This is equivalent to (58), and completes the proof.

Note. In terms of the components of $\mathbf{f}$ and $\mathbf{g}$, (65) becomes

$$\sum_{j=1}^{n} (D_j f_i)(\mathbf{a}, \mathbf{b})(D_k g_j)(\mathbf{b}) = -(D_{n+k} f_i)(\mathbf{a}, \mathbf{b})$$

or

$$\sum_{j=1}^{n} \left(\frac{\partial f_i}{\partial x_j}\right)\left(\frac{\partial g_j}{\partial y_k}\right) = -\left(\frac{\partial f_i}{\partial y_k}\right)$$

where $1 \le i \le n$, $1 \le k \le m$.

For each k, this is a system of n linear equations in which the derivatives $\partial g_j / \partial y_k$ $(1 \le j \le n)$ are the unknowns.

9.29 Example Take $n = 2$, $m = 3$, and consider the mapping $\mathbf{f} = (f_1, f_2)$ of R^5 into R^2 given by

$$f_1(x_1, x_2, y_1, y_2, y_3) = 2e^{x_1} + x_2 y_1 - 4y_2 + 3$$
$$f_2(x_1, x_2, y_1, y_2, y_3) = x_2 \cos x_1 - 6x_1 + 2y_1 - y_3 .$$

If $\mathbf{a} = (0, 1)$ and $\mathbf{b} = (3, 2, 7)$, then $\mathbf{f}(\mathbf{a}, \mathbf{b}) = \mathbf{0}$.

With respect to the standard bases, the matrix of the transformation $A = \mathbf{f}'(\mathbf{a}, \mathbf{b})$ is

$$[A] = \begin{bmatrix} 2 & 3 & 1 & -4 & 0 \\ -6 & 1 & 2 & 0 & -1 \end{bmatrix}.$$

Hence

$$[A_x] = \begin{bmatrix} 2 & 3 \\ -6 & 1 \end{bmatrix}, \qquad [A_y] = \begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix}.$$

We see that the column vectors of $[A_x]$ are independent. Hence A_x is invertible and the implicit function theorem asserts the existence of a $\mathscr{C}'$-mapping $\mathbf{g}$, defined in a neighborhood of $(3, 2, 7)$, such that $\mathbf{g}(3, 2, 7) = (0, 1)$ and $\mathbf{f}(\mathbf{g}(\mathbf{y}), \mathbf{y}) = \mathbf{0}$.

We can use (58) to compute $\mathbf{g}'(3, 2, 7)$: Since

$$[(A_x)^{-1}] = [A_x]^{-1} = \frac{1}{20}\begin{bmatrix} 1 & -3 \\ 6 & 2 \end{bmatrix}$$

(58) gives

$$[\mathbf{g}'(3, 2, 7)] = -\frac{1}{20}\begin{bmatrix} 1 & -3 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} 1 & -4 & 0 \\ 2 & 0 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{1}{5} & -\frac{3}{20} \\ -\frac{1}{2} & \frac{6}{5} & \frac{1}{10} \end{bmatrix}.$$

In terms of partial derivatives, the conclusion is that

$$D_1 g_1 = \tfrac{1}{4} \qquad D_2 g_1 = \tfrac{1}{5} \qquad D_3 g_1 = -\tfrac{3}{20}$$
$$D_1 g_2 = -\tfrac{1}{2} \qquad D_2 g_2 = \tfrac{6}{5} \qquad D_3 g_2 = \tfrac{1}{10}$$

at the point (3, 2, 7).

THE RANK THEOREM

Although this theorem is not as important as the inverse function theorem or the implicit function theorem, we include it as another interesting illustration of the general principle that the local behavior of a continuously differentiable mapping $\mathbf{F}$ near a point $\mathbf{x}$ is similar to that of the linear transformation $\mathbf{F}'(\mathbf{x})$.

Before stating it, we need a few more facts about linear transformations.

9.30 Definitions Suppose X and Y are vector spaces, and $A \in L(X, Y)$, as in Definition 9.6. The *null space* of A, $\mathscr{N}(A)$, is the set of all $\mathbf{x} \in X$ at which $A\mathbf{x} = \mathbf{0}$. It is clear that $\mathscr{N}(A)$ is a vector space in X.

Likewise, the *range* of A, $\mathscr{R}(A)$, is a vector space in Y.

The *rank of* A is defined to be the dimension of $\mathscr{R}(A)$.

For example, the invertible elements of $L(R^n)$ are precisely those whose rank is n. This follows from Theorem 9.5.

If $A \in L(X, Y)$ and A has rank 0, then $A\mathbf{x} = \mathbf{0}$ for all $x \in A$, hence $\mathscr{N}(A) = X$. In this connection, see Exercise 25.

9.31 Projections Let X be a vector space. An operator $P \in L(X)$ is said to be a *projection in* X if $P^2 = P$.

More explicitly, the requirement is that $P(P\mathbf{x}) = P\mathbf{x}$ for every $\mathbf{x} \in X$. In other words, P fixes every vector in its range $\mathscr{R}(P)$.

