Starting with S_0, we thus construct sets $S_1, \ldots, S_r$. The last of these consists of $\mathbf{y}_1, \ldots, \mathbf{y}_r$, and our construction shows that it spans X. But Q is independent; hence $\mathbf{y}_{r+1}$ is not in the span of S_r. This contradiction establishes the theorem.

Corollary $\dim R^n = n$.

Proof Since $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ spans R^n, the theorem shows that $\dim R^n \le n$. Since $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is independent, $\dim R^n \ge n$.

9.3 Theorem *Suppose X is a vector space, and* $\dim X = n$.

(*a*) *A set E of n vectors in X spans X if and only if E is independent.*
(*b*) *X has a basis, and every basis consists of n vectors.*
(*c*) *If $1 \le r \le n$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$ is an independent set in X, then X has a basis containing $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$.*

Proof Suppose $E = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. Since $\dim X = n$, the set $\{\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{y}\}$ is dependent, for every $\mathbf{y} \in X$. If E is independent, it follows that $\mathbf{y}$ is in the span of E; hence E spans X. Conversely, if E is dependent, one of its members can be removed without changing the span of E. Hence E cannot span X, by Theorem 9.2. This proves (*a*).

Since $\dim X = n$, X contains an independent set of n vectors, and (*a*) shows that every such set is a basis of X; (*b*) now follows from 9.1(*d*) and 9.2.

To prove (*c*), let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a basis of X. The set

$$S = \{\mathbf{y}_1, \ldots, \mathbf{y}_r, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$$

spans X and is dependent, since it contains more than n vectors. The argument used in the proof of Theorem 9.2 shows that one of the $\mathbf{x}_i$'s is a linear combination of the other members of S. If we remove this $\mathbf{x}_i$ from S, the remaining set still spans X. This process can be repeated r times and leads to a basis of X which contains $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$, by (*a*).

9.4 Definitions A mapping A of a vector space X into a vector space Y is said to be a *linear transformation* if

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2, \qquad A(c\mathbf{x}) = cA\mathbf{x}$$

for all $\mathbf{x}$, $\mathbf{x}_1$, $\mathbf{x}_2 \in X$ and all scalars c. Note that one often writes $A\mathbf{x}$ instead of $A(\mathbf{x})$ if A is linear.

Observe that $A\mathbf{0} = \mathbf{0}$ if A is linear. Observe also that a linear transformation A of X into Y is completely determined by its action on any basis: If

$\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is a basis of X, then every $\mathbf{x} \in X$ has a unique representation of the form

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{x}_i,$$

and the linearity of A allows us to compute $A\mathbf{x}$ from the vectors $A\mathbf{x}_1, \ldots, A\mathbf{x}_n$ and the coordinates $c_1, \ldots, c_n$ by the formula

$$A\mathbf{x} = \sum_{i=1}^{n} c_i A\mathbf{x}_i.$$

Linear transformations of X into X are often called *linear operators* on X. If A is a linear operator on X which (i) is one-to-one and (ii) maps X onto X, we say that A is *invertible*. In this case we can define an operator A^{-1} on X by requiring that $A^{-1}(A\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in X$. It is trivial to verify that we then also have $A(A^{-1}\mathbf{x}) = \mathbf{x}$, for all $\mathbf{x} \in X$, and that A^{-1} is linear.

An important fact about linear operators on finite-dimensional vector spaces is that each of the above conditions (i) and (ii) implies the other:

9.5 Theorem *A linear operator A on a finite-dimensional vector space X is one-to-one if and only if the range of A is all of X.*

Proof Let $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a basis of X. The linearity of A shows that its range $\mathscr{R}(A)$ is the span of the set $Q = \{A\mathbf{x}_1, \ldots, A\mathbf{x}_n\}$. We therefore infer from Theorem 9.3(a) that $\mathscr{R}(A) = X$ if and only if Q is independent. We have to prove that this happens if and only if A is one-to-one.

