$ for any $\langle\alpha_1, t\rangle, \langle\alpha_2, t\rangle$ in this sphere by the mean value theorem. We next choose δ as in the preceding theorem so that $\|F(\alpha_0, t)\| \leqq b$ on $I_0 = (t_0 - \delta, t_0 + \delta)$, $\delta c < 1$ and $\delta b < (1 - \delta c)r$, and set $V = \mathcal{BC}(I_0, W)$, $U = S_r(\bar{\alpha}_0) \subset V$. We choose $\epsilon < \delta$ and set $B = S_\epsilon(\alpha_0) \subset W$ and $J = S_\epsilon(t_0) = (t_0 - \epsilon, t_0 + \epsilon) \subset \mathbb{R}$. Finally we define $K : U \times (B \times J) \to V$ by setting $g = K(f, \alpha_1, t_1) \iff g(t) = \alpha_1 + \int_{t_1}^{t} F(f(s), s)ds$ for all t in $I_0 = (t_0 - \delta, t_0 + \delta)$.

Now the mapping $h \longrightarrow k$ defined by $k(t) = \int_{t_1}^{t} h(s)ds$ for all $s \in I_0$ is a linear mapping of V into V and bounded by lub $\{|t - t_1| : t \in I_0\} \leqq \delta + \epsilon$. By §3.11 the integrand map $f \longrightarrow h$ defined by $h(s) = F(f(s), s)$ is continuously differentiable on U with differential bounded by c. Combining these two maps we see that $dK^1_{\langle f,\alpha_1,t_1\rangle}$ exists on $U \times B \times J$, is continuous there, and bounded by $C = (\delta + \epsilon)c$. Now $\Delta K^2_{\langle f,\alpha_1,t_1\rangle}(\xi) = \xi$, so that $dK^2 = I$, and $\Delta K^3_{\langle f,\alpha_1,t_1\rangle}(h) = -\int_{t_1}^{t_1+h} F(f(s), s)ds$ so that $dK^3_{\langle f,\alpha_1,t_1\rangle}(h) = hF(f(t_1), t_1)$. Since these three partial differentials are all continuous on $U \times B \times J$, it follows from §3.9 that K is continuously differentiable there.

Finally, $\|\bar{\alpha}_0 - K(\bar{\alpha}_0, \alpha_1, t_1)\|_\infty = \|(\alpha_0 - \alpha_1) - \int_{t_1}^{t} F(\bar{\alpha}_0, s)ds\|_\infty \leqq \|\alpha_0 - \alpha_1\| + \delta b < \epsilon + \delta b$. We now require that ϵ be small enough so that $C = (\delta + \epsilon)c < 1$ and so that $\epsilon + \delta b \leqq (1 - C)r = (1 - (\delta + \epsilon)c)r$. Since we already have $\| dK^1 \| \leqq C$ on $U \times B \times J$ we can now, at last, apply §3.12 and conclude that the solution $f \in U$ is a continuously differentiable function of the initial point $\langle \alpha_1, t_1 \rangle$ in $B \times J$.

§16. Global solutions

The solutions we have found for the differential equation $d\alpha/dt = F(\alpha, t)$ are defined only in sufficiently small neighborhoods of the initial point t_0 and are accordingly called local solutions. Now if we run along to a point $\langle \alpha_1, t_1 \rangle$ near the end of such a local solution and then consider the local solution about $\langle \alpha_1, t_1 \rangle$ it will first of all have to agree on the approach side with our first solution, because there is only one solution going through $\langle \alpha_1, t_1 \rangle$, and secondly it will in general extend further in the other direction than the first solution, so that the two local solutions fit together to make a solution on a larger t interval than either gives separately. We can continue in this way to extend our original solution to what might be called a global solution, made up of a patch work of matching local solutions. These notions are somewhat vague as described above and we now turn to a more precise construction of a global solution.

Given $\langle \alpha_0, t_0 \rangle \subset A \times I$, let $\mathcal{F}$ be the family of all solutions through $\langle \alpha_0, t_0 \rangle$. Thus $g \in \mathcal{F} \Longleftrightarrow g$ is a solution on an interval $J \subset I$, $t_0 \in J$ and $g(t_0) = \alpha_0$.

We show that the union f of all the functions $g \in \mathcal{F}$ is a function. Since each function g is a set of ordered pairs, f is just the union of a family of sets; a point $\langle t, \alpha \rangle$ is in f if and only if $\langle t, \alpha \rangle$ is in some g in $\mathcal{F}$. If f is not a function it contains two distinct pairs $\langle s, \alpha_1 \rangle$ and $\langle s, \alpha_2 \rangle$ with the same first element s. Suppose that $s > t_0$ and that $\langle s, \alpha_1 \rangle \in g_1$, $\langle s, \alpha_2 \rangle \in g_2$. Set

$x = \text{glb}\,\{t : t > t_0 \text{ and } g_1(t) \neq g_2(t)\}$. Then $x < s$ and $g_1(t) = g_2(t)$ for $t \in [t_0, x)$. Since g_1 and g_2 are continuous, $g_1(x) = g_2(x)$. Call this common value α, and choose r such that $S_r(\alpha) \subset A$ and ϵ so that $g_1(t)$ and $g_2(t) \in S_r(\alpha)$ for all $t \in (x - \epsilon, x + \epsilon)$. Also take ϵ small enough so that the differential equation has a unique solution on $(x-\epsilon, x+\epsilon)$ through $\langle \alpha, x \rangle$ with values in $S_r(\alpha)$. Since the restrictions of g_1 and g_2 to $(x-\epsilon, x+\epsilon)$ are such solutions, $g_1 = g_2$ on $(x-\epsilon, x+\epsilon)$ and this contradicts the definition of x as $\text{glb}\,\{t > t_0 : g_1(t) \neq g_2(t)\}$.

Thus f is a function. It is a solution because around any x in its domain f agrees with some $g \in \mathcal{F}$. By the way f was defined we see that f is the unique maximum solution through $\langle \alpha_0, t_0 \rangle$. We have thus proved the following theorem.

<u>Theorem 18</u>. Let $F : A \times I \longrightarrow V$ be a function satisfying the hypotheses of either §3.14 or §3.15. Then through each point $\langle \alpha_0, t_0 \rangle \in A \times I$ there exists a uniquely determined maximal solution to the differential equation $d\alpha/dt = F(\alpha, t)$.

We shall examine the nature of a maximal solution under somewhat stronger hypotheses than we used in §3.14. We shall suppose both that $\|F(\alpha_1, t) - F(\alpha_2, t)\| \leqq c\|\alpha_1 - \alpha_2\|$ for all $\alpha_1, \alpha_2 \in A$ and all $t \in I$, and also that for some α_0, $\|F(\alpha_0, t)\|$ is bounded on I. If $\|F(\alpha_0, t)\| \leqq b$ on I then $\|F(\alpha, t)\| \leqq c\|\alpha - \alpha_0\| + b$ and we have control of the size of F everywhere. In particular $\|F(\alpha_1, t)\|$ is bounded by b_1 on I where $b_1 = c\|\alpha_1 - \alpha_0\| + b$.

It g is any solution through $\langle \alpha_0, t_0 \rangle$ then of course

$g(t) - \alpha_0 = \int_{t_0}^{t} F(g(s), s)ds$ for all t in the domain J of g. Suppose that $t > t_0$. Then $\|g(t) - \alpha_0\| \leqq \int_{t_0}^{t} \|F(g(s), s) - F(\alpha_0, s)\| ds + \int_{t_0}^{t} \|F(\alpha_0, s)\|$ $\leqq c\int_{t_0}^{t} \|g(s) - \alpha_0\| ds + b(t - t_0)$. If the function θ is defined for $t > t_0$ by $\theta(t - t_0) = \|g(t) - \alpha_0\|$, then θ is a continuous, non-negative real-valued function defined on an interval $[0, \ell)$ and satisfying $\theta(x) \leqq c\int_0^x \theta(y)\,dy + bx$ for all $x \in [0, \ell)$. Now it can be shown that this inequality forces the inequality $\theta(x) \leqq \frac{b}{c}(e^{cx} - 1)$ (see). Therefore,

$$\|g(t) - \alpha_0\| \leqq \frac{b}{c}(e^{c|t-t_0|} - 1)$$

for all $t \in J$. This inequality can also be proved directly by a closer examination of the iterated sequence of the fixed point theorem in this special situation.

We now show that the impossibility of extending the maximal f any further can be explained loosely as the fact that f already goes right up to the boundary of the domain $A \times I$ on which F is defined.

Theorem 19. If the right endpoint d of the domain interval J of the maximal solution f is less than the right endpoint of I then $f(t)$ approaches the boundary of A as $t \longrightarrow d$.

Proof. Given $\epsilon > 0$, set $\epsilon_1 = \epsilon/k$, where k is to be determined later. The construction process used in §3.14 to obtain the local solution through $\langle \alpha_1, t_1 \rangle$ involved first choosing any r such that $S_r(\alpha_1) \subset A$ and then choosing the solution domain radius δ to be any number such that $\delta c < 1$ and $\delta b_1 < (1 - \delta c)r$, where in the present situation $b_1 = c\|\alpha_1 - \alpha_0\| + b$, and $\|\alpha_1 - \alpha_0\| = \|f(t_1) - \alpha_0\| \leqq \frac{b}{c}(\ell^{c(d-t_0)} - 1)$. For simplicity we shall take $\delta < 1/2c$ so that the second inequality will be satisfied if $\delta b_1 < r/2$. Substituting for b_1 this becomes $\delta < r/k$ where $k = 2b\,\ell^{c(d-t_0)}$. Now if $t_1 > d - \epsilon_1$ then δ must be $< \epsilon_1$, for the solution through $\langle \alpha_1, t_1 \rangle$ on $(t_1 - \delta, t_1 + \delta)$ cannot extend beyond d, and so $t_1 + \delta \leqq d$, implying that $\delta \leqq d - t_1 < \epsilon_1$. Therefore r/k must be $\leqq \epsilon_1$, and $r \leqq \epsilon_1 k = \epsilon$. Since r was any number such that $S_r(\alpha_1) \subset A$, the fact that r must necessarily be $\leqq \epsilon$ means that $\rho(f(t_1), \partial A) \leqq \epsilon$. Thus for every $\epsilon > 0$ we have found $\epsilon_1 > 0$ such that $|d - t_1| < \epsilon_1 \Rightarrow \rho(f(t_1), \partial A) \leqq \epsilon$. That is, $f(t) \longrightarrow \partial A$ as $t \longrightarrow d$, q. e. d.

It is also possible to make the continuous and differentiable dependence of the solution on its initial value $\langle \alpha_0, t_0 \rangle$ into a global affair. The following is the theorem. We shall not go into its proof here.

<u>Theorem 20.</u> Let f be the maximal solution through $\langle \alpha_0, t_0 \rangle$, with domain J, and let $[a, b]$ be any closed subinterval of J containing t_0. Then there exists $\epsilon > 0$ such that for every $\langle \alpha_1, t_1 \rangle \in S_\epsilon(\langle \alpha_0, t_0 \rangle)$ the domain of the global solution through $\langle \alpha_1, t_1 \rangle$ includes $[a, b]$, and the restriction of this solution to $[a, b]$ is a continuous function of $\langle \alpha_1, t_1 \rangle$. If F satisfies the hypotheses of §3.15 then this dependence is continuously differentiable.

§17. The linear equation

If $F(\alpha, t)$ is linear in α for each fixed t, the hypothesis of a uniform Lipschitz constant c under which we operated in §3.14 becomes $\|F(\alpha, t)\| \leqq c\|\alpha\|$ for every $\langle\alpha, t\rangle \in V \times I$. It now follows that the solution function varies linearly with the initial value α_0.

Theorem 21. Let V be a Banach space, I an open interval in $\mathbb{R}$ and F a continuous mapping of $V \times I$ into V. Suppose also that $F(\alpha, t)$ is linear as a function of α for each $t \in I$ and that there exists a constant c such that $\|F(\alpha, t)\| \leqq c\|\alpha\|$ for all $\langle\alpha, t\rangle \in V \times I$. Fix $t_0 \in I$. Then for each $\beta \in V$ there exists a unique solution $f_\beta : I \longrightarrow V$ to the differential equation $d\alpha/dt = F(\alpha, t)$ such that $f_\beta(t_0) = \beta$, and the mapping $\beta \longmapsto f_\beta$ is a bounded linear mapping of V into $\mathcal{BC}(I, V)$.

Proof. Since $\|F(\beta, t)\| \leqq c\|\beta\|$ for all $t \in I$ the maximal solution f_β through $\langle\beta, t_0\rangle$ satisfies $\|f_\beta(t)\| \leqq \|\beta\| e^{c|t-t_0|}$. The domain of f_β is the whole of I because V has no boundary for the solution to run into, and the above inequality gives $\|f_\beta\|_\infty \leqq k\|\beta\|$ where $k = e^{c(d-a)}$ and $I = (a, d)$. This is the boundedness of $\beta \longmapsto f_\beta$. Finally $(f_\beta + f_\gamma)'(t) = f_\beta'(t) + f_\gamma'(t) = F(f_\beta(t), t) + F(f_\gamma(t), t) = F(f_\beta(t) + f_\gamma(t), t)$. Thus $f_\beta + f_\gamma$ is a solution, and since $(f_\beta + f_\gamma)(t_0) = \beta + \gamma$, it follows that $f_\beta + f_\gamma = f_{\beta+\gamma}$. It follows similarly that $xf_\beta = f_{x\beta}$, and the mapping $\beta \longmapsto f_\beta$ is therefore linear, q. e. d.

§18. The n^{th} order equation

Let $A_1, A_2, \ldots, A_n$ be open subsets of a Banach space W, let I be an open interval in $\mathbb{R}$ and let $G : A_1 \times A_2 \times \cdots \times A_n \times I \longrightarrow W$ be continuous. We consider the differential equation

$$\frac{d^n \alpha}{dt^n} = G(\alpha, d\alpha/dt, \ldots, d^{n-1}\alpha/dt^{n-1}, t)$$

A function $f : J \longrightarrow A_1$ is a solution to this equation if J is an open subinterval of I, f has continuous derivatives on J up to the n^{th} order, $f^{(i-1)}[J] \subset A_i$, $i = 1, \ldots, n$, and

$$f^{(n)}(t) = G(f(t), f'(t), \ldots, f^{(n-1)}(t), t)$$

for $t \in J$. An initial value is now a point

$$\langle \beta_1, \beta_2, \cdots, \beta_n, t_0 \rangle \in A_1 \times \cdots \times A_n \times I .$$

The basic theorem is almost the same as before. To simplify our notation, let $\underline{\alpha}$ be the n-tuple $\langle \alpha_1, \alpha_2, \ldots, \alpha_n \rangle$ in $W^n = V$ and set $\underline{A} = \prod_1^n A_i$. Also set $f^{(\underline{n})} = \langle f, f', \ldots, f^{(n-1)} \rangle$.

Theorem 22. Let $G : \underline{A} \times I \longrightarrow W$ be as above and suppose, in addition, that there is a constant c such that $\| G(\underline{\alpha}, t) - G(\underline{\beta}, t) \| \leqq c \| \underline{\alpha} - \underline{\beta} \|$ for all $\underline{\alpha}, \underline{\beta} \in \underline{A} \times I$. Then for any $\langle \underline{\beta}, t_0 \rangle \in \underline{A} \times I$ there exist $r, \delta > 0$ and a unique function $f : (t_0 - \delta, t_0 + \delta) \longrightarrow S_r(\beta_1) \subset W$ such that f is a solution to the above n^{th} order equation and such that $f^{(\underline{n})}(t_0) = \underline{\beta}$.