Here are some elementary properties of projections:

(*a*) *If P is a projection in X, then every $\mathbf{x} \in X$ has a unique representation of the form*

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$$

where $\mathbf{x}_1 \in \mathscr{R}(P)$, $\mathbf{x}_2 \in \mathscr{N}(P)$.

To obtain the representation, put $\mathbf{x}_1 = P\mathbf{x}$, $\mathbf{x}_2 = \mathbf{x} - \mathbf{x}_1$. Then $P\mathbf{x}_2 = P\mathbf{x} - P\mathbf{x}_1 = P\mathbf{x} - P^2\mathbf{x} = \mathbf{0}$. As regards the uniqueness, apply P to the equation $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$. Since $\mathbf{x}_1 \in \mathscr{R}(P)$, $P\mathbf{x}_1 = \mathbf{x}_1$; since $P\mathbf{x}_2 = \mathbf{0}$, it follows that $\mathbf{x}_1 = P\mathbf{x}$.

(*b*) *If X is a finite-dimensional vector space and if X_1 is a vector space in X, then there is a projection P in X with $\mathscr{R}(P) = X_1$.*

If X_1 contains only $\mathbf{0}$, this is trivial: put $P\mathbf{x} = \mathbf{0}$ for all $\mathbf{x} \in X$.

Assume $\dim X_1 = k > 0$. By Theorem 9.3, X has then a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ such that $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is a basis of X_1. Define

$$P(c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n) = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$$

for arbitrary scalars $c_1, \ldots, c_n$.

Then $P\mathbf{x} = \mathbf{x}$ for every $\mathbf{x} \in X_1$, and $X_1 = \mathscr{R}(P)$.

Note that $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ is a basis of $\mathscr{N}(P)$. Note also that there are infinitely many projections in X, with range X_1, if $0 < \dim X_1 < \dim X$.

9.32 Theorem *Suppose m, n, r are nonnegative integers, $m \geq r$, $n \geq r$, $\mathbf{F}$ is a $\mathscr{C}'$-mapping of an open set $E \subset R^n$ into R^m, and $\mathbf{F}'(\mathbf{x})$ has rank r for every $\mathbf{x} \in E$.*

Fix $\mathbf{a} \in E$, put $A = \mathbf{F}'(\mathbf{a})$, let Y_1 be the range of A, and let P be a projection in R^m whose range is Y_1. Let Y_2 be the null space of P.

Then there are open sets U and V in R^n, with $\mathbf{a} \in U$, $U \subset E$, and there is a 1-1 $\mathscr{C}'$-mapping $\mathbf{H}$ of V onto U (whose inverse is also of class $\mathscr{C}'$) such that

$$\mathbf{F}(\mathbf{H}(\mathbf{x})) = A\mathbf{x} + \varphi(A\mathbf{x}) \qquad (\mathbf{x} \in V) \tag{66}$$

where φ is a $\mathscr{C}'$-mapping of the open set $A(V) \subset Y_1$ into Y_2.

After the proof we shall give a more geometric description of the information that (66) contains.

Proof If $r = 0$, Theorem 9.19 shows that $\mathbf{F}(\mathbf{x})$ is constant in a neighborhood U of $\mathbf{a}$, and (66) holds trivially, with $V = U$, $\mathbf{H}(\mathbf{x}) = \mathbf{x}$, $\varphi(\mathbf{0}) = \mathbf{F}(\mathbf{a})$.

From now on we assume $r > 0$. Since $\dim Y_1 = r$, Y_1 has a basis $\{\mathbf{y}_1, \ldots, \mathbf{y}_r\}$. Choose $\mathbf{z}_i \in R^n$ so that $A\mathbf{z}_i = \mathbf{y}_i$ $(1 \leq i \leq r)$, and define a linear mapping S of Y_1 into R^n by setting

$$S(c_1\mathbf{y}_1 + \cdots + c_r\mathbf{y}_r) = c_1\mathbf{z}_1 + \cdots + c_r\mathbf{z}_r \tag{67}$$

for all scalars $c_1, \ldots, c_r$.

Then $AS\mathbf{y}_i = A\mathbf{z}_i = \mathbf{y}_i$ for $1 \leq i \leq r$. Thus

$$AS\mathbf{y} = \mathbf{y} \qquad (\mathbf{y} \in Y_1). \tag{68}$$

Define a mapping $\mathbf{G}$ of E into R^n by setting

$$\mathbf{G}(\mathbf{x}) = \mathbf{x} + SP[\mathbf{F}(\mathbf{x}) - A\mathbf{x}] \qquad (\mathbf{x} \in E). \tag{69}$$

Since $\mathbf{F}'(\mathbf{a}) = A$, differentiation of (69) shows that $\mathbf{G}'(\mathbf{a}) = I$, the identity operator on R^n. By the inverse function theorem, there are open sets U and V in R^n, with $\mathbf{a} \in U$, such that $\mathbf{G}$ is a 1-1 mapping of U onto V whose inverse $\mathbf{H}$ is also of class $\mathscr{C}'$. Moreover, by shrinking U and V, if necessary, we can arrange it so that V is convex and $\mathbf{H}'(\mathbf{x})$ is invertible for every $\mathbf{x} \in V$.