Suppose A is one-to-one and $\Sigma c_i A\mathbf{x}_i = \mathbf{0}$. Then $A(\Sigma c_i \mathbf{x}_i) = \mathbf{0}$, hence $\Sigma c_i \mathbf{x}_i = \mathbf{0}$, hence $c_1 = \cdots = c_n = 0$, and we conclude that Q is independent.

Conversely, suppose Q is independent and $A(\Sigma c_i \mathbf{x}_i) = \mathbf{0}$. Then $\Sigma c_i A\mathbf{x}_i = \mathbf{0}$, hence $c_1 = \cdots = c_n = 0$, and we conclude: $A\mathbf{x} = \mathbf{0}$ only if $\mathbf{x} = \mathbf{0}$. If now $A\mathbf{x} = A\mathbf{y}$, then $A(\mathbf{x} - \mathbf{y}) = A\mathbf{x} - A\mathbf{y} = \mathbf{0}$, so that $\mathbf{x} - \mathbf{y} = \mathbf{0}$, and this says that A is one-to-one.

9.6 Definitions

(a) Let $L(X, Y)$ be the set of all linear transformations of the vector space X into the vector space Y. Instead of $L(X, X)$, we shall simply write $L(X)$. If $A_1, A_2 \in L(X, Y)$ and if c_1, c_2 are scalars, define $c_1 A_1 + c_2 A_2$ by

$$(c_1 A_1 + c_2 A_2)\mathbf{x} = c_1 A_1 \mathbf{x} + c_2 A_2 \mathbf{x} \qquad (\mathbf{x} \in X).$$

It is then clear that $c_1 A_1 + c_2 A_2 \in L(X, Y)$.

(b) If X, Y, Z are vector spaces, and if $A \in L(X, Y)$ and $B \in L(Y, Z)$, we define their *product* BA to be the composition of A and B:

$$(BA)\mathbf{x} = B(A\mathbf{x}) \qquad (\mathbf{x} \in X).$$

Then $BA \in L(X, Z)$.

Note that BA need not be the same as AB, even if $X = Y = Z$.

(*c*) For $A \in L(R^n, R^m)$, define the *norm* $\|A\|$ of A to be the sup of all numbers $|A\mathbf{x}|$, where $\mathbf{x}$ ranges over all vectors in R^n with $|\mathbf{x}| \le 1$.

Observe that the inequality

$$|A\mathbf{x}| \le \|A\| \, |\mathbf{x}|$$

holds for all $\mathbf{x} \in R^n$. Also, if λ is such that $|A\mathbf{x}| \le \lambda |\mathbf{x}|$ for all $\mathbf{x} \in R^n$, then $\|A\| \le \lambda$.

9.7 Theorem

(*a*) *If $A \in L(R^n, R^m)$, then $\|A\| < \infty$ and A is a uniformly continuous mapping of R^n into R^m.*

(*b*) *If $A, B \in L(R^n, R^m)$ and c is a scalar, then*

$$\|A + B\| \le \|A\| + \|B\|, \qquad \|cA\| = |c| \, \|A\|.$$

With the distance between A and B defined as $\|A - B\|$, $L(R^n, R^m)$ is a metric space.

(*c*) *If $A \in L(R^n, R^m)$ and $B \in L(R^m, R^k)$, then*

$$\|BA\| \le \|B\| \, \|A\|.$$

Proof

(*a*) Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the standard basis in R^n and suppose $\mathbf{x} = \Sigma c_i \mathbf{e}_i$, $|\mathbf{x}| \le 1$, so that $|c_i| \le 1$ for $i = 1, \ldots, n$. Then

$$|A\mathbf{x}| = \left|\sum c_i A\mathbf{e}_i\right| \le \sum |c_i| \, |A\mathbf{e}_i| \le \sum |A\mathbf{e}_i|$$

so that

$$\|A\| \le \sum_{i=1}^{n} |A\mathbf{e}_i| < \infty.$$

Since $|A\mathbf{x} - A\mathbf{y}| \le \|A\| \, |\mathbf{x} - \mathbf{y}|$ if $\mathbf{x}, \mathbf{y} \in R^n$, we see that A is uniformly continuous.