Proof. There is an ancient and standard device for reducing a single

n^{th} order equation to a system of first order equations. The idea is to replace the single equation

$$d^n\alpha/dt^n = G(\alpha, d\alpha/dt, \ldots, d^{(n-1)}\alpha/dt^{n-1}, t)$$

by the system of equations

$$d\alpha_1/dt = \alpha_2$$
$$d\alpha_2/dt = \alpha_3$$
$$\vdots$$
$$d\alpha_{n-1}/dt = \alpha_n$$
$$d\alpha_n/dt = G(\alpha_1, \alpha_2, \ldots, \alpha_n, t)$$

Clearly, a system of functions $\alpha_1 = f_1(t), \ldots, \alpha_n = f_n(t)$ satisfies this set of equations if and only if $f_1(t)$ satisfies the original n^{th} order equation.

For us, this device will simply amount to replacing the n^{th} order equation involving the space W by a single first order equation involving the space $V = W^n$. We define the functions $F^i : \underline{A} \times I \longrightarrow W$ by setting $F^i(\underline{\alpha}, t) = \alpha_{i+1}$ for $i = 1, \ldots, n-1$, and $F^n(\underline{\alpha}, t) = G(\underline{\alpha}, t)$. Then $F : \underline{A} \times I \longrightarrow V$ is simply the n-tuplet $\langle F^1, \ldots, F^n \rangle$. That is, $F = \sum_1^n \theta_i \circ F^i$, where θ_i is the injection of W into $V = W^n$ as the i^{th} coordinate subspace. Now the system of first order equations. is entirely equivalent to the single first order equation

$$d\underline{\alpha}/dt = F(\underline{\alpha}, t)$$

And the n-tuplet initial condition $f^{(\underline{n})}(t_0) = \underline{\beta}$ is now just $\underline{f}'(t_0) = \underline{\beta}$, where, of course, $\underline{f} = \langle f_1, \ldots, f_n \rangle$. Finally, the Lipschitz inequality $\| G(\underline{\alpha}, t) - G(\underline{\beta}, t) \| \leqq c \| \underline{\alpha} - \underline{\beta} \|$ implies that $\| F(\underline{\alpha}, t) - F(\underline{\beta}, t) \| \leqq c' \| \underline{\alpha} - \underline{\beta} \|$, where $c' = c+1$. This is because $\| F^i(\underline{\alpha}, t) - F^i(\underline{\beta}, t) \| = \| \alpha^{i+1} - \beta^{i+1} \|$ if $i < n$, and it supposes that we have used the sum norm in $V = W^n : \| \underline{\alpha} \| = \sum_1^n \| \alpha_i \|$. But now our n^{th} order theorem for G has turned into the first order theorem for F and so follows from §3. 14.

*§19. More on the linear equation

Now let X_0 be the Banach space $\mathcal{BC}(I, V)$ of all bounded continuous functions on I into V, under the uniform norm $\|f\|_\infty = \text{lub}\{\|f(t)\| : t \in I\}$, and let X_1 be the subspace (not closed) $\mathcal{BC}^1(I, V)$ of functions having a bounded continuous derivative on I. According to the above theorem, if F is linear in α then the set N of global solutions of the differential equation $d\alpha/dt - F(\alpha, t) = 0$ is a subspace of X_1. Writing the equation in this way suggests pretty clearly that N is just the nullspace of a linear mapping $T : X_1 \longrightarrow X_0$ defined by $d\alpha/dt - F(\alpha, t)$. Thus we define T by setting $g = T(f)$ if and only if $g(t) = f'(t) - F(f(t), t)$. If $f \in \mathcal{BC}^1(I, V) = X_1$ then $T(f) \in \mathcal{BC}(I, V) = X_0$ and the linearity of T follows from the linearity of differentiation and the assumed linearity of F. T is not bounded in the uniform norm because $f \longmapsto f^1$ is not bounded. (However if we fix $t_0 \in I$ and set $|||f||| = \|f(t_0)\| + \|f'\|_\infty$, then $|||f|||$ is a new norm on X_1 and it is not hard to check that T is bounded with respect to it.)

An important property of T is that it is surjective ; every $g \in \mathcal{BC}(I, V)$ is the image under T of at least one $f \in \mathcal{BC}^1(I, V)$. To see this we set $G(\alpha, t) = F(\alpha, t) + g(t)$ and consider the new equation (no longer linear). $d\alpha/dt = G(\alpha, t)$. By our general theory it has a unique solution f through any initial point $\langle \alpha_0, t_0 \rangle$, and, just as for F, its global extension has domain I and is (exponentially) bounded. Because $f'(t) = F(f(t), t) + g(t)$, the derivative f' is also bounded ($\|f'(t)\| \leqq \|F(f(t), t)\| + \|g(t)\| \leqq c\|f(t)\| + \|g(t)\|$ and so $\|f'\|_{\infty} \leqq c\|f\|_{\infty} + \|g\|_{\infty}$).

According to §1.3c, a general solution to the inhomogeneous equation $T(f) = g$ is a linear mapping of X_0 into X_1 which is a right inverse of T, and is completely determined by the complementary subspace M that is its range. Now for each initial point t_0 the subspace $M_{t_0} = \{f \in X_1 : f(t_0) = 0\}$ is such a complement of N in X_1. For if $f \in M_{t_0} \cap N$ then f is a solution to $d\alpha/dt = F(\alpha, t)$ having value 0 at t_0. Since the zero function is also such a solution, and since this solution is unique, we must have $f = 0$. Thus $M_{t_0} \cap N = \{0\}$. To see that $M_{t_0} + N = X_1$ we take any $f \in X_1$ and let g be the unique element of N through $\langle \alpha_0, t_0 \rangle = \langle f(t_0), t_0 \rangle$. If $h = f - g$ then $h(t_0) = 0$. Thus $f = g + h$ where $g \in N$ and $h \in M_{t_0}$, and we have finished proving that M_{t_0} and N are complementary. Altogether we have proved the following theorem.

<u>Theorem 24</u>. Let V be a Banach space, I an open interval in $\mathbb{R}$ and F a continuous map of $V \times I$ into V. Suppose also that $F(\alpha, t)$ is linear as a function of α for each $t \in I$ and that there exists a constant c such that $\|F(\alpha, t)\| \leqq c\|\alpha\|$ for every $\langle \alpha, t \rangle \in V \times I$. Define a mapping T by setting $g = T(f) \iff g(t) = f'(t) - F(f(t), t)$ for all $t \in I$. Then T is a linear mapping of $\mathcal{BC}^1(I, V)$ onto $\mathcal{BC}(I, V)$. For each $t_0 \in I$ the mapping $f \longmapsto f(t_0)$ is an isomorphism of the null-space N(T) onto V. The subspace $M_{t_0} = \{f \in X_1 : f(t_0) = 0\}$ is a complement of N in V, and therefore determines a general solution transformation for the inhomogeneous equation $T(f) = g$.

If J is a closed subinterval of I and we confine our attention to solutions over J, then we can locate our initial condition $\langle \alpha_0, t_0 \rangle$ at an endpoint of J if we wish. If t_0 is the left-hand endpoint of J this merely means that the local solution on $(t_0-\delta, t_0+\delta)$ is now restricted to $[t_0, t_0+\delta)$.

It is also possible to start with the original domain of F in the form $A \times J$, where J is a closed interval. The derivative f' of a parametrized arc f makes perfectly good sense as a one-sided notion; we only have to notice that the limit of the difference quotient $\Delta f/\Delta t$ is now being taken at a point not interior to the domain of Δf. Generally, however, it is permissible to think of a one-sided derivation or differential of f at a boundary point α_0 of its domain as being given by the ordinary derivative or differential of an extension of f over a larger domain having α_0 as an interior point, as we did above in taking the closed interval J as a subinterval of the open interval I.

If J is the closed interval $[a, b]$, then according to our discussion above the subspaces M_a and M_b in particular are complements in X_1 of the nullspace N of our differential operator $T : X_1 \longrightarrow X_0$. We remember that M_a is the nullspace of the coordinate mapping of $X_1 = \mathcal{BC}^1(J, V)$ into V defined by $f \longmapsto f(a)$. The problem that will occupy our attention in the next chapter can be roughly described as follows. Let V_1 and V_2 be closed subspaces of V and let M_a^1 and M_b^2 be their inverse images in X_1 under the a and b coordinate maps. Thus, for example, $M_a^1 = \{f \in X_1 : f(a) \in V_1\}$. Now if V_1 is larger than $\{0\}$ then M_a^1 is too large to be a complement of N in X_1 since it includes the complement M_a as a proper subspace. In the same way M_b^2 is too large. But it may happen that $M = M_a^1 \cap M_b^2$ is a complement of N. If so it is a new type of complement, different from any initial values complement M_{t_0}. In this situation the requirements that $f(a) \in V_1$ and $f(b) \in V_2$ are called boundary conditions, and the problem of solving the equation $T(f) = g$ subject to these conditions is called a boundary value problem. Thus the solution to the boundary value problem is the inverse of the isomorphism $T \upharpoonright M : M \longrightarrow X_0$ where M is the complement of N determined by the boundary conditions.

In the special situation that will concern us (T self-adjoint) the actual computation of this inverse will be accomplished by using an infinite basis for X_0 (and X_1) consisting entirely of "eigen functions" of T. The exact situation cannot be described without more machinery, but it corresponds to finding a basis in a finite dimensional space with respect to which the matrix of T is diagonal.

Chapter IV

§1. Bilinear functionals

Let V and W be any two vector spaces and suppose that $f : V \times W \longrightarrow \mathbb{R}$ is bilinear in the sense that $f(\alpha, \beta)$ is linear in each variable when the other is held fixed. This is an entirely different notion from that of being linear on the product vector space $V \times W$. The function xy is bilinear on $\mathbb{R} \times \mathbb{R}$ but not linear, whereas $x+y$ is linear but not bilinear. Indeed, a linear functional on $V \times W$ cannot be bilinear unless it is 0.

A bilinear $f : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ has an obvious matrix associated with it, namely $t_{ij} = f(\delta^i, \delta^j)$. Then $f(\underline{x}, \underline{y}) = f(\sum_{i=1}^n x_i\delta^i, \sum_{j=1}^m y_j\delta^j)$
$= \sum_{i=1}^n x_i f(\delta^i, \sum_{j=1}^m y_j \delta^j) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j f(\delta^i, \delta^j) = \sum_{i,j} t_{ij} x_i y_j$, giving the value of $f(\underline{x}, \underline{y})$ in terms of the matrix of f and the components of $\underline{x}$ and $\underline{y}$.

If $\{\alpha_i\}_1^n$ and $\{\beta_j\}_1^m$ are bases for the finite dimensional spaces V and W, then relative to these bases a bilinear functional $f : V \times W \rightarrow \mathbb{R}$ has similarly the matrix $t_{ij} = f(\alpha_i, \beta_j)$. If $\xi = \sum_1^n x_i\alpha_i$ and $\eta = \sum_1^m y_j\beta_j$ then $f(\xi, \eta) = \sum_{i,j} t_{ij}x_iy_j$, just as above in the Cartesian case.

We are used to associating matrices with linear transformations, and it might occur to the reader that probably some linear transformation is associated with a bilinear functional. We shall look into this shortly.

If V and W are infinite dimensional we shall assume as usual that they carry norms and that a bilinear functional on $V \times W$ is bounded. Here this means that there exists a constant C such that $|f(\alpha, \beta)| \leqq C\|\alpha\| \|\beta\|$ for all $<\alpha, \beta> \in V \times W$.

The space of all bilinear functionals on $V \times W$ is easily checked to be a vector space. We designate it $V^* \otimes W^*$ and call it the tensor-product of V^* and W^*. We come now to the correspondence with linear transformations.

Theorem 1 . The vector spaces $V^* \otimes W^*$, $\mathrm{Hom}(V, W^*)$ and $\mathrm{Hom}(W, V^*)$ are naturally isomorphic.

Proof. This is nothing but duality again. An element $f \in V^* \otimes W^*$ is a function of two variables, and therefore determines a mapping $\alpha \longmapsto f_\alpha$, where $f_\alpha(\beta) = f(\alpha, \beta)$. But now $f_\alpha \in W^*$. It is linear by the linearity of f as a function of β when α is held fixed, and the inequality $|f_\alpha(\beta)| = |f(\alpha, \beta)| \leqq C\|\alpha\| \|\beta\|$ shows that f_α is bounded and that $\|f_\alpha\| \leqq C\|\alpha\|$. Moreover, the mapping $\alpha \longmapsto f_\alpha$ is an element of $\mathrm{Hom}(V, W^*)$. It is linear by the linearity of f in its first variable. Thus, $f_{(\alpha_1+\alpha_2)}(\beta) = f(\alpha_1 + \alpha_2, \beta) = f(\alpha_1, \beta) + f(\alpha_2, \beta) = f_{\alpha_1} + f_{\alpha_2}(\beta)$, so that $f_{(\alpha_1+\alpha_2)} = f_{\alpha_1} + f_{\alpha_2}$. And it is bounded by C since we saw above that $\|f_\alpha\| \leqq C\|\alpha\|$.

Conversely, if $T \in \mathrm{Hom}(V, W^*)$ then we define $f \in V^* \otimes W^*$ by $f(\alpha, \beta) = (T(\alpha))(\beta)$. The bilinearity of f and its boundedness will be evident to the reader. Also, T is the transformation then defined by f, since $f_\alpha(\beta) = f(\alpha, \beta) = (T(\alpha))(\beta) \Rightarrow f_\alpha = T(\alpha) \Rightarrow$ T is the mapping $\alpha \longmapsto f_\alpha$.

Finally, this bijection between $V^* \otimes W^*$ and $\mathrm{Hom}(V, W^*)$ is an isomorphism. If f_T is the bilinear functional corresponding to T then $f_{(T+S)} = f_T + f_S$, for $f_{(T+S)}(\alpha, \beta) = ((T+S)(\alpha))(\beta) = (T(\alpha) + S(\alpha))(\beta) = (T(\alpha))(\beta) + (S(\alpha))(\beta) = f_T(\alpha, \beta) + f_S(\alpha, \beta)$. And similarly for homogeneity.

The isomorphism of $V^* \otimes W^*$ with $\mathrm{Hom}(W, V^*)$ follows in exactly the same way by reversing the roles of the variables. Thus if g_β is defined by $g_\beta(\alpha) = f(\alpha, \beta)$ then $g_\beta \in V^*$, $\beta \longmapsto g_\beta$ is a linear mapping of W into V^*, etc. We are thus finished with the proof.

Before looking for bases in $V^* \otimes W^*$ we define a bilinear functional $\gamma \otimes \lambda$ from any two functionals $\gamma \in V^*$ and $\lambda \in W^*$ by $(\gamma \otimes \lambda)(\xi, \eta) = \gamma(\xi)\xi(\eta)$. We call $\gamma \otimes \lambda$ the tensor product of the functionals γ and λ and call any bilinear functional having this form <u>elementary</u>. It is not too hard to see that $f \in V^* \otimes W^*$ is elementary if and only if the corresponding $T \in \mathrm{Hom}(V, W^*)$ is a dyad.

If V and W are finite dimensional, with dimensions m and n respectively, then the above isomorphism of $V^* \otimes W^*$ with $\mathrm{Hom}(V, W^*)$ shows that the dimension of $V^* \otimes W^*$ is mn. We now describe the basis determined by given bases in V and W.

<u>Theorem 2.</u> Let $\{\alpha_i\}_1^m$ and $\{\beta_j\}_1^n$ be any bases for V and W, and let their dual bases in V^* and W^* be $\{\mu_i\}_1^m$ and $\{\nu_j\}_1^n$. Then the mn elementary bilinear functionals $\{\mu_i \otimes \nu_j\}$ form the

corresponding basis for $V^* \otimes W^*$.

<u>Proof.</u> Since $\mu_i \otimes \nu_j(\xi, \eta) = \mu_i(\xi)\, \nu_j(\eta) = x_i y_j$, the matrix expansion $f(\xi, \eta) = \sum_{i,j} t_{ij} x_i y_j$ becomes $f(\xi, \eta) = \sum_{i,j} t_{ij}\, \mu_i \otimes \nu_j(\xi, \eta)$ or

$$f = \sum_{i,j} t_{ij}\, \mu_i \otimes \nu_j .$$

The set $\{\mu_i \otimes \nu_j\}$ thus spans $V^* \otimes W^*$. Since it contains the same number of elements (mn) as the dimension of $V^* \otimes W^*$, it forms a basis.

Of course independence can also be checked directly: If $\sum_{i,j} t_{ij}\mu_i \otimes \nu_i = 0$ then for every pair $\langle k, \ell \rangle$, $t_{k\ell} = \sum_{i,j} t_{ij}\, \mu_i \otimes \nu_j (\alpha_k, \beta_\ell) = 0.$

§2. Multilinear functionals.

All of the above considerations generalize to multilinear functionals $f : V_1 \times \cdots \times V_n \longrightarrow \mathbb{R}$. We change notation, just as we do in replacing the traditional $<x, y> \in \mathbb{R}^2$ by $\underline{x} = <x_1, \ldots, x_n> \in \mathbb{R}^n$. Thus we write $f(\alpha_1, \ldots, \alpha_n) = f(\underline{\alpha})$, where $\underline{\alpha} = <\alpha_1, \ldots, \alpha_n> \in V_1 \times \cdots \times V_n$. Our requirement now is that $f(\alpha_1, \ldots, \alpha_n)$ be a linear functional of α_j when α_i is held fixed for each $i \neq j$. The set of all such functionals is a vector space, called the tensor product of the dual spaces $V_1^*, \ldots, V_n^*$, and designated $V_1^* \otimes \cdots \otimes V_n^*$.

Just as before, there are natural isomorphisms between these tensor product spaces and various Hom spaces. For example, $V_1^* \otimes \cdots \otimes V_n^*$ and $\mathrm{Hom}(V_1, V_2^* \otimes \cdots \otimes V_n^*)$ are naturally isomorphic. And there are further isomorphisms of a variety not encountered in the bilinear case. However, it will not be necessary for us to look into these questions.

We define elementary multilinear functionals as before. If $\lambda_i \in V_i^*$, $i = 1, \ldots, n$, and $\underline{\xi} = <\xi_1, \ldots, \xi_n>$, then $(\lambda_1 \otimes \cdots \otimes \lambda_n)(\underline{\xi}) = \lambda_1(\xi_1) \cdots \lambda_n(\xi_n)$.

To keep our notation as simple as possible, and also because it is the case of most interest to us, we shall consider the question of bases only when $V_1 = V_2 = \cdots = V_n = V$. In this case $(V^*)^{ⓝ} = V^* \otimes \cdots \otimes V^*$ (n factors) is called the space of covariant tensors of order n (over V).

If $\{\alpha_i\}_1^m$ is a basis for V and $f \in (V^*)^{ⓝ}$ then we can expand the value $f(\underline{\xi}) = f(\xi_1, \ldots, \xi_n)$ with respect to the basis expansions of the vectors ξ_i just as we did when f was bilinear, but now the result is notationally more complex. If we set $\xi_i = \sum_{j=1}^{m} x_i^j \alpha_j$ for $i = 1, \ldots, n$ and use the linearity of the $f(\xi_1, \ldots, \xi_n)$ in its separate variables one variable at a time, we get

$$f(\xi_1, \ldots, \xi_n) = \sum x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n} f(\alpha_{p_1}, \alpha_{p_2}, \ldots, \alpha_{p_n}),$$

where the sum is taken over all n-tuples $\underline{p} = \langle p_1, \ldots, p_n \rangle$, such that $1 \leqq p_i \leqq m$ for each i from 1 to m. The set of all these n-tuples is just the set of all functions on $\{1, \ldots, n\}$ into $\{1, \ldots, m\}$. We have designated this set $\overline{m}^{\overline{n}}$, using the notation $\overline{n} = \{1, \ldots, n\}$, and the scope of the above sum can thus be indicated in the formula as follows:

$$f(\xi_1, \ldots, \xi_n) = \sum_{\underline{p} \in \overline{m}^{\overline{n}}} x_1^{p_1} \cdots x_n^{p_n} f(\alpha_{p_1}, \ldots, \alpha_{p_n})$$

A strict proof of this formula would require an induction on n, and will be left to the interested reader. At the inductive step he will have to rewrite a double sum $\sum_{\underline{p} \in \overline{m}^{\overline{n}}} \sum_{j \in \overline{m}}$ by the single sum $\sum_{\underline{q} \in \overline{m}^{\overline{n+1}}}$ using the fact that an ordered pair $\langle \underline{p}, j \rangle \in \overline{m}^{\overline{n}} \times \overline{m}$ is equivalent to an (n+1)-tuplet $\underline{q} \in \overline{m}^{\overline{n+1}}$, where $q_i = p_i$ for $i = 1, \ldots, n$ and $q_{n+1} = j$.

If $\{\mu_i\}_1^m$ is the dual basis for V^* and $\underline{q} \in \overline{m}^{\overline{n}}$, let $\mu_{\underline{q}}$ be the elementary functional $\mu_{q_1} \otimes \cdots \otimes \mu_{q_n}$. Thus $\mu_{\underline{q}}(\alpha_{p_1}, \ldots, \alpha_{p_n}) = \pi_1^n \mu_{q_i}(\alpha_{p_i})$ $= 0$ unless $\underline{p} = \underline{q}$, in which case its value is 1. More generally,

$\mu_{\underline{q}}(\xi_1, \ldots, \xi_n) = \mu_{q_1}(\xi_1) \cdots \mu_{q_n}(\xi_n) = x_1^{q_1} \cdots x_n^{q_n}$. Therefore, if we set $t_{\underline{q}} = f(\alpha_{q_1}, \ldots, \alpha_{q_n})$, the general expansion now appears as

$$f(\xi_1, \ldots, \xi_n) = \sum_{\underline{p} \in \overline{m}^{\overline{n}}} t_{\underline{p}} \, \mu_{\underline{p}}(\xi_1, \ldots, \xi_n)$$

or $f = \sum_{\underline{p}} t_{\underline{p}} \mu_{\underline{p}}$, which is the same formula we obtained in the bilinear case, but using more sophisticated notation. The functionals $\{\mu_{\underline{p}} : \underline{p} \in \overline{m}^{\overline{n}}\}$ thus span $V^{*(n)}$. They are also independent. For, if $\sum t_{\underline{p}} \mu_{\underline{p}} = 0$ then for each $\underline{q}$, $t_{\underline{q}} = \sum t_{\underline{p}} \mu_{\underline{p}}(\alpha_{q_1}, \ldots, \alpha_{q_n}) = 0$. We have proved the following theorem.

<u>Theorem 3</u>. The set $\{\mu_{\underline{p}} : \underline{p} \in \overline{m}^{\overline{n}}\}$ is a basis for $(V^*)^{(n)}$. For any $f \in (V^*)^{(n)}$ its coordinate function $\{t_{\underline{p}}\}$ is defined by $t_{\underline{p}} = f(\alpha_{p_1}, \ldots, \alpha_{p_n})$. Thus $f = \sum t_{\underline{p}} \mu_{\underline{p}}$ and $f(\xi_1, \ldots, \xi_n) = \sum t_{\underline{p}} \mu_{\underline{p}}(\xi_1, \ldots, \xi_n) = \sum t_{\underline{p}} x_1^{p_1} \cdots x_n^{p_n}$ for any $f \in V^{*(n)}$ and any $\langle \xi_1, \ldots, \xi_n \rangle \in V^n$.

<u>Corollary.</u> The dimension of $(V^*)^n$ is m^n.

<u>Proof.</u> There are m^n functions in $\overline{m}^{\overline{n}}$, so the basis $\{\mu_{\underline{p}} : \underline{p} \in \overline{m}^{\overline{n}}\}$ has m^n elements.

§3. Permutations

A permutation on a set S is a bijection $f: S \longrightarrow S$. If $\mathcal{S}(S)$ is the set of all permutations on S then, clearly, $\mathcal{S} = \mathcal{S}(S)$ is closed under composition ($\sigma, \rho \in \mathcal{S} \Rightarrow \sigma\circ\rho \in \mathcal{S}$) and taking inverses ($\sigma \in \mathcal{S} \Rightarrow \sigma^{-1} \in \mathcal{S}$). Also, the identity map I is in $\mathcal{S}$ and, of course, the composition operation is associative. Together these statements say exactly that $\mathcal{S}$ is a group under composition. The simplest kind of permutation other than I is one which interchanges a pair of elements of S and leaves every other element fixed. Such a permutation is called a transposition.

We now take S to be the finite set $\overline{n} = \{1,\ldots,n\}$ and set $\mathcal{S}_n = \mathcal{S}(\overline{n})$. We shall assume from elementary group theory the fundamental fact that there exists a homomorphism (denoted 'sgn') of $\mathcal{S}_n$ onto the two element multiplicative group $\{1, -1\}$ such that $\operatorname{sgn}\sigma = -1$ whenever σ is a transposition. Being a homomorphism means preserving the group operation, so we are saying simply that $\operatorname{sgn}(\sigma_1 \circ \sigma_2) = (\operatorname{sgn}\sigma_1)(\operatorname{sgn}\sigma_2)$ for all $\sigma_1, \sigma_2 \in \mathcal{S}_n$. The permutation σ is called even if $\operatorname{sgn}\sigma = 1$ and odd if $\operatorname{sgn}\sigma = -1$. It is not hard to see that any permutation can be expressed as a product of transpositions, and in more than one way. The existence of the mapping sgn is equivalent to the fact that the number of transpositions in a factoring of σ is either always even (if σ is even) or always odd (if σ is odd).

A more elementary fact of group theory that we shall need is that if (in the present context) ρ is a fixed element of $\mathcal{S}_n$ then the mapping $\sigma \longmapsto \sigma \circ \rho$ is a bijection $\mathcal{S}_n \longrightarrow \mathcal{S}_n$. It is surjective because any σ' can be written $\sigma' = (\sigma' \circ \rho^{-1}) \circ \rho$ and it is injective because $\sigma_1 \circ \rho = \sigma_2 \circ \rho \implies (\sigma_1 \circ \rho) \circ \rho^{-1} = (\sigma_2 \circ \rho) \circ \rho^{-1} \implies \sigma_1 = \sigma_2$. Similarly, the mapping $\sigma \longmapsto \rho \circ \sigma$ (ρ fixed) is bijective.

We also need the fact that there are $n!$ elements in $\mathcal{S}_n$. This is the elementary count from high school algebra. In defining an element $\sigma \in \mathcal{S}_n$, $\sigma(1)$ can be chosen in n ways. For each of these choices $\sigma(2)$ can be chosen in $(n-1)$ ways, so that $< \sigma(1), \sigma(2) >$ can be chosen in $n(n-1)$ ways. For each of these choices $\sigma(3)$ can be chosen in $n-2$ ways, etc. Altogether σ can be chosen in $n(n-1)(n-2) \cdots 1 = n!$ ways.

In the sequal we shall often write '$\rho\sigma$' instead of '$\rho \circ \sigma$', just as we occasionally wrote 'ST' instead of '$S \circ T$' for the composition of linear maps.

§4. The permutation representation on $(V^*)^{(n)}$.

If $\underline{\xi} = \langle \xi_1, \dots, \xi_n \rangle \in V^n$ and $\sigma \in \mathcal{S}_n$, then we can "apply σ to $\underline{\xi}$", or "permute the elements of $\langle \xi_1, \dots, \xi_n \rangle$ through σ". We mean, of course, replacing $\langle \xi_1, \dots, \xi_n \rangle$ by $\langle \xi_{\sigma(1)}, \dots, \xi_{\sigma(n)} \rangle$ i.e., replacing $\underline{\xi}$ by $\underline{\xi} \circ \sigma$.

Permuting the variables changes a functional $f \in V^{*(n)}$ into a new such functional. Specifically, given $f \in V^{*(n)}$ and $\sigma \in \mathcal{S}_n$, we define f^σ by

$$f^\sigma(\underline{\xi}) = f(\underline{\xi} \circ \sigma^{-1})$$

The reason for using σ^{-1} instead of σ is, in part, that it gives us the following formula.

Lemma 1. $f^{(\sigma_1 \sigma_2)} = (f^{\sigma_1})^{\sigma_2}$.

Proof. $f^{(\sigma_1 \sigma_2)}(\xi) = f(\underline{\xi} \circ (\sigma_1 \circ \sigma_2)^{-1}) = f(\underline{\xi} \circ (\sigma_2^{-1} \circ \sigma_1^{-1})) =$
$f((\underline{\xi} \circ \sigma_2^{-1}) \circ \sigma_1^{-1}) = f^{\sigma_1}(\underline{\xi} \circ \sigma_2^{-1}) = (f^{\sigma_1})^{\sigma_2}(\underline{\xi})$.

Theorem 4. For each $\sigma \in \mathcal{S}_n$ the mapping T_σ defined by $f \mapsto f^\sigma$ is a linear isomorphism of $(V^*)^{(n)}$ onto itself. The mapping $\sigma \mapsto T_\sigma$ is an anti-homomorphism of the group $\mathcal{S}_n$ into the group of non-singular elements of $\mathrm{Hom}((V^*)^{(n)})$.

Proof. Permuting the variables does not alter the property of multilinearity, so T_σ maps $(V^*)^{(n)}$ into itself. It is linear, since $(af+bg)^\sigma = af^\sigma + bg^\sigma$. And $T_{\rho\sigma} = T_\sigma \circ T_\rho$, because $f^{\rho\sigma} = (f^\rho)^\sigma$.

Thus $\sigma \longmapsto T_\sigma$ preserves products but in the reverse order. This is what is meant by an <u>anti</u>-homomorphism. Finally, $T_{(\sigma^{-1})} \circ T_\sigma = T_{(\sigma\sigma^{-1})} = T_I = I$, so that T_σ is invertible (non-singular, an isomorphism).

The mapping $\sigma \longmapsto T_\sigma$ is a <u>representation</u> (really an <u>anti</u>-representation) of the group $\mathcal{S}_n$ by linear transformations on $(V^*)^{(n)}$.

<u>Lemma 2.</u> Each T_σ carries the basis $\{\mu_{\underline{p}}\}$ into itself and so is a permutation on the basis.

<u>Proof.</u> We have $(\mu_{\underline{p}})^\sigma(\underline{\xi}) = \mu_{\underline{p}}(\underline{\xi} \circ \sigma^{-1}) = \prod_{i=1}^{n} \mu_{p_i}(\xi_{\sigma^{-1}(i)})$. Setting $j = \sigma^{-1}(i)$, and so having $i = \sigma(j)$, this product can be rewritten $\prod_{j=1}^{n} \mu_{p_{\sigma(j)}}(\xi_j) = \mu_{\underline{p}\circ\sigma}(\underline{\xi})$. Thus

$$(\mu_{\underline{p}})^\sigma = \mu_{\underline{p}\circ\sigma},$$

and since $\underline{p} \longmapsto \underline{p}\circ\sigma$ is a permutation on $\overline{m}^n$ we are done.

§5. The subspace $\mathcal{A}^n$ of alternating tensors

Definition. A covariant tensor $f \in V^{*(n)}$ is symmetric $\Longleftrightarrow f^{\sigma} = f$ for all $\sigma \in \mathcal{S}_n$.

If f is bilinear ($f \in V^{*(2)}$) this is just the condition $f(\xi, \eta) = f(\eta, \xi)$ for all $\xi, \eta \in V$.