(*b*) The inequality in (*b*) follows from

$$|(A + B)\mathbf{x}| = |A\mathbf{x} + B\mathbf{x}| \le |A\mathbf{x}| + |B\mathbf{x}| \le (\|A\| + \|B\|) \, |\mathbf{x}|.$$

The second part of (*b*) is proved in the same manner. If

$$A, B, C \in L(R^n, R^m),$$

we have the triangle inequality

$$\|A - C\| = \|(A - B) + (B - C)\| \le \|A - B\| + \|B - C\|,$$

and it is easily verified that $\|A - B\|$ has the other properties of a metric (Definition 2.15).

(*c*) Finally, (*c*) follows from

$$|(BA)\mathbf{x}| = |B(A\mathbf{x})| \leq \|B\| \, |A\mathbf{x}| \leq \|B\| \, \|A\| \, |\mathbf{x}|.$$

Since we now have metrics in the spaces $L(R^n, R^m)$, the concepts of open set, continuity, etc., make sense for these spaces. Our next theorem utilizes these concepts.

9.8 Theorem *Let Ω be the set of all invertible linear operators on R^n.*

(*a*) *If $A \in \Omega$, $B \in L(R^n)$, and*

$$\|B - A\| \cdot \|A^{-1}\| < 1,$$

then $B \in \Omega$.

(*b*) *Ω is an open subset of $L(R^n)$, and the mapping $A \to A^{-1}$ is continuous on Ω.*

(This mapping is also obviously a $1 - 1$ mapping of Ω onto Ω, which is its own inverse.)

Proof

(*a*) Put $\|A^{-1}\| = 1/\alpha$, put $\|B - A\| = \beta$. Then $\beta < \alpha$. For every $\mathbf{x} \in R^n$,

$$\begin{aligned} \alpha|\mathbf{x}| &= \alpha|A^{-1}A\mathbf{x}| \leq \alpha\|A^{-1}\| \cdot |A\mathbf{x}| \\ &= |A\mathbf{x}| \leq |(A - B)\mathbf{x}| + |B\mathbf{x}| \leq \beta|\mathbf{x}| + |B\mathbf{x}|, \end{aligned}$$

so that

$$(1) \qquad (\alpha - \beta)|\mathbf{x}| \leq |B\mathbf{x}| \qquad (\mathbf{x} \in R^n).$$

Since $\alpha - \beta > 0$, (1) shows that $B\mathbf{x} \neq 0$ if $\mathbf{x} \neq 0$. Hence B is $1 - 1$. By Theorem 9.5, $B \in \Omega$. This holds for all B with $\|B - A\| < \alpha$. Thus we have (*a*) and the fact that Ω is open.

(*b*) Next, replace $\mathbf{x}$ by $B^{-1}\mathbf{y}$ in (1). The resulting inequality

$$(2) \qquad (\alpha - \beta)|B^{-1}\mathbf{y}| \leq |BB^{-1}\mathbf{y}| = |\mathbf{y}| \qquad (\mathbf{y} \in R^n)$$

shows that $\|B^{-1}\| \leq (\alpha - \beta)^{-1}$. The identity

$$B^{-1} - A^{-1} = B^{-1}(A - B)A^{-1},$$

combined with Theorem 9.7(*c*), implies therefore that

$$\|B^{-1} - A^{-1}\| \leq \|B^{-1}\| \, \|A - B\| \, \|A^{-1}\| \leq \frac{\beta}{\alpha(\alpha - \beta)}.$$

This establishes the continuity assertion made in (*b*), since $\beta \to 0$ as $B \to A$.