Definition. A covariant tensor $f \in V^{*(n)}$ is anti-symmetric or alternating $\Longleftrightarrow f^{\sigma} = (\operatorname{sgn} \sigma) f$ for all $\sigma \in \mathcal{S}_n$. Since each σ is a product of transpositions, this can be expressed also as the fact that f just changes sign if two of its arguments are interchanged. In the case of a bilinear functional it is the condition $f(\xi, \eta) = -f(\eta, \xi)$ for all $\xi, \eta \in V$. The set of all symmetric elements of $V^{*(n)}$ is clearly a subspace, as also is the (for us) more important set $\mathcal{A}^n$ of all alternating elements. There is an important linear projection of $V^{*(n)}$ onto $\mathcal{A}^n$ which we now describe.

Lemma . The mapping $f \longrightarrow \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} (\operatorname{sgn} \sigma) f^{\sigma}$ is a projection Ω of $V^{*(n)}$ onto $\mathcal{A}^n$.

Proof. We first check that $\Omega f \in \mathcal{A}^n$ for every $f \in V^{*(n)}$. We have $(\Omega f)^{\rho} = \frac{1}{n!} \sum_{\sigma} \operatorname{sgn} \quad f^{\sigma\rho}$. Now $\operatorname{sgn} \sigma = (\operatorname{sgn} \sigma\rho)(\operatorname{sgn} \rho)$. Setting $\sigma' = \sigma \circ \rho$ and remembering that $\sigma \longmapsto \sigma'$ is a bijection $\mathcal{S}_n \longrightarrow \mathcal{S}_n$ we thus have $(\Omega f)^{\rho} = \frac{(\operatorname{sgn} \rho)}{n!} \sum_{\sigma'} \operatorname{sgn} \sigma' f^{\sigma'} = (\operatorname{sgn} \rho)(\Omega f)$. Thus $\Omega f \in \mathcal{A}^n$.

If f is already in $\mathcal{A}^n$ then $f^\sigma = (\text{sgn}\,\sigma)f$ and $\Omega f = \frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n} f$. Since $\mathcal{S}_n$ has $n!$ elements, $\Omega f = f$. Thus Ω is a projection of $V^{*ⓝ}$ onto $\mathcal{A}^n$.

Lemma 4. $\Omega(f^\rho) = (\text{sgn}\,\rho)\Omega f$.

Proof. The formula for $\Omega(f^\rho)$ is the same as that for $(\Omega f)^\rho$ except that $\rho\sigma$ replaced $\sigma\rho$. The proof is thus the same as the one occurring in the above lemma.

Theorem 5. The vector space $\mathcal{A}^n$ of alternating n-linear functionals over the m-dimensional vector space V has dimension $\binom{m}{n}$.

Proof. If $f\in\mathcal{A}^n$ and $f = \sum_{\underline{p}} t_{\underline{p}}\mu_{\underline{p}}$, then, since $f^\sigma = (\text{sgn}\,\sigma)f$ for any $\sigma\in\mathcal{S}_n$, we have $\sum_{\underline{p}} t_{\underline{p}}\mu_{\underline{p}\circ\sigma} = \sum_{\underline{p}}(\text{sgn}\,\sigma)t_{\underline{p}}\mu_{\underline{p}}$. Setting $\underline{p}\circ\sigma = \underline{q}$, the left sum becomes $\sum_{\underline{q}} t_{\underline{q}\circ\sigma^{-1}}\mu_{\underline{q}}$, and since the basis expansion is unique we must have $t_{\underline{q}\circ\sigma^{-1}} = \text{sgn}\,\sigma\, t_{\underline{q}}$, or $t_{\underline{p}} = (\text{sgn}\,\sigma)t_{\underline{p}\circ\sigma}$ for all $\underline{p}\in\overline{m}^n$. Working backward we see, conversely, that this condition implies $f^\sigma = (\text{sgn}\,\sigma)f$. Thus $f\in\mathcal{A}^n$ if and only if its coordinate function $t_{\underline{p}}$ satisfies the identity:

$$t_{\underline{p}} = (\text{sgn}\,\sigma)\, t_{\underline{p}\circ\sigma} \qquad \text{for all } \underline{p}\in\overline{m}^n.$$

This has many consequences. For one thing, $t_{\underline{p}} = 0$ unless $\underline{p}$ is one-to-one (injective). For if $p_i = p_j$ and σ is the transposition interchanging i and j then $\underline{p}\circ\sigma = \underline{p}$, $t_{\underline{p}} = (\text{sgn}\,\sigma)\,t_{\underline{p}\circ\sigma} = -t_{\underline{p}}$, and so $t_{\underline{p}} = 0$. Since no $\underline{p}$ can be injective if $n > m$, we see that in this case the only element of $\mathcal{A}^n$ is the zero functional. Thus $n > m \Longrightarrow \dim \mathcal{A}^n = 0$.

Now suppose that $n \leqq m$. For any injective $\underline{p}$, the set $\{\underline{p} \circ \sigma : \sigma \in \mathcal{S}_n\}$ consists of all the (injective) n-tuples with the same range set as $\underline{p}$. There are clearly $n!$ of them. Exactly one $\underline{q} = \underline{p} \circ \sigma$ counts off the range set in its natural order, i.e., satisfies $q_1 < q_2 < \cdots < q_n$. We select this unique $\underline{q}$ as the representative of all the elements $\underline{p} \circ \sigma$ having this range. The collection C of these canonical (representative) q's is thus in one-to-one correspondence with the collection of all (range) subsets of $\overline{m} = \{1, \ldots, m\}$ of size n.

Each injective $\underline{p} \in \overline{m}^{\overline{n}}$ is uniquely expressible as $\underline{p} = \underline{q} \circ \sigma$ for some $\underline{q} \in C$, $\sigma \in \mathcal{S}_n$. Thus each f in $\mathcal{A}^n$ is the sum $\sum_{\underline{q} \in C} \left(\sum_{\sigma \in \mathcal{S}_n} \right) t_{\underline{q} \circ \sigma} \mu_{\underline{q} \circ \sigma}$. Since $t_{\underline{q} \circ \sigma} = (\operatorname{sgn} \sigma)\, t_{\underline{q}}$, this sum can be rewritten $\sum_{\underline{q} \in C} t_{\underline{q}} \left(\sum_\sigma (\operatorname{sgn} \sigma)\, \mu_{\underline{q} \circ \sigma} \right) = \sum_{\underline{q} \in C} t_{\underline{q}}\, \nu_{\underline{q}}$ where we have set $\nu_{\underline{q}} = \sum_\sigma (\operatorname{sgn} \sigma)\, \mu_{\underline{q} \circ \sigma} = n!\, \Omega(\mu_{\underline{q}})$.

We are just about done. Each $\nu_{\underline{q}}$ is alternating, being in the range of Ω, and the expansion

$$f = \sum_{\underline{q} \in C} t_{\underline{q}}\, \nu_{\underline{q}}$$

which we have just found to be valid for every $f \in \mathcal{A}^n$ shows that the set $\{\nu_{\underline{q}} : \underline{q} \in C\}$ spans $\mathcal{A}^n$. It is also independent, since

$\sum_{\underline{q} \in C} t_{\underline{q}} \nu_{\underline{q}} = \sum_{\underline{p} \in \overline{m}^{\overline{n}}} t_{\underline{p}} \mu_{\underline{p}}$ and the set $\{\mu_{\underline{p}}\}$ is independent.
It is therefore a basis for $\mathcal{a}^n$.

Now the total number of injective mappings $\underline{p}$ of $\overline{n} = \{1, \ldots, n\}$ into $\overline{m} = \{1, \ldots, m\}$ is $m(m-1) \cdots (m-n+1)$, for the first element can be chosen in m ways, the second in n-1 ways, and so on down through n choices, the last element having $m-(n-1) = m-n+1$ possibilities. We have seen above that the number of these $\underline{p}$'s with a given range is n! Therefore the number of different range sets is $m(m-1) \cdots (m-n+1)/n! = m!/n!(m-n)! = \binom{m}{n}$. Andthis is the number of elements $\underline{q} \in C$.

§6. The determinant.

We saw in §5 that the dimension of the space $\mathcal{A}^m$ of exterior m-forms over an m-dimensional V is $\binom{m}{m} = 1$. Thus, to within scalar multiples there is only one alternating m-linear functional D on $V = \mathbb{R}^m$, and we can adjust the constant so that $D(\delta_1, \ldots, \delta_n) = 1$. This uniquely determined m-form is the determinant functional, and its value $D(\underline{x}_1, \ldots, \underline{x}_m)$ at the m-tuple $< \underline{x}_1, \ldots, \underline{x}_n >$ is the determinant of the matrix $\{x^i_j\}$ (where $\underline{x}_j = \begin{pmatrix} x^1_j \\ x^2_j \\ \vdots \\ x^m_j \end{pmatrix}$).

Now let $\dim V = m$ and let f be any non-zero exterior m-form on V. For any $T \in \operatorname{Hom} V$ the functional f_T defined by $f_T(\xi_1, \ldots, \xi_n) = f(T\xi_1, \ldots, T\xi_n)$ also belongs to $\mathcal{A}^m$. Since $\mathcal{A}^m$ is one-dimensional, $f_T = k_T f$ for some constant k_T. Moreover, k_T is independent of f, since if $g_T = k'_T g$ and $g = cf$ we must have $cf_T = k'_T cf$ and $k'_T = k_T$. This unique constant is called the determinant of T; we shall designate it $\Delta(T)$. Notice that $\Delta(T)$ is defined independently of any basis for V.

Theorem 6. $\Delta(S \circ T) = \Delta(S)\,\Delta(T)$.

Proof. $\Delta(S \circ T)\, f(\xi_1, \ldots, \xi_n) = f((S \circ T)(\xi_1), \ldots, (S \circ T)\xi_n) = f(S(T(\xi_1)), \ldots, S(T(\xi_n))) = \Delta(S)\, f(T(\xi_1), \ldots, T(\xi_n)) = \Delta(T)\Delta(S) f(\xi_1, \ldots, \xi_n)$.

<u>Corollary 1.</u> If s and t are $m \times m$ matrices then $D(s \cdot t) = D(s)\,D(t)$.

<u>Proof.</u> An $m \times m$ matrix t is equivalent to a transformation $T \in \text{Hom}(\mathbb{R}^m)$, the columns of t being the m-tuples $T(\delta_1), \ldots, T(\delta_m)$. Thus $D(t) = D(\underline{t}_1, \ldots, \underline{t}_m) = D(T(\delta_1), \ldots, T(\delta_n)) = \Delta(T) D(\delta_1, \ldots, \delta_n) = \Delta(T)$. The Corollary now follows from the lemma.

<u>Corollary 2.</u> $D(t) = 0$ if and only if t is singular.

<u>Proof.</u> If t is non-singular then t^{-1} exists and $D(t)\,D(t^{-1}) = D(tt^{-1}) = D(I) = 1$. In particular $D(t) \neq 0$. If t is singular some column, say $\underline{t}_1$, is a linear combination of the others, $\underline{t}_1 = \sum_2^m c_i \underline{t}_i$, and $D(\underline{t}_1, \ldots, \underline{t}_m) = \sum_2^m c_i D(\underline{t}_i, \underline{t}_2, \ldots, \underline{t}_m) = 0$ since each term in the sum evaluates D at an m-tuple having two identical elements and so is 0 by the alternating property.

§7. The exterior algebra.

Our final job is to introduce a multiplication operation between alternating n-linear functionals (also now called exterior n-forms). We first extend the tensor product operation that we have used to fashion elementary covariant tensors out of functionals.

Definition. If $f \in (V^*)^{(n)}$ and $g \in (V^*)^{(\ell)}$ then $f \otimes g$ is that element of $(V^*)^{(n+\ell)}$ defined as follows:

$$f \otimes g(\xi_1, \ldots, \xi_{n+\ell}) = f(\xi_1, \ldots, \xi_n)\, g(\xi_{n+1}, \ldots, \xi_{n+\ell})$$

We naturally ask how this operation combines with the projection Ω of $(V^*)^{(n)}$ onto $\mathcal{A}^n$.

Lemma 5. $\Omega(f \otimes g) = \Omega(f \otimes \Omega g) = \Omega(\Omega f \otimes g)$.

Proof. We have $\Omega(f \otimes \Omega g) = \frac{1}{(n+\ell)!} \sum_\sigma (\operatorname{sgn} \sigma)(f \otimes \Omega g)^\sigma$

$$= \frac{1}{(n+\ell)!} \sum_\sigma (\operatorname{sgn} \sigma)\Big(f \otimes \frac{1}{\ell!} \sum_\rho (\operatorname{sgn} \rho) g^\rho\Big)^\sigma$$

$$= \frac{1}{(n+\ell)!\,\ell!} \sum_{\sigma, \rho} (\operatorname{sgn} \sigma)(\operatorname{sgn} \rho)(f \otimes g^\rho)^\sigma$$

We can regard ρ as acting on the full $n+\ell$ places of $f \otimes g$ by taking it as the identity on the first n places. Then $(f \otimes g^\rho)^\sigma = (f \otimes g)^{\rho\sigma}$. Set $\rho\sigma = \sigma'$. For each σ' there are exactly $\ell!$ pairs $<\rho, \sigma>$ with $\rho\sigma = \sigma'$, namely, the pairs $\{<\rho, \rho^{-1}\sigma'> : \rho \in \mathcal{S}_\ell\}$. Thus the above sum is $1/(n+\ell)! \sum_{\sigma'} (\operatorname{sgn} \sigma')(f \otimes g)^{\sigma'} = \Omega(f \otimes g)$. The proof for $\Omega(\Omega f \otimes g)$ is essentially the same.

Definition. $f \wedge g = \binom{n+\ell}{n} \Omega(f \otimes g)$.

Lemma 6. $f_1 \wedge f_2 \wedge \cdots \wedge f_k = \frac{n!}{n_1!\, n_2! \cdots n_k!} \Omega(f_1 \otimes \cdots \otimes f_k)$ where n_i is the order of f_i, $i = 1, \ldots, k$, and $n = \sum_1^k n_i$.

Proof. This is simply an induction, using the definition of the wedge operation $\wedge$ and the above lemma.

Corollary. If $\lambda_i \in V^*$, $i = 1, \ldots, n$ then $\lambda_1 \wedge \cdots \wedge \lambda_n = n!\, \Omega(\lambda_1 \otimes \cdots \otimes \lambda_n)$. In particular, if $q_1 < \cdots < q_n$ and $\{\mu_i\}_1^m$ is a basis for V^*, then $\mu_{q_1} \wedge \cdots \wedge \mu_{q_n} = n!\,\Omega(\mu_{\underline{q}})$ = the basis element $\nu_{\underline{q}}$ of $\mathcal{Q}^n$.

Lemma 7. If $f \in \mathcal{Q}^n$, $g \in \mathcal{Q}^\ell$ then $g \wedge f = (-1)^{\ell n} f \wedge g$. In particular $\lambda \wedge \lambda = 0$ for $\lambda \in V^*$.

Proof. We have $g \otimes f = (f \otimes g)^\sigma$ where σ is the permutation moving each of the last ℓ places over each of the first n places. Thus σ is the product of ℓn transposition, $(\operatorname{sgn} \sigma) = (-1)^{\ell n}$, and $\Omega(g \otimes f) = \Omega((f \otimes g)^\sigma) = (\operatorname{sgn} \sigma)\, \Omega(f \otimes g) = (-1)^{\ell n}\, \Omega(f \otimes g)$. We multiply by $\binom{n+\ell}{n}$ and have the lemma.

Corollary. If $\{\lambda_i\}_1^n \subset V^*$ then $\lambda_1 \wedge \cdots \wedge \lambda_n = 0$ if and only if the sequence $\{\lambda_i\}_1^n$ is dependent.

Proof. If $\{\lambda_i\}$ is independent it can be extended to a basis for V^* and then $\lambda_1 \wedge \cdots \wedge \lambda_n$ is some basic vector $\nu_{\underline{q}}$ of $\mathcal{Q}^n$ by the above corollary. In particular $\lambda_1 \wedge \cdots \wedge \lambda_n \neq 0$.

If $\{\lambda_i\}$ is dependent then one of its elements, say λ_1, is a linear combination of the rest, $\lambda_1 = \sum_2^n c_i \lambda_i$ and $\lambda_1 \wedge \lambda_2 \wedge \cdots \wedge \lambda_n = \sum_{i=2}^n c_i \lambda_i \wedge (\lambda_2 \wedge \cdots \wedge \lambda_n)$. Each of these terms repeats λ_i and so is 0 by the lemma and the above corollary.

Lemma 8. The mapping $\langle f, g \rangle \longmapsto f \wedge g$ is a bilinear mapping of $\mathcal{A}^n \times \mathcal{A}^\ell$ into $\mathcal{A}^{n+\ell}$.

Proof. This follows at once from the obvious bilinearity of $f \otimes g$.

Chapter 5. Scalar products and self-adjoint transformations.

§1. Scalar products.

A <u>scalar product</u> on a vector space V is a real-valued function on $V \times V$ to $\mathbb{R}$, its value at the pair $\langle \xi, \eta \rangle$ ordinarily being designated (ξ, η), such that:

(a) (ξ, η) is linear in ξ when η is held fixed ;

(b) $(\xi, \eta) = (\eta, \xi)$;

(c) $(\xi, \xi) > 0$ if $\xi \neq 0$.

If (c) is replaced by the weaker condition

(c') $(\xi, \xi) \geq 0$ for all $\xi \in V$

then (ξ, η) is called a <u>pseudo scalar product.</u>

Two important examples of scalar products are :

$$(\underline{x}, \underline{y}) = \sum_1^n x_i y_i \quad \text{when } V = \mathbb{R}^n, \quad \text{and}$$

$$(f, g) = \int_a^b f(t)\, g(t)\, dt \quad \text{when } V = \mathcal{C}([a, b]).$$

There are a number of very elementary but important properties which must be established at the very beginning of the theory and which give it its special flavor. These constitute a collection of lemmas and remarks which we shall treat informally as a series of numbered paragraphs.

1. It follows from (a) and (b) that a pseudo-scalar product is also linear in the second variable when the first variable is held fixed, and

therefore fits into the general theory of bilinear functionals as a special kind of symmetric doubly covariant tensor.

2. The Schwarz inequality

$$|(\xi, \eta)| \leqq (\xi, \xi)^{1/2}(\eta, \eta)^{1/2}$$

is valid for any pesudo-scalar product.

Proof. We have $0 \leqq (\xi - t\eta, \xi - t\eta) = (\xi, \xi) - 2t(\xi, \eta) + t^2(\eta, \eta)$ for every $t \in \mathbb{R}$. Since this quadratic in t cannot have distinct roots the usual b^2-$4ac$ formula implies that $4(\xi, \eta)^2 - 4(\xi, \xi)(\eta, \eta) \leqq 0$, which is equivalent to the Schwarz inequality. We can also see this directly. If $(\eta, \eta) > 0$ and if we set $t = (\xi, \eta)/(\eta, \eta)$ in the above quadratic inequality then the resulting expression simplifies to the Schwarz inequality. If $(\eta, \eta) = 0$ then (ξ, η) must also $= 0$ (or else the inequality is clearly false for some t) and now the inequality holds trivially.

3. $\|\xi\| = (\xi, \xi)^{1/2}$ is a pseudonorm, and hence is a norm if and only if (ξ, η) is a scalar product.

Proof. $\|\xi + \eta\|^2 = (\xi + \eta, \xi + \eta) = \|\xi\|^2 + 2(\xi, \eta) + \|\eta\|^2 \leqq \|\xi\|^2 + 2\|\xi\|\,\|\eta\| + \|\eta\|^2$ (by Schwarz) $= (\|\xi\| + \|\eta\|)^2$, proving the triangle inequality. Also, $\|c\xi\| = (c\xi, c\xi)^{1/2} = (c^2(\xi, \xi))^{1/2} = |c|\,\|\xi\|$.

Notice that the Schwarz inequality $|(\xi, \eta)| \leqq \|\xi\|\,\|\eta\|$ is now just the statement that the bilinear functional (ξ, η) is bounded with respect to the scalar product norm.

4. A normed linear space V in which the norm is a scalar product norm is called a <u>pre Hilbert space</u>. If V is complete in this norm it

is a Hilbert space. The two examples of scalar products mentioned earlier give us the norms that we have called 2-norms:

$$\|\underline{x}\|_2 = \left(\sum_1^n x_i^2\right)^{1/2} \text{ for } \underline{x} \in \mathbb{R}^n, \text{ and}$$

$$\|f\|_2 = \left(\int_a^b f^2\right)^{1/2} \text{ for } f \in \mathcal{C}([a, b]) .$$

We know that $\mathbb{R}^n$ is complete in any norm and therefore $\mathbb{R}^n$ is a Hilbert space under the 2-norm. But $\mathcal{C}([a, b])$ is incomplete in the 2-norm (although we shall not show this) and is therefore a pre-Hilbert space but not a Hilbert space in this norm. Remember, however, that $\mathcal{C}([a, b])$ is complete in the uniform norm $\|f\|_\infty$.

Scalar product norms have in some sense the smoothest possible unit spheres. For example, if V has dimension 2 or 3 then the surface of the unit sphere in a scalar product norm is an ellipse or ellipsoid (these notions being independent of coordinate systems !). We shall see this later on.

5. A scalar product is differentiable with respect to its own norm (on $V \times V$, of course). For, if we set $f(\xi, \eta) = (\xi, \eta)$, then $\Delta f_{\langle\alpha,\beta\rangle}(\xi, \eta) = (\alpha + \xi, \beta + \eta) - (\alpha, \beta) = (\xi, \beta) + (\alpha, \eta) + (\xi, \eta)$. But $\ell(\langle\xi,\eta\rangle) = (\xi, \beta) + (\alpha, \eta)$ is linear on $V \times V$. And if we use the maximum norm on $V \times V$, then ℓ is bounded by $\|\alpha\| + \|\beta\|$, for $|\ell(\langle\xi, \eta\rangle)| \leq \|\xi\| \|\beta\| + \|\alpha\| \|\eta\|$ (by Schwarz) $\leq (\|\alpha\| + \|\beta\|) \|\langle\xi, \eta\rangle\|$. Thus $\ell \in V^*$. Finally $|(\xi, \eta)| \leq \|\xi\| \|\eta\| \leq \|\langle\xi, \eta\rangle\|^2 = \mathcal{O}(\|\langle\xi,\eta\rangle\|)$. Thus $\Delta f_{\langle\alpha,\beta\rangle} = \ell + \mathcal{O}$ and $df_{\langle\alpha,\beta\rangle} = \ell$.

6. We need from general bilinear theory the fact that if for each $\beta \in V$ we define $\theta_\beta : V \longrightarrow \mathbb{R}$ by $\theta_\beta(\xi) = (\xi, \beta)$ then $\theta_\beta \in V^*$ and the mapping θ taking β into θ_β is a linear mapping of V into V^*. Of course this can be directly checked in the present context. If (ξ, η) is a scalar product then $\theta_\beta(\beta) = \|\beta\|^2 > 0$ if $\beta \neq 0$ and θ is injective. If V is finite dimensional it follows that <u>θ is an isomorphism.</u> Actually, θ is an isomorphism when V is any Hilbert space, but this is much harder to show when V is infinite dimensional.

7. Two vectors α and β are <u>orthogonal,</u> written $\alpha \perp \beta$, if and only if $(\alpha, \beta) = 0$. Since $\theta_\beta(\alpha) = (\alpha, \beta)$, this is exactly the condition that the functional $\theta_\beta \in V^*$ be orthogonal to α in the old sense. For any two subsets $A, B \subset V$ we write $A \perp B$ if and only if $\alpha \perp \beta$ for every $\langle \alpha, \beta \rangle \in A \times B$, and set $A^\perp = \{\beta \in V : \beta \perp A\}$. $A^\perp$ is always a closed subspace of V. (Check this.) Also,

$$\beta \perp A \Rightarrow \beta \perp L(A) \Rightarrow \beta \perp \overline{L(A)}.$$

8. The Pythagorean theorem : $\alpha \perp \beta \Longleftrightarrow \|\alpha + \beta\|^2 = \|\alpha\|^2 + \|\beta\|^2$. For, $\|\alpha + \beta\|^2 = \|\alpha\|^2 + 2(\alpha, \beta) + \|\beta\|^2$, and this is $\|\alpha\|^2 + \|\beta\|^2$ $\Longleftrightarrow (\alpha, \beta) = 0$. By induction we then have that if the sequence $\{\alpha_i\}_1^n$ is orthogonal, in the sense of being pairwise orthogonal, then $\|\sum_1^n \alpha_i\|^2 = \sum_1^n \|\alpha_i\|^2$.

9. If $\{\phi_i\}_1^n$ is an orthogonal set of non-zero vectors and $\alpha = \sum_1^n a_i\phi_i$ then $a_j = (\alpha, \phi_j)/\|\phi_j\|^2$. (For $(\alpha, \phi_j) = \sum_{i=1}^n a_i(\phi_i, \phi_j) = a_j(\phi_j, \phi_j)$). It follows in particular that $\{\phi_i\}_1^n$, is independent $(\alpha = 0 \Rightarrow (\alpha, \phi_j) = 0 \Rightarrow$

$a_j = 0$, all j). Generally an orthogonal set is normalized, replacing ϕ_i by $\psi_i = \phi_i/\|\phi_i\|$, which has norm 1. If $\{\phi_i\}_1^n$ is thus orthonormal and $\alpha = \sum_1^n a_i\phi_i$ then $a_j = (\alpha, \phi_j)$ for each j.

10. If B is an orthogonal set of non-zero vectors and α is any vector in V then the function $a : B \longrightarrow \mathbb{R}$ defined by $a_\beta = (\alpha, \beta)/\|\beta\|^2$ for all $\beta \in B$ is the Fourier coefficient function of α with respect to B.

If $\{\phi_i\}_1^n$ is a finite orthogonal set, with linear span M, $\alpha \in V$, and $\{a_i\}_1^n$ are the Fourier coefficients of α, then $\alpha - \sum_1^n a_i\phi_i \perp M$. In particular, $V = M \oplus M^\perp$.

Proof. Set $\beta = \alpha - \sum_1^n a_i\phi_i$. Then $(\beta, \phi_j) = (\alpha, \phi_j) - \sum_1^n a_i(\phi_i, \phi_j)$ $= (\alpha, \phi_j) - a_j(\phi_j, \phi_j) = 0$. Thus $\beta \perp \{\phi_i\}_1^n$ and so $\beta \perp L(\{\phi_i\}_1^n) = M$. Since $\sum_1^n a_i\phi_i \in M$, we have $\alpha = \sum a_i\phi_i + \beta \in M + M^\perp$. Finally, if $\alpha \in M \cap M^\perp$ then $\alpha \perp \alpha$, $(\alpha, \alpha) = 0$ and so $\alpha = 0$. Thus $V = M \oplus M^\perp$.

11. If V is a finite dimensional Hilbert space then orthogonal bases can be constructed very easily. We take ϕ_1 as any non-zero vector in V and set $M_1 = L(\{\phi_1\})$. Since $V = M_1 \oplus M_1^\perp$ we can take ϕ_2 as any non-zero vector in $M_1^\perp$. Set $M_2 = L(\{\phi_1, \phi_2\})$, take ϕ_3 in $M_2^\perp$, etc. We thus define an orthogonal (and hence independent) sequence ; therefore $M_n = V$ when $n = \dim V$.

12. A basis $\{\phi_i\}_1^n$ for V is orthonormal if and only if it equals its dual basis under the identification θ of V with V*. For orthonormality is the condition $(\phi_i, \phi_j) = \delta_j^i$, and since $(\phi_i, \phi_j) = \theta_{\phi_j}(\phi_i)$, this says exactly that the functional in V* corresponding to ϕ_j is the j^{th} dual

basis vector. An equivalent statement is that the matrix of θ with respect to these two bases is the identity matrix!

Finally, since $\mathbb{R}^n$ is naturally isomorphic to $(\mathbb{R}^n)^*$, the scalar product effecting this natural identification can be considered to be the standard scalar product on $\mathbb{R}^n$. (See §2.8). It is, of course, $(\underline{x}, \underline{y}) = \sum_{i=1}^n x_i y_i$, the first of our sample scalar products.

13. Suppose that $\{\phi_i\}_1^n$ is orthogonal, $M = L(\{\phi_i\})$, and that $\{a_i\}_1^n$ are the Fourier coefficients of a vector α. Then $\sigma = \sum_1^n a_i\phi_i$ is the best approximation to α in M. For, $\beta \in M \Rightarrow$

$$\|\alpha - \beta\|^2 = \|(\alpha - \sigma) + (\sigma - \beta)\|^2 = \|\alpha - \sigma\|^2 + \|\sigma - \beta\|^2$$

(by 8 and 10) $> \|\alpha - \sigma\|^2$ unless $\beta = \sigma$.

14. If $\{\phi_i\}$ is an infinite orthonormal sequence and $\{a_i\}$ is the Fourier coefficient sequence of $\alpha \in V$, then

$$\sum_1^\infty |a_i|^2 \leq \|\alpha\|^2 \qquad \text{(\underline{Bessel's inequality})}$$

and $\alpha = \sum_1^\infty a_i\phi_i \iff \sum_1^\infty |a_i|^2 = \|\alpha\|^2$ (Parseval's equation).

Proof. Set $\sigma_n = \sum_1^n a_i\phi_i$. Then $\|\sigma_n\|^2 =$

$$\left(\sum_{i=1}^n a_i\phi_i, \sum_{j=1}^n a_j\phi_j\right) = \sum_{i,j} a_i a_j(\phi_i, \phi_j) = \sum_{i=1}^n |a_i|^2 .$$

Therefore $\|\alpha\|^2 = \|\alpha - \sigma_n\|^2 + \|\sigma_n\|^2$ (as in 10) $= \|\alpha - \sigma_n\|^2 + \sum_1^n |a_i|^2$. From this identity we have first that $\sum_1^n |a_i|^2 \leqq \|\alpha\|^2$ for every n, proving Bessel's inequality, and, second, that $\sigma_n \longrightarrow \alpha$ $(\|\alpha - \sigma_n\| \longrightarrow 0) \iff \sum_1^n |a_i|^2 \longrightarrow \|\alpha\|^2$, proving the Parseval identity.

15. An infinite orthonormal sequence is a basis for V if and only if $\alpha = \sum_1^\infty a_i\phi_i$ for each $\alpha \in V$. We now show that if V is a Hilbert space then an orthonormal sequence $\{\phi_i\}$ is a basis if and only if $\{\phi_i\}^\perp = \{0\}$.

Proof. Take any $\alpha \in V$. The sequence $\{\sigma_n\}$ of partial sums (as in 14) is always Cauchy, since $\|\sigma_n - \sigma_m\|^2 = \|\sum_{m+1}^n a_i\phi_i\|^2 = \sum_{m+1}^n |a_i|^2 < \epsilon$ to give $\|\sigma_n - \sigma_m\|^2 < \epsilon$ for all $n, m \geqq N$. If V is complete $\{\sigma_n\}$ converges to some $\sigma \in V$. But then, for every j, $(\alpha - \sigma, \phi_j) = \lim(\alpha - \sigma_n, \phi_j) = 0$, since $\alpha - \sigma_n \perp \phi_j$ if $n \geqq j$. Therefore $\alpha - \sigma \perp \{\phi_i\}$, and if the only vector orthogonal to all ϕ_i is 0, then $\alpha = \sigma = \sum_1^\infty a_i\phi_i$. Thus $\{\phi_i\}$ is a basis. Conversely, if there exists $\alpha \neq 0$ with $\alpha \perp \{\phi_i\}$, then the Fourier coefficients of α are all 0 and α cannot be the sum of its Fourier series, so that in this case $\{\phi_i\}$ is not a basis.