9.9 Matrices Suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ are bases of vector spaces X and Y, respectively. Then every $A \in L(X, Y)$ determines a set of numbers a_{ij} such that

$$A\mathbf{x}_j = \sum_{i=1}^{m} a_{ij}\mathbf{y}_i \qquad (1 \le j \le n). \tag{3}$$

It is convenient to visualize these numbers in a rectangular array of m rows and n columns, called an *m by n matrix*:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Observe that the coordinates a_{ij} of the vector $A\mathbf{x}_j$ (with respect to the basis $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$) appear in the jth column of $[A]$. The vectors $A\mathbf{x}_j$ are therefore sometimes called the *column vectors* of $[A]$. With this terminology, the *range of A is spanned by the column vectors of* $[A]$.

If $\mathbf{x} = \Sigma c_j \mathbf{x}_j$, the linearity of A, combined with (3), shows that

$$A\mathbf{x} = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}c_j\right)\mathbf{y}_i. \tag{4}$$

Thus the coordinates of $A\mathbf{x}$ are $\Sigma_j a_{ij}c_j$. Note that in (3) the summation ranges over the first subscript of a_{ij}, but that we sum over the second subscript when computing coordinates.

Suppose next that an m by n matrix is given, with real entries a_{ij}. If A is then defined by (4), it is clear that $A \in L(X, Y)$ and that $[A]$ is the given matrix. Thus there is a natural 1-1 correspondence between $L(X, Y)$ and the set of all real m by n matrices. We emphasize, though, that $[A]$ depends not only on A but also on the choice of bases in X and Y. The same A may give rise to many different matrices if we change bases, and vice versa. We shall not pursue this observation any further, since we shall usually work with fixed bases. (Some remarks on this may be found in Sec. 9.37.)

If Z is a third vector space, with basis $\{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$, if A is given by (3), and if

$$B\mathbf{y}_i = \sum_k b_{ki}\mathbf{z}_k, \qquad (BA)\mathbf{x}_j = \sum_k c_{kj}\mathbf{z}_k,$$

then $A \in L(X, Y)$, $B \in L(Y, Z)$, $BA \in L(X, Z)$, and since

$$\begin{aligned} B(A\mathbf{x}_j) &= B\sum_i a_{ij}\mathbf{y}_i = \sum_i a_{ij}B\mathbf{y}_i \\ &= \sum a_{ij}\sum_k b_{ki}\mathbf{z}_k = \sum_k \left(\sum_i b_{ki}a_{ij}\right)\mathbf{z}_k, \end{aligned}$$

the independence of $\{\mathbf{z}_1, \ldots, \mathbf{z}_p\}$ implies that

$$c_{kj} = \sum_i b_{ki} a_{ij} \qquad (1 \le k \le p,\ 1 \le j \le n). \tag{5}$$

This shows how to compute the p by n matrix $[BA]$ from $[B]$ and $[A]$. If we define the product $[B][A]$ to be $[BA]$, then (5) describes the usual rule of matrix multiplication.

Finally, suppose $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ are standard bases of R^n and R^m, and A is given by (4). The Schwarz inequality shows that

$$|A\mathbf{x}|^2 = \sum_i \Big(\sum_j a_{ij} c_j\Big)^2 \le \sum_i \Big(\sum_j a_{ij}^2 \cdot \sum_j c_j^2\Big) = \sum_{i,j} a_{ij}^2 |\mathbf{x}|^2.$$

Thus

$$\|A\| \le \Big\{\sum_{i,j} a_{ij}^2\Big\}^{1/2}. \tag{6}$$

If we apply (6) to $B - A$ in place of A, where $A, B \in L(R^n, R^m)$, we see that if the matrix elements a_{ij} are continuous functions of a parameter, then the same is true of A. More precisely:

If S is a metric space, if $a_{11}, \ldots, a_{mn}$ are real continuous functions on S, and if, for each $p \in S$, A_p is the linear transformation of R^n into R^m whose matrix has entries $a_{ij}(p)$, then the mapping $p \to A_p$ is a continuous mapping of S into $L(R^n, R^m)$.