16. If $\beta = \sum_1^\infty b_i\phi_i$ for each β in a dense subset $B \subset V$ then $\{\phi_i\}$ is a basis. Take any $\alpha \in V$. Given ϵ choose $\beta \in B$ such that $\|\alpha - \beta\| < \epsilon/2$ and take n so that $\|\beta - \sigma_n\| < \epsilon/2$ where $\sigma_n = \sum_1^n b_i\phi_i$. Then $\|\alpha - \sigma_n\| < \epsilon$. Therefore $\|\alpha - \sum_1^n a_i\phi_i\| < \epsilon$, by (13). Therefore $\alpha = \sum_1^\infty a_i\phi_i$ and so $\{\phi_i\}$ is a basis.

§2. Self-adjoint transformations.

Definition. If V is a preHilbert space then $T \in \mathrm{Hom}(V, V)$ is self-adjoint if and only if $(T\alpha, \beta) = (\alpha, T\beta)$ for every $\alpha, \beta \in V$. The set of all self-adjoint transformations will be designated SA.

Self-adjointness suggests that T ought to be equal to its own adjoint under the identification θ of V with V^*. We check this now. Since $(\alpha, \beta) = \theta_\beta(\alpha)$, we have $(T\alpha, \beta) = (\alpha, T\beta)$, all $\alpha, \beta \in V \iff \theta_\beta(T\alpha) = \theta_{T\beta}(\alpha)$, all $\alpha, \beta \in V \iff (T^*(\theta_\beta))(\alpha) = \theta_{T\beta}(\alpha)$, all $\alpha, \beta \in V \iff T^*(\theta_\beta) = \theta_{T\beta}$ all $\beta \in V \iff T^* \circ \theta = \theta \circ T \iff T^* = \theta \circ T \circ \theta^{-1}$, which is the asserted identification.

Lemma 1. If V is a finite dimensional Hilbert space and $\{\phi_i\}_1^n$ is an orthonormal basis for V, then $T \in \mathrm{Hom}(V)$ is self-adjoint if and only if the matrix $\{t_{ij}\}$ of T with respect to $\{\phi_i\}$ is symmetric $(t = t^*)$.

Proof. Passing from transformations to their matrices we have that $T^* = \theta \circ T \circ \theta^{-1} \iff t^* = i_n t\, i_n = t$ (since the matrix of θ is the identity i_n).

We can also argue directly. Since $T(\phi_i) = \sum_j t_{ji}\phi_j$ and $(T\alpha, \beta) = (\alpha, T\beta)$ if T is self adjoint, we have $t_{ki} = (T(\phi_i), \phi_k) = (\phi_i, T(\phi_k)) = t_{ik}$, and so t is symmetric.

Lemma 2. SA is a vector subspace of $\mathrm{Hom}(V)$. If $S, T \in SA$ then $S \circ T \in SA \iff S \circ T = T \circ S$. In particular, $T \in SA \Rightarrow T^n \in SA$

and hence $P(T) \in SA$ for any polynomial P.

Proof. It is clear that SA is a subspace. If $S, T \in SA$ then $(S(T\alpha), \beta) = (T\alpha, S\beta) = (\alpha, T(S\beta))$ for all $\alpha, \beta \in V$. But $S \circ T \in SA \iff (S(T\alpha), \beta) = (\alpha, S(T\beta))$. Thus $S \circ T \in SA \iff (\alpha, T(S\beta)) = (\alpha, S(T(\beta))$, all $\alpha, \beta \in V$. $\iff T(S\beta) = S(T\beta)$, all $\beta \in V \iff T \circ S = S \circ T$.

Lemma 3. If $T \in SA$ then $N(T) = R(T)^{\perp}$.

Proof. $\alpha \in N(T) \iff T\alpha = 0 \iff (T\alpha, \beta) = 0$ for all $\beta \in V$ $\iff (\alpha, T\beta) = 0$, all $\beta \iff \alpha \perp R(T)$.

Lemma 4. $T \in SA$ and $x \neq 0 \implies T^2 + x^2 \neq 0$. Therefore, T cannot satisfy a polynomial identeity $P(T) = 0$ where P is an irreducible quadratic polynomial.

Proof. $((T^2 + x^2)\alpha, \alpha) = (T^2\alpha, \alpha) + (x^2\alpha, \alpha) = (T\alpha, T\alpha) + (x\alpha, x\alpha) = \|T\alpha\|^2 + \|x\alpha\|^2 \geqq \|x\alpha\|^2 > 0$ if $x \neq 0$. Therefore $T^2 + x^2 \neq 0$. Any irreducible quadratic $P(t) = t^2 + bt + c$ can be rewritten $(t + y)^2 + x^2$ with $x \neq 0$, so that $P(T) = (T + y)^2 + x^2 \neq 0$.

Lemma 5. $T \in SA$ and $T^n = 0 \Rightarrow T = 0$.

Proof. $T^n = 0 \Rightarrow (T^n\alpha, \beta) = 0$, all $\alpha, \beta \in V$. We suppose of course that $n \geqq 2$. Taking $\beta = T^{n-2}\alpha$ and transferring one T by self adjointness, we obtain $\|T^{n-1}\alpha\|^2 = 0$ for all α, and so $T^{n-1} = 0$. Continuing inductively we finally get $T = 0$.

Theorem 1. If $T \in SA$ and $P(T) = 0$ for some polynomial P then P factors into linear factors (i.e., $P = \prod_1^m P_i$, where $P_i(t) =$

$(t - r_i)^{n_i}$, and $r_i \neq r_j$ for $i \neq j$), the subspaces $N_i = N(P_i)$ are orthogonal, and $P = r_i I$ on N_i.

Proof. This follows from §1.4 D and the above two lemmas. We first factor P as far as possible into relatively prime factors, getting $P = \prod_1^m P_i$, where the P_i are relatively prime and no P_i can be further factored into relatively prime factors. We assume the algebraic fact that the only irreducible polynomials (real coefficients and roots) are <u>linear</u> ($P(t) = t - r$) or <u>quadratic</u> ($P(t) = t^2 + bt + c$). Therefore each P_i above is of the form $P_i = Q_i^{n_i}$ where Q_i is linear or quadratic. It follows from §1.4B that $V = \bigoplus_1^m N_i$, and that $T[N_i] \subset N_i$ for each i. Now the restriction of T to N_i is self-adjoint on N_i as a preHilbert space, and it satisfies $Q_i(T)^{n_i} \neq 0$ there. Since $Q_i(T)$ is self-adjoint (Lemma 2) it follows from the above lemma that $Q_i(T) = 0$ on N_i. Then Lemma 4 implies that Q_i cannot be an irreducible quadratic and therefore is linear. Thus $Q_i(t) = t - r_i$ and $T - r_i = 0$ on N_i. Finally, if $\alpha \in N_i$ and $\beta \in N_j$, $i \neq j$, then $r_i(\alpha, \beta) = (r_i\alpha, \beta) = (T\alpha, \beta) = (\alpha, T\beta) = (\alpha, r_j\beta) = r_j(\alpha, \beta)$. Since $r_i \neq r_j$, $(\alpha, \beta) = 0$. Thus $N_i \perp N_j$ if $i \neq j$. This concludes the proof of the theorem.

This theorem is also a corollary of the much more sophisticated theorem on compact operators in §5.4.

<u>Definition.</u> If $\alpha \neq 0$ and $T(\alpha) = c\alpha$ then α is called an eigenvector (or a proper vector) for T and c is the corresponding eigenvalue (proper value).

Theorem 2. If V is a finite dimensional Hilbert space and $T \in SA$ then V has an orthonormal basis consisting entirely of eigenvectors of T.

Proof. If $n = d(V)$ then we know that Hom(V) has dimension n^2 and therefore the set of $n^2 + 1$ vectors $\{T^i\}_0^{n^2}$ in Hom(V) is dependent. This is exactly the same thing as saying that $P(T) = 0$ for some polynomial P of degree $\leqq n^2$. Now the above theorem gives us $V = \bigoplus_1^m N_i$ with $N_i \perp N_j$ if $i \neq j$ and $T = r_i I$ on N_i. If B_i is any orthonormal basis for N_i and $B = \bigcup_1^m B_i$ then B is an orthonormal basis for V. Since $T(\beta) = r_i\beta$ if $\beta \in B_i$, the theorem is proved.

Although the above theorem constructs the eigenvalues r_i and eigensubspaces N_i in a non-unique way, they nevertheless are unique.

Theorem 3. In the context of the above theorem, if $\alpha \neq 0$ and $T(\alpha) = r\alpha$ for some $r \in \mathbb{R}$ then $\alpha \in N_i$ (and so $r = r_i$) for some i.

Proof. We have $\alpha = \sum_1^n c_i\alpha_i$, with $\alpha_j \in N_j$ for all j. Then $\sum_1^n rc_i\alpha_i = r\alpha = T(\alpha) = T(\sum c_i\alpha_i) = \sum c_i T(\alpha_i) = \sum_1^n c_i r_i \alpha_i$. Therefore $rc_i = r_i c_i$ for all i. Choose j such that $c_j \neq 0$. Then $r = r_j$, and since $r \neq r_i$ for all $i \neq j$ it follows that $c_i = 0$ for all $i \neq j$. Thus $\alpha = c_j\alpha_j \in N_j$.

If V is a finite dimensional vector space and we are given $T \in \mathrm{Hom}\, V$ then we know how to compute related mappings such as T^2 and T^{-1} (if it exists), and vectors $T\alpha$, $T^{-1}\alpha$, etc., by choosing a basis for V and then computing matrix products, inverses (when they exist) and so on. Some of these computations, particularly those related to inverses, can be quite arduous. One enormous advantage of a basis consisting of eigen-vectors for T is that it trivializes all of these calculations.

To see this, let $\{\beta_n\}$ be a basis of V consisting entirely of eigen-vectors for T and let $\{r_n\}$ be the corresponding eigen-values. To compute $T\xi$ we write down the basis expansion for ξ, $\xi = \sum_1^n x_i\beta_i$ and then $T\xi = \sum_1^n r_i x_i \beta_i$. T^2 has the same eigen-vectors, but with eigen-values $\{r_i^2\}$. Thus $T^2\alpha = \sum_1^n r_i^2 x_i \beta_i$. T^{-1} exists if and only if no $r_i = 0$, in which case it has the same eigen-vectors with eigen-values $\{1/r_i\}$. Thus $T^{-1}\xi = \sum_1^n (x_i/r_i)\beta_i$. If $P(t) = \sum_0^m a_n t^n$ is any polynomial, then $P(T)$ takes β_i into $P(r_i)\beta_i$. Thus $P(T)\xi = \sum_0^m P(r_i)\, x_i\beta_i$. By now the point should be amply clear.

The additional value of orthonormality in a basis is already clear from the last section. Basically it enables us to compute the coefficients $\{x_i\}$ of ξ by scalar products : $x_i = (\xi, \beta_i)$.

Another virtue of an orthonormal basis of eigen-values of a self-adjoint T is that in certain infinite dimensional situations, where V is a pre-Hilbert space but not a Hilbert space, this is the easiest way of showing that V has a basis. See § 6 .

§3. Orthogonal transformations

Assuming that V is a Hilbert space and that therefore $\theta: V \longrightarrow V^*$ is an isomorphism, we can of course replace the adjoint $T^* \in \text{Hom } V^*$ of any $T \in \text{Hom } V$ by the corresponding transformation $\theta^{-1} \circ T^* \circ \theta \in \text{Hom } V$. In Hilbert space theory it is this mapping that is called the adjoint of T. Then, exactly as in our discussion of a self-adjoint T, we see that

$$(T\alpha, \beta) = (\alpha, T^*\beta) \quad \text{for all } \alpha, \beta \in V,$$

and that T^* is uniquely defined by this identity. And T is self-adjoint $\Longleftrightarrow T = T^*$.

Another very important type of transformation on a Hilbert space is one that preserves the scalar product.

Definition. A transformation $T \in \text{Hom } V$ is orthogonal if and only if $(T\alpha, T\beta) = (\alpha, \beta)$ for all $\alpha, \beta \in V$.

By the basic adjoint identity above this is entirely equivalent to $(\alpha, T^*T\beta) = (\alpha, \beta)$, for all α, β, and hence to $T^*T = I$. An orthogonal T is injective, since $\|T\alpha\|^2 = \|\alpha\|^2$, and therefore invertible if V is finite dimensional. Whether T is finite dimensional or not, if T is invertible then the above condition becomes $T^* = T^{-1}$.

If $T \in \text{Hom } \mathbb{R}^n$ the matrix form of the equation $T^*T = I$ is of course $t^*t = i_n$, and if this is written out it becomes

$$\sum_{k=1}^{n} t_{ki}t_{kj} = \delta^i_j \quad \text{for all } i, j$$

which simply says that the columns of t form an orthonormal set (and hence a basis) in $\mathbb{R}^n$. We thus have:

Theorem 4. A transformation $T \in \operatorname{Hom} \mathbb{R}^n$ is orthogonal if and only if the image of the standard basis $\{\delta^i\}_1^n$ under T is another orthonormal basis (with respect to the standard scalar product).

The necessity of this condition is, of course, obvious from the scalar-product-preserving definition of orthogonality, and the sufficiency can also be checked directly using the basis expansions of α and β.

We can now state the eigen basis theorem in different terms. By a diagonal matrix we shall mean a matrix which is zero everywhere except on the main diagonal.

Theorem 5. Let $t = \{t_{ij}\}$ be a symmetric $n \times n$ matrix. Then there exists an orthogonal $n \times n$ matrix b such that $b^{-1}tb$ is a diagonal matrix.

Proof. Since the transformation $T \in \operatorname{Hom} \mathbb{R}^n$ defined by t is self-adjoint there exists an orthonormal basis $\{\underline{b}^i\}_1^n$ of eigenvectors of T, with corresponding eigen values $\{r_i\}_1^n$. Let B be the orthogonal transformation defined by $B(\delta^j) = \underline{b}^j$, $j = 1, \ldots, n$. (The n-tuple $\underline{b}^j$ are the columns of the matrix $b = \{b_{ij}\}$ of B.) Then $(B^{-1}\ T\ B)(\delta^j) = r_j\delta^j$. Since $(B^{-1} \circ T \circ B)(\delta^j)$ is the j^{th} column of $b^{-1}tb$, we see that $s = b^{-1}tb$ is diagonal, with $s_{jj} = r_j$.

This theorem simply recognizes that since the standard basis in $\mathbb{R}^n$ is not an eigen basis for T (in general) the change of basis matrix b taking the standard basis into an eigenbasis results in a matrix for T which displays its eigenvectors and eigenvalues.

For later applications we are also going to want the following result.

Theorem 6 . Any invertible $T \in \text{Hom } V$ on a finite dimensional Hilbert space V can be expressed in the form $T = RS$, where R is orthogonal and S is self-adjoint.

Proof. For any T, T^*T is self-adjoint since $(T^*T)^* = T^*T^{**} = T^*T$. Let $\{\phi_i\}_1^n$ be an orthonormal eigen-basis and $\{r_i\}_1^n$ the corresponding eigen-values of T^*T. Then $0 < \|T\phi_i\|^2 = (T^*T\phi_i, \phi_i) = (r_i\phi_i, \phi_i) = r_i$, for each i. Since all the eigen-values of T^*T are thus positive, we can define a positive square root $S = (T^*T)^{1/2}$ by $S\phi_i = (r_i)^{1/2}\phi_i$, $i = 1, 2, \ldots, n$. It is clear that $S^2 = T^*T$ and that S is self-adjoint.

Then $A = ST^{-1}$ is orthogonal, for $(ST^{-1}\alpha, ST^{-1}\beta) = (T^{-1}\alpha, S^2T^{-1}\beta)$ $= (T^{-1}\alpha, T^*TT^{-1}\beta) = (T^{-1}\alpha, T^*\beta) = (T\ T^{-1}\alpha, \beta) = (\alpha, \beta)$. Since $T = A^{-1}S$, we set $R = A^{-1}$ and have the theorem.

It is not hard to see that the above factorization of T is unique. Also, by starting with TT^*, we can express T in the form $T = SR$, where S is self-adjoint and positive, and R is orthogonal.

Corollary. Any non-singular $n \times n$ matrix t can be expressed as $t = udv$, where u and v are orthogonal and d is diagonal.

Proof. From the theorem we have $t = rs$, where r is orthogonal and s is symmetric. By §3, $s = bdb^{-1}$ where d is diagonal and b is orthogonal. Thus $t = rs = (rb)db^{-1} = udv$, where $u = rb$ and $v = b^{-1}$ are both orthogonal.

§4. Compact transformations

The preceding theory breaks down when V is an infinite dimensional Hilbert space. A self-adjoint transformation T does not in general have enough eigenvectors to form a basis for V, and a more sophisticated analysis, allowing for a "continuous spectrum" as well as "discrete spectrum" is necessary. This enriched situation is the reason for the need of further study of Hilbert space theory at the graduate level, and is one of the sources of complexity in the mathematical structure of quantum mechanics.

However, there is one very important special case in which the eigenbasis theorm is available, and we shall spend the rest of the chapter studying this special kind of self-adjoint T and its occurrence in the classical theory of ordinary differential equations.

Definition. Let V and W be any normed linear spaces and let S be the unit sphere in V. A transformation $T \in \mathrm{Hom}(V, W)$ is compact if and only if the closure of $T[S]$ is a sequentically compact subset of W.

Theorem 7. Let V be any preHilbert space and let $T \in \mathrm{Hom}\, V$ be self-adjoint and compact. Then there exists a uniquely determined sequence $\{r_n\} \subset \mathbb{R}$ and an orthonormal sequence $\{\phi_n\} \subset V$ such that:

(1) $|r_n| \downarrow 0$ or $\{r_n\}$ and $\{\phi_n\}$ are finite;

(2) $T(\phi_n) = r_n\phi_n$, all n;

(3) $\{\phi_n\}$ is a basis for the range of T.

Proof. Set $m = \|T\|$ and choose α_n such that $\|\alpha_n\| = 1$ and $\|T(\alpha_n)\| \longrightarrow m$. Since T is compact we can suppose (passing to a subsequence if necessary) that $\{T(\alpha_n)\}$ converges, say $T(\alpha_n) \longrightarrow \beta$. Then $\|\beta\| = \lim \|T\alpha_n\| = m$. But now $\|T(\beta)\| \leqq \|T\| \|\beta\| \leqq m^2$. Also $m^2 = (\beta, \beta) = \lim(\beta, T(\alpha_n)) = \lim(T\beta, \alpha_n) \leqq \|T\beta\|$ (by Schwarz and the fact that $\|\alpha_n\| = 1$). Thus $\|T(\beta)\| = m^2$. But then $\|(T^2 - m^2)\beta\|^2 = ((T^2 - m^2)(\beta), (T^2 - m^2)(\beta)) = \|T^2\beta\|^2 - 2m^2\|T\beta\|^2 + m^4\|\beta\|^2 \leqq \|T^2\|^2\|\beta\|^2 - m^6 \leqq m^6 - m^6 = 0$. Thus $T^2\beta = m^2\beta$. Set $\alpha = \beta/\|\beta\|$.

We have thus found a vector α such that $\|\alpha\| = 1$ and $0 = (m^2 - T^2)(\alpha) = (m - T)(m + T)(\alpha)$. Then either $(m + T)(\alpha) = 0$, in which case $T(\alpha) = -m\alpha$ or $(m + T)(\alpha) = \gamma \neq 0$ and $(m - T)\gamma = 0$, in which case $T\gamma = m\gamma$. Thus there exists a vector ϕ_1 (α or $\gamma/\|\gamma\|$) such that $\|\phi_1\| = 1$ and $T(\phi_1) = r_1\phi_1$, where $|r_1| = m$.

For notation consistency we set $m_1 = m$, $V_1 = V$, and now set $V_2 = \{\phi_1\}^\perp$. Then $T[V_2] \subset V_2$, since $\alpha \perp \phi_1 \Rightarrow (\alpha, \phi_1) = 0 \Rightarrow 0 = (\alpha, T\phi_1) = (T\alpha, \phi_1) \Rightarrow T\alpha \perp \phi_1$. Thus $T \restriction V_2$ is compact and self-adjoint, and if $m_2 = \|T \restriction V_2\|$ there exists ϕ_2 with $\|\phi_2\| = 1$ and $T(\phi_2) = r_2\phi_2$ where $|r_2| = m_2$. We continue inductively, obtaining an orthonormal sequence $\{\phi_n\} \subset V$ and a sequence $\{r_n\} \subset \mathbb{R}$ such that $T\phi_n = r_n\phi_n$ and $|r_n| = \|T \restriction V_n\|$, where $V_n = \{\phi_1, \ldots, \phi_{n-1}\}^\perp$.

We suppose for the moment, this being the most interesting case, that $r_n \neq 0$ for all n. Then we claim that $|r_n| \longrightarrow 0$. For $|r_n|$ is decreasing in any case, and if it does not converge to 0 then there exists $b > 0$ such that $|r_n| \geqq b$ for all n. Then $\|T(\phi_i) - T(\phi_j)\|^2 =$

$\| r_i\phi_i - r_j\phi_j \|^2 = \| r_i\phi_i \|^2 + \| r_j\phi_j \|^2 = r_i^2 + r_j^2 \geqq 2b^2$ for all $i \neq j$, and the sequence $\{T(\phi_i)\}$ can have no convergent subsequence, contradicting the compactness of T. Therefore $|r_n| \downarrow 0$.

Finally, if $\beta = T(\alpha)$ and $\{b_n\}$ and $\{a_n\}$ are the Fourier coefficients of β and α, then $b_n = (\beta, \phi_n) = (T(\alpha), \phi_n) = (\alpha, T(\phi_n)) = (\alpha, r_n\phi_n) = r_n(\alpha, \phi_n) = r_n a_n$, and $\| \beta - \sum_1^n b_i\phi_i \| = \| T(\alpha - \sum_1^n a_i\phi_i)\| \leqq |r_{n+1}| \| \alpha - \sum_1^n a_i\phi_i \|$ (since $\alpha - \sum_1^n a_i\phi_i \in V_{n+1}$ by § 1.10) $\leqq |r_{n+1}| \|\alpha\| \longrightarrow 0$. Thus $\beta = \sum_1^\infty b_i\phi_i$ for every β in R(T) and $\{\phi_i\}$ is a basis for R(T). Also $N(T) = R(T)^\perp = \{\phi_i\}^\perp = \bigcap_1^\infty V_i$.

If some $r_i = 0$ then there is a first n such that $r_n = 0$. In this case $\| T \restriction V_n \| = |r_n| = 0$, so that $V_n \subset N(T)$, whereas $\phi_i = T(\phi_i)/r_i \in R(T)$ for $i < n$, so that $\{\phi_1, \ldots, \phi_{n-1}\} \subset R(T) \subset N(T)^\perp$. Since $V = \{\phi_1, \ldots, \phi_{n-1}\} \oplus V_n$, it follows that $V_n = N(T)$ and $\{\phi_i\}_1^{n-1} = R(T)$.

§5. Equicontinuity

Let A and B be metric spaces. A subset $\mathcal{F} \subset B^A$ is equicontinuous at $p_0 \in A \iff (\forall \epsilon > 0)(\exists \delta > 0)(\forall p \in A)(\forall f \in \mathcal{F})(\rho(p, p_0) < \delta \implies \rho(f(p), f(p_0)) < \epsilon)$. That is all the functions in $\mathcal{F}$ are continuous at p_0, and the same δ works for them all. $\mathcal{F}$ is uniformly equicontinuous on A $\iff (\forall \epsilon > 0)(\exists \delta > 0)(\forall p, q \in A)(\forall f \in \mathcal{F})(\rho(p, q) < \delta \implies \rho(f(p), f(q)) < \epsilon)$.

Theorem 8. If A and B are totally bounded metric spaces and if $\mathcal{F}$ is a uniformly equicontinuous subfamily of B^A then $\mathcal{F}$ is totally bounded in the uniform metric.

Proof. Given $\epsilon > 0$, choose δ so that for all $f \in \mathcal{F}$ and all $p_1, p_2 \in A$, $\rho(p_1, p_2) < \delta \implies \rho(f(p_1), f(p_2)) < \epsilon/4$. Let D be a finite subset of A which is δ-dense in A and let E be a finite subset of B which is $\epsilon/4$-dense in B. Let F be the set E^D of all functions on D into E. F is of course finite; in fact, $\#F = n^m$ where $m = \#D$ and $n = \#E$. If $f \in \mathcal{F}$ and $p \in D$ then by the definition of E there exists $q \in E$ such that $\rho(q, f(p)) < \epsilon/4$. If we choose such a $q \in E$ for each $p \in D$ we have constructed a function $g \in F$ such that $\rho(f(p), g(p)) < \epsilon/4$ for every $p \in D$. For each $g \in F$ choose one $f_g \in \mathcal{F}$ if there is one such that $\rho(g, f_g) < \epsilon/4$ on D, and let $\mathcal{F}_0$ be the subset of $\mathcal{F}$ so chosen. Clearly, $\#\mathcal{F}_0 \leqq \#F = n^m$. We claim that $\mathcal{F}_0$ is ϵ-dense in $\mathcal{F}$. To see this take any $f \in \mathcal{F}$ and construct $g \in F$ as above such that $\rho(g, f) < \epsilon/4$ on D. Then f_g exists, and we claim that $\rho(f, f_g) < \epsilon$. Since $\rho(g, f_g) < \epsilon/4$ on D and

$\rho(f, g) < \epsilon/4$ on D it follows that $\rho(f(p), f_g(p)) < \epsilon/2$ for each $p \in D$. Then for any $p' \in A$ we have only to choose $p \in D$ such that $\rho(p', p) < \delta$ and have $\rho(f(p'), f_g(p')) \leqq \rho(f(p'), f(p)) + \rho(f(p), f_g(p)) + \rho(f_g(p), f_g(p')) \leqq \epsilon/4 + \epsilon/2 + \epsilon/4 = \epsilon$.

§6. The Sturm-Liouville Problem

In our theory of linear differential equations we considered a function $F : W \times I \longrightarrow W$ which was continuous and linear in the first variable for each fixed value of $t \in I$. More generally, for the n-th order case, we had $G : W^n \times I \longrightarrow W$ such that $G(\alpha_1, \ldots, \alpha_n, t) = G(\underline{\alpha}, t)$ is a linear in the n-tuple variable $\underline{\alpha} = <\alpha_1, \ldots, \alpha_n>$ for each fixed t.

We now consider the special case $W = \mathbb{R}$. For each fixed t, G is now a linear map of $\mathbb{R}^n$ into $\mathbb{R}$, i.e., an element of $(\mathbb{R}^n)^*$, and its coordinate set with respect to the standard basis is an n-tuple $\underline{k} = <k_1, \ldots, k_n>$ (its image in $\mathbb{R}^n$ under the natural isomorphism). Since the linear map varies continuously with t the n-tuple $\underline{k}$ varies continuously with t. Thus, when we take t into account we have an n-tuple $\underline{k}(t) = <k_1(t), \ldots, k_n(t)>$ of continuous real-valued functions on I such that

$$G(x_1, \ldots, x_n, t) = \sum_{i=1}^{n} k_i(t)\, x_i \,.$$

As in our earlier general discussion the solution space N of the n-th order differential equation

$$\frac{d^n\alpha}{dt^n} = G(\alpha, \ldots, d^{n-1}\alpha/dt^{n-1}, t)$$

is just the null space of the linear transformation $T : \mathcal{C}^n(I : \mathbb{R}) \longrightarrow \mathcal{C}^0(I, \mathbb{R})$ defined by

$$(Tf)(t) = f^{(n)}(t) - k_n(t)f^{(n-1)}(t) - \ldots - k_1(t)\, f(t) \,.$$

If we shift indices to coincide with the order of the derivative and let $f^{(n)}$ also have a coefficient function, our n-th order linear differential operator T appears as

$$Tf = c_n(t)f^{(n)}t + \dots + c_0f(t) .$$

In what is called the regular case the function $c_n(t)$ is never zero, and so can be divided out to give the form we have studied. We shall also restrict our attention to a closed interval $[a, b] \subset I$, so that we can consider initial conditions at the endpoints if we wish.

Our present study of T will be by means of the scalar product $(f, g) = \int_a^b f(t)g(t)\,dt$.

The general problem is the usual one of solving $Tf = g$ by finding a right inverse S of T. And, also as usual, we find S by finding a complement M of N(T). Now, however, it turns out that if T is "formally self-adjoint" then suitable choices of M will make the associated right inverses S self-adjoint and compact, and the eigen-vectors of S, computed as those solutions of the homogeneous equation $Tf - rf = 0$ which lie in M, then allow the same (relatively) easy handling of S by virtue of § 4 that they gave us earlier in the finite dimensional situation (§ 2).

We first consider the notion of 'formal adjoint' for a differential operator T. The ordinary formula for integration by parts,

$$\int_a^b f'g = fg\Big]_a^b - \int_a^b fg' ,$$

allows the derivatives of f occurring in the scalar product (Tf, g) to be shifted one at a time to g. At the end, f is undifferentiated and g is acted on by a certain n-th order linear differential operator R. The endpoint evaluations, like the above $fg\Big]_a^b$, that accumulate step by step can be described as

$$B(f, g)\Big]_a^b = \sum_{0 \le i+j < n} k_{ij}(x) f^{(i)}(x) g^{(j)}(x) \Big]_a^b ,$$

where the coefficient functions $k_{ij}(x)$ are linear combinations of the coefficient functions $c_i(x)$ and their derivatives. Thus

$$(Tf, g) = (f, Rg) + B(f, g)\Big]_a^b .$$

The operator R is called the formal adjoint of T and if $R = T$ we say that T is formally self-adjoint.

Every application of the integration by parts formula introduces a sign change, and the reader may be able to see that the leading coefficient of R is $(-1)^n c_n(t)$. Assuming this, we see that a necessary condition for formal self-adjointness is that n be even, so that R and T have the same first terms.

From now on we are going to consider only the second order case. However, almost everything that we are going to do works perfectly well for the general case, the price of generality being only additional notational complexity.

We now compute the formal adjoint of the second order operator $Tf = c_2 f'' + c_1 f' + c_0 f$. We have $(Tf, g) = \int_a^b c_2 f'' g + \int_a^b c_1 f' g + \int_a^b c_0 fg$,

$$\int_a^b c_1 f' g = c_1 fg\Big]_a^b - \int_a^b f(c_1 g)'$$

$$\int_a^b c_2 f'' g = c_2 f' g\Big]_a^b - \int_a^b f'(c_2 g)'$$

$$= \Big[c_2 f' g - f(c_2 g)'\Big]_a^b + \int_a^b f(c_2 g)'' \quad \text{, giving}$$

$$(f, R.g) = \int f((c_2 g)'' - (c_1 g)' + (c_0 g)), \text{ and}$$

$$B(f, g) = c_2(f'g - g'f) + (c_1 - c_2')fg.$$

Thus $R.g = c_2 g'' + (2c_2' - c_1)g' + (c_2'' - c_1' + c_0)g$, and $R = L \iff 2c_2' - c_1 = c_1$ (and $c_2'' - c_1' = 0$) $\iff c_2' = c_1$. Thus T is formally self-adjoint if and only if

$Tf = c_2 f'' + c_2' f' + c_0 f = (c_2 f')' + c_0 f$, in which case $B(f, g) = c_2(f'g - g'f)$.