DIFFERENTIATION

9.10 Preliminaries In order to arrive at a definition of the derivative of a function whose domain is R^n (or an open subset of R^n), let us take another look at the familiar case $n = 1$, and let us see how to interpret the derivative in that case in a way which will naturally extend to $n > 1$.

If f is a real function with domain $(a, b) \subset R^1$ and if $x \in (a, b)$, then $f'(x)$ is usually defined to be the real number

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}, \tag{7}$$

provided, of course, that this limit exists. Thus

$$f(x+h) - f(x) = f'(x)h + r(h) \tag{8}$$

where the "remainder" $r(h)$ is small, in the sense that

$$\lim_{h \to 0} \frac{r(h)}{h} = 0. \tag{9}$$

Note that (8) expresses the difference $f(x+h)-f(x)$ as the sum of the *linear function* that takes h to $f'(x)h$, plus a small remainder.

We can therefore regard the derivative of f at x, not as a real number, but as the linear operator on R^1 that takes h to $f'(x)h$.

[Observe that every real number α gives rise to a linear operator on R^1; the operator in question is simply multiplication by α. Conversely, every linear function that carries R^1 to R^1 is multiplication by some real number. It is this natural 1-1 correspondence between R^1 and $L(R^1)$ which motivates the preceding statements.]

Let us next consider a function $\mathbf{f}$ that maps $(a, b) \subset R^1$ into R^m. In that case, $\mathbf{f}'(x)$ was defined to be that vector $\mathbf{y} \in R^m$ (if there is one) for which

$$\lim_{h\to 0}\left\{\frac{\mathbf{f}(x+h)-\mathbf{f}(x)}{h}-\mathbf{y}\right\}=\mathbf{0}. \tag{10}$$

We can again rewrite this in the form

$$\mathbf{f}(x+h)-\mathbf{f}(x)=h\mathbf{y}+\mathbf{r}(h), \tag{11}$$

where $\mathbf{r}(h)/h \to \mathbf{0}$ as $h \to 0$. The main term on the right side of (11) is again a *linear* function of h. Every $\mathbf{y} \in R^m$ induces a linear transformation of R^1 into R^m, by associating to each $h \in R^1$ the vector $h\mathbf{y} \in R^m$. This identification of R^m with $L(R^1, R^m)$ allows us to regard $\mathbf{f}'(x)$ as a member of $L(R^1, R^m)$.

Thus, if $\mathbf{f}$ is a differentiable mapping of $(a, b) \subset R^1$ into R^m, and if $x \in (a, b)$, then $\mathbf{f}'(x)$ is the linear transformation of R^1 into R^m that satisfies

$$\lim_{h\to 0}\frac{\mathbf{f}(x+h)-\mathbf{f}(x)-\mathbf{f}'(x)h}{h}=\mathbf{0}, \tag{12}$$

or, equivalently,

$$\lim_{h\to 0}\frac{|\mathbf{f}(x+h)-\mathbf{f}(x)-\mathbf{f}'(x)h|}{|h|}=0. \tag{13}$$

We are now ready for the case $n > 1$.

9.11 Definition Suppose E is an open set in R^n, $\mathbf{f}$ maps E into R^m, and $\mathbf{x} \in E$. If there exists a linear transformation A of R^n into R^m such that

$$\lim_{\mathbf{h}\to \mathbf{0}}\frac{|\mathbf{f}(\mathbf{x}+\mathbf{h})-\mathbf{f}(\mathbf{x})-A\mathbf{h}|}{|\mathbf{h}|}=0, \tag{14}$$

then we say that $\mathbf{f}$ is *differentiable at* $\mathbf{x}$, and we write

$$\mathbf{f}'(\mathbf{x})=A. \tag{15}$$

If $\mathbf{f}$ is differentiable at every $\mathbf{x} \in E$, we say that $\mathbf{f}$ is *differentiable in* E.