From now on we shall suppose that T is formally self-adjoint.

Supposing that we have found a solution of the inhomogeneous equation $Tf = g$ in the form of a complement M of $N(T)$ in $\mathcal{C}^n([a, b])$, we next enquire as to the self-adjointness of the inverse operator $S : \mathcal{C}^0 \longrightarrow M$ (considered as an element of Hom $\mathcal{C}^0$). For any $u, v \in \mathcal{C}^0$ set $f = Su$ and $g = Sv$, so that $f, g \in M$ and $u = Tf$, $v = Tg$. Then $(u, Sv) = (Tf, g) = (f, Tg) + B(f, g)\Big]_a^b = (Su, v) + B(f, g)\Big]_a^b$. Thus S is self-adjoint if and only if $f, g \in M \Rightarrow B(f, g)\Big]_a^b = 0$. Of course we must also show that S is bounded, but in the present case it will actually turn out to be compact.

The only kind of complement we know about so far is the "single point" subspace $M_{t_0} = \{f \in \mathcal{C}^2 : f(t_0) = f'(t_0) = 0\}$. It is clear, however, that if f and g satisfy only two such conditions at a single point, then in general $B(f, g)]_a^b$, involving as it does the values of f and g and their first derivatives at both a and b, will not be zero. We therefore try for a subspace M defined by two conditions which involve both endpoints. Each condition will be the requirement that a linear functional ℓ on $\mathcal{C}^2$, whose value $\ell(f)$ depends on the values of f and f' at a and b, be zero. Such a functional ℓ will be the composition of the linear mapping $f \longrightarrow \langle f(a), f'(a), f(b), f'(b)\rangle$ of $\mathcal{C}^2$ into $\mathbb{R}^4$ with an element of $(\mathbb{R}^4)^*$, and therefore will be of the form $\ell(f) = k_1f(a) + k_2f'(a) + k_3f(b) + k_4f'(b)$. Thus $M = \{f \in \mathcal{C}^2 : \ell_1(f) = 0 \text{ and } \ell_2(f) = 0\}$ where ℓ_1 and ℓ_2 are both of the above form. The boundary condition $\ell_1(f) = \ell_2(f) = 0$ is called a self-adjoint if $f, g \in M \Rightarrow B(f, g)]_a^b = c_2(f'g - g'f]_a^b = 0$. We know then that the right inverse S defined by M (if M is a complement of N) is self-adjoint.

Instead of trying to find out what self-adjointness implies about an otherwise arbitrary pair ℓ_1 and ℓ_2, we shall confine ourselves to a few special boundary conditions that obviously are self-adjoint. We list them below.

1. $f \in M \iff f(a) = f(b) = 0$ (i.e., $\ell_1(f) = f(a)$ and $\ell_2(f) = f(b)$).
2. $f \in M \iff f'(a) = f'(b) = 0$.
3. More generally, $f'(a) = k_1f(a)$, $f'(b) = k_2f(b)$. (In fact ℓ_1 can be any ℓ that depends on values at a only, and ℓ_2 any ℓ depending

only on b. Then $\ell_1(f) = \ell_1(g) = 0 \Rightarrow L\{\langle f(a), f'(a)\rangle, \langle g(a), g'(a)\rangle\}$ has dimansion $\leqq 1 \Rightarrow f'g - g'f\big|_a = 0$, and similarly for ℓ_2 and b, so that this split pair of endpoint conditions makes $B(f, g)]_a^b = 0$ by making the values of B at a and at b separately 0.)

4. If $c_2(a) = c_2(b)$ then take $f \in M \iff f(a) = f(b)$ and $f'(a) = f'(b)$.

<u>Lemma 6.</u> Suppose that M is defined by one of the self-adjoint boundary conditions 1, 2, or 4 above, that $c_2(t) \geqq m > 0$ on [a, b] *and that* $\lambda = \|c_0\|_\infty + 1$. Then

$$|((T - \lambda)f, f)| \geqq m\|f'\|_2^2 + \|f\|_2^2$$

for all $f \in M$.

<u>Proof.</u> We have $((\lambda - T)f, f) = -\int_a^b (c_2f')'f + \int_a^b (\lambda - c_0)f^2$ $= -c_2f'f\Big]_a^b + \int_a^b c_2(f')^2 + \int_a^b (\lambda - c_0)f^2$. Under any of the conditions 1, 2 or 4, $c_2f'f\Big]_a^b = 0$, and the two integral terms are clearly bounded below by $m\|f'\|_2^2$ and $\|f\|_2^2$ respectively.

<u>Remark.</u> By using the Schwarz inequality somewhat as in the next theorem we can show that for any $c > 0$ there exists $k > 0$ such that

$$\|f\|_\infty^2 \leqq c\|f'\|_2^2 + k\|f\|_2^2$$

for any $f \in \mathcal{C}^1([a, b])$. Taking $c = m/2$ in the above situation and

$\lambda = \|c_0\|_\infty + k + 1$ we can get the lemma also for the omitted case 3, but with m replaced by $m/2$.

Corollary. In each of the cases 1, 2 and 4 M is a complement of $N(T - \lambda)$ and therefore determines a right inverse S to $T - \lambda$.

Proof. The inequality of the lemma shows that $T - \lambda$ is injective on M. Now M is the nullspace of the mapping $f \mapsto < \ell_1(f), \ell_2(f) >$ of $\mathcal{C}^2$ onto $\mathbb{R}^2$. Since $M \cap N = \{0\}$, this mapping is injective on the two-dimensional space $N = N(T - \lambda)$ and therefore is an isomorphism on N. For any $f \in \mathcal{C}^2([a, b])$ we can thus find $g \in N$ such that $< \ell_1(g), \ell_2(g) > = < \ell_1(f), \ell_2(f) >$. Then $f - g \in M$, and $f = g + (f-g)$, showing that $\mathcal{C}^2 = N + M$, q. e. d.

We come now to our main theorem. It says that the right inverse S of $T - \lambda$ determined by the subspace M above is a compact self-adjoint mapping of the preHilbert space $\mathcal{C}^0([a, b])$ into itself, and is therefore endowed with all the rich eigen-value structures of §4 . First, some classical terminology. A Sturm-Liouville system on $[a, b]$ is a formally self-adjoint second order differential operator $Tf = (c_2 f')' + c_0 f$ defined over the closed interval $[a, b]$, together with a self-adjoint boundary condition $\ell_1(f) = \ell_2(f) = 0$ for that interval. If $c_2(t)$ is never zero on $[a, b]$ the system is called regular. If $c_2(a)$ or $c_2(b)$ is zero, or if the interval $[a, b]$ is replaced by an infinite interval such as $[a, \infty)$, then the system is called singular.

Theorem 9. If T; ℓ_1, ℓ_2 is a regular Sturm-Liouille system on $[a, b]$, with $c_2(t) > 0$ there, then the subspace

$M = \{f \in \mathcal{C}^2[a, b]\ ;\ \ell_1(f) = \ell_2(f) = 0\}$ is a complement of $N(T-\lambda)$ if λ is taken sufficiently large, and the right inverse of $T - \lambda$ thus determined by M is a compact self-adjoint mapping of the preHilbert space $\mathcal{C}^0([a, b])$ into itself.

Proof. The proof depends on the inequality of the above lemma. Since we have proved this inequality only for the boundary conditions 1, 2 and 4 our proof will be complete only for those cases.

Set $g = (T - \lambda)f$. Since $\|g\|_2 \|f\|_2 \geqq |((T - \lambda)f, f)|$ by the Schwarz inequality, we see from the lemma first that $\|f\|_2^2 \leqq \|g\|_2\|f\|_2$, so that $\|f\|_2 \leqq \|g\|_2$, and then that $m\|f'\|_2^2 \leqq \|g\|_2 \|f\|_2 \leqq \|g\|_2^2$, so that $\|f'\|_2 \leqq \|g\|_2/\sqrt{m}$.

We have already checked that the right inverse S of the formally self-adjoint $T-\lambda$ defined by M is self-adjoint, and there remains to be shown that the set $S[U] = \{f : \|g\|_2 \leqq 1\}$ has compact closure. We have, first, by Schwarz, that

$$|f(y) - f(x)| \leqq \int_x^y |f'| \leqq \|f'\|_2 \, |y-x|^{1/2} \leqq \frac{|y-x|^{1/2}}{\sqrt{m}}$$

Thus $S[U]$ is equicontinuous. Taking y and x where $|f|$ assumes its maximum and minimumvalues, we have $\|f\|_\infty - \min|f| \leqq (b-a)^{1/2}/\sqrt{m}$. Since $1 \geqq \|g\|_2 \geqq \|f\|_2 \geqq (\min|f|)(b-a)^{1/2}$, we have $\|f\|_\infty \leqq C$, where $C = 1/(b-a)^{1/2} + (b-a)^{1/2}/\sqrt{m}$.

Thus $S[U]$ is a set of equicontinuous functions mapping the compact set $[a,b]$ into the compact set $[-C, C]$ and is therefore totally-bounded in the <u>uniform norm</u> by §5 . Since $\mathcal{C}^0$ is complete in the uniform norm, every sequence in $S[U]$ has a subsequence uniformly converging to some $f \in \mathcal{C}^0$. Since $\|f\|_2 \leq (b-a)^{1/2} \|f\|_\infty$, this subsequence also converges to f in the 2-norm and this is equivalent to the desired compactness.

Corollary . There exists an orthonormal sequence $\{\phi_n\}$ consisting entirely of eigen-vectors of T and forming a basis for M. Moreover the Fourier expansion of any $f \in M$ with respect to the basis $\{\phi_n\}$ converges uniformly to f (as well as in the 2-norm).

Proof. By §4 there exists an eigenbasis for S on the range of S, which is M. Since $S\phi_n = r_n\phi_n$, we have $(T - \lambda)(r_n\phi_n) = \phi_n$ and $T\phi_n = ((1 + \lambda r_n)/r_n)\phi_n$. The uniformity of the series convergence comes out of the following general consideration.

Lemma 7 . Suppose in §4 that the self-adjoint operator T is compact from the scalar product norm p to another norm q on V, and suppose that $p \leq cq$ for some c. Then T is compact (from p to p) and the eigen-basis expansion $\sum b_n\phi_n$ of an element $\beta \in R(T)$ converges to β in both norms.

Proof. Let U be the unit sphere of V in the scalar product norm. By the hypothesis of the lemma, T[U] has compact q-closure and therefore compact p-closure . We thus have the eigen-basis theorem.

Now let $\beta = T(\alpha)$ have the Fourier series $\sum b_i\phi_i$, where $T(\phi_i) = r_i\phi_i$. Then $b_i = r_i a_i$ where $\{a_i\}$ are the Fourier coefficients of α. Since the sequence of partial sums $\sum_1^n a_i\phi_i$ is p-bounded (Bessel's inequality), the sequence $\left\{\sum_1^n b_i\phi_i\right\} = \left\{T(\sum_1^n a_i\phi_i)\right\}$ is q-totally-bounded . Any subsequence of it therefore has a subsubsequence q-converging to some element $\gamma \in V$. Since it then p-converges to γ, γ must be β. Thus every subsequence has a subsubsequence q-converging to β and so $\left\{\sum_1^n b_i\phi_i\right\}$ itself q-converges to β, q.e.d.

§7. Fourier Series

There are not many regular Sturm-Liouville systems whose associated orthonormal eigen-bases have proved to be important in actual calculations. Most orthonormal bases that are used, such as those due to Bessel, Legendre, Hermite and Laguerre, arise from singular Sturm-Liouville systems and are therefore beyond the limitations we have set for this discussion. However, the best known example of all, Fourier series, is available to us.

We shall consider the constant coefficient operator $Tf = D^2 f$, which is clearly both formally self-adjoint and regular, and either the boundary conditions $f(0) = f(\pi) = 0$ on $[0, \pi]$ (type 1) or the periodic boundary condition $f(-\pi) = f(\pi)$, $f'(-\pi) = f'(\pi)$ on $[-\pi, \pi]$ (type 4).

To solve the first problem we have to find the solutions of $f'' - \lambda f = 0$ which satisfy $f(0) = f(\pi) = 0$. If $\lambda > 0$ then we know that the two-dimensional solution space is spanned by $\{e^{rx}, e^{-rx}\}$ where $r = \lambda^{1/2}$. But if $c_1 e^{rx} + c_2 e^{-rx}$ is 0 at both 0 and π then $c_1 = c_2 = 0$ (because the pairs $<1, 1>$ and $\langle e^{r\pi}, e^{-r\pi} \rangle$ are independent). Therefore there are no solutions satisfying the boundary conditions when $\lambda > 0$. If $\lambda = 0$ then $f(x) = c_1 x + c_0$ and again $c_1 = c_0 = 0$.

If $\lambda < 0$ then the solution space is spanned by $\{\sin rx, \cos rx\}$ where $r = (-\lambda)^{1/2}$. Now if $c_1 \sin rx + c_2 \cos rx$ is 0 at $x = 0$ and $x = \pi$ we get, first, that $c_2 = 0$, and second that $r\pi = n\pi$ for some integer n. Thus the eigen-functions for the first system form the set $\{\sin nx\}_1^{\infty}$ and the corresponding eigen-values of D^2 are $\{n^2\}_1^{\infty}$.

We therefore have the following corollary of the Sturm-Liouville theorem and the 2-norm density of M in $\mathcal{C}^0$.

Theorem . The sequence $\{\sin nx\}_1^{\infty}$ is an orthogonal basis for the preHilbert space $\mathcal{C}^0([0, \pi])$. If $f \in \mathcal{C}^2([0, \pi])$ and $f(0) = f(\pi) = 0$ then the Fourier series for f converges uniformly to f.

We now consider the second boundary problem. The computations are now a little more complicated, but again, if $f(x) = c_1 e^{rx} + c_2 e^{-rx}$ and if $f(-\pi) = f(\pi)$ and $f'(-\pi) = f'(\pi)$ then $f = 0$. For now we have $c_1 e^{-r\pi} + c_2 e^{r\pi} = c_1 e^{r\pi} + c_2 e^{-r\pi}$, giving, $c_1 = c_2$, and $c_1 re^{-r\pi} - c_2 re^{r\pi} = c_1 re^{r\pi} - c_2 re^{-r\pi}$, giving, $c_1(e^{r\pi} - e^{-r\pi}) = 0$ and so $c_1 = 0$. Again $f(x) = c_1 x + c_0$ is ruled out. Finally, if $f(x) = c_1 \sin rx + c_2 \cos rx$, our boundary conditions become

$$2c_1 \sin r\pi = 0 \quad \text{and} \quad 2rc_2 \sin r\pi = 0,$$

so that again $r = n$, but this time the full solution space of $(D^2 + n^2)f = 0$ satisfies the boundary condition.

Theorem . The set $\{\sin nx\}_1^{\infty} \cup \{\cos nx\}_0^{\infty}$ forms an orthogonal basis for the preHilbert space $\mathcal{C}^0([-\pi, \pi])$. If $f \in \mathcal{C}^2([-\pi, \pi])$ and $f(-\pi) = f(\pi)$, $f'(-\pi) = f'(\pi)$ then the Fourier series for f converges to f uniformly on $[-\pi, \pi]$.

Remaining proof. This theorem follows from our general Sturm-Liouville discussion except for the orthogonality of sin nx and cos nx. We have $(\sin nx, \cos nx) = \int_{-\pi}^{\pi} \sin nt \cos nt \, dt = \frac{1}{2} \int_{-\pi}^{\pi} \sin 2nt \, dt = -\frac{1}{4n} \cos 2nx \Big]_{-\pi}^{\pi} = 0$. Or we can simply remark that the first integrand is an odd function and therefore its integral over any symmetric interval [-a, a] is necessarily zero.

The orthogonality of eigen-vectors having different eigen-values follows of course as in the proof of the theorem in §2.