Note that $ASPA = A$, since $PA = A$ and (68) holds. Therefore (69) gives

$$AG(\mathbf{x}) = PF(\mathbf{x}) \qquad (\mathbf{x} \in E). \tag{70}$$

In particular, (70) holds for $\mathbf{x} \in U$. If we replace $\mathbf{x}$ by $\mathbf{H}(\mathbf{x})$, we obtain

$$P\mathbf{F}(\mathbf{H}(\mathbf{x})) = A\mathbf{x} \qquad (\mathbf{x} \in V). \tag{71}$$

Define

$$\psi(\mathbf{x}) = \mathbf{F}(\mathbf{H}(\mathbf{x})) - A\mathbf{x} \qquad (\mathbf{x} \in V). \tag{72}$$

Since $PA = A$, (71) implies that $P\psi(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in V$. Thus ψ is a $\mathscr{C}'$-mapping of V into Y_2.

Since V is open, it is clear that $A(V)$ is an open subset of its range $\mathscr{R}(A) = Y_1$.

To complete the proof, i.e., to go from (72) to (66), we have to show that there is a $\mathscr{C}'$-mapping φ of $A(V)$ into Y_2 which satisfies

$$\varphi(A\mathbf{x}) = \psi(\mathbf{x}) \qquad (\mathbf{x} \in V). \tag{73}$$

As a step toward (73), we will first prove that

$$\psi(\mathbf{x}_1) = \psi(\mathbf{x}_2) \tag{74}$$

if $\mathbf{x}_1 \in V$, $\mathbf{x}_2 \in V$, $A\mathbf{x}_1 = A\mathbf{x}_2$.

Put $\Phi(\mathbf{x}) = \mathbf{F}(\mathbf{H}(\mathbf{x}))$, for $\mathbf{x} \in V$. Since $\mathbf{H}'(\mathbf{x})$ has rank n for every $\mathbf{x} \in V$, and $\mathbf{F}'(\mathbf{x})$ has rank r for every $\mathbf{x} \in U$, it follows that

$$\operatorname{rank} \Phi'(\mathbf{x}) = \operatorname{rank} \mathbf{F}'(\mathbf{H}(\mathbf{x}))\mathbf{H}'(\mathbf{x}) = r \qquad (\mathbf{x} \in V). \tag{75}$$

Fix $\mathbf{x} \in V$. Let M be the range of $\Phi'(\mathbf{x})$. Then $M \subset R^m$, $\dim M = r$. By (71),

$$P\Phi'(\mathbf{x}) = A. \tag{76}$$

Thus P maps M onto $\mathscr{R}(A) = Y_1$. Since M and Y_1 have the same dimension, it follows that P (restricted to M) is 1-1.

Suppose now that $A\mathbf{h} = \mathbf{0}$. Then $P\Phi'(\mathbf{x})\mathbf{h} = \mathbf{0}$, by (76). But $\Phi'(\mathbf{x})\mathbf{h} \in M$, and P is 1-1 on M. Hence $\Phi'(\mathbf{x})\mathbf{h} = \mathbf{0}$. A look at (72) shows now that we have proved the following:

If $\mathbf{x} \in V$ *and* $A\mathbf{h} = \mathbf{0}$, *then* $\psi'(\mathbf{x})\mathbf{h} = \mathbf{0}$.

We can now prove (74). Suppose $\mathbf{x}_1 \in V$, $\mathbf{x}_2 \in V$, $A\mathbf{x}_1 = A\mathbf{x}_2$. Put $\mathbf{h} = \mathbf{x}_2 - \mathbf{x}_1$ and define

$$\mathbf{g}(t) = \psi(\mathbf{x}_1 + t\mathbf{h}) \qquad (0 \le t \le 1). \tag{77}$$

The convexity of V shows that $\mathbf{x}_1 + t\mathbf{h} \in V$ for these t. Hence

$$\mathbf{g}'(t) = \psi'(\mathbf{x}_1 + t\mathbf{h})\mathbf{h} = \mathbf{0} \qquad (0 \le t \le 1), \tag{78}$$

so that $\mathbf{g}(1) = \mathbf{g}(0)$. But $\mathbf{g}(1) = \psi(\mathbf{x}_2)$ and $\mathbf{g}(0) = \psi(\mathbf{x}_1)$. This proves (74).

By (74), $\psi(\mathbf{x})$ depends only on $A\mathbf{x}$, for $\mathbf{x} \in V$. Hence (73) defines φ unambiguously in $A(V)$. It only remains to be proved that $\varphi \in \mathscr{C}'$.

Fix $\mathbf{y}_0 \in A(V)$, fix $\mathbf{x}_0 \in V$ so that $A\mathbf{x}_0 = \mathbf{y}_0$. Since V is open, $\mathbf{y}_0$ has a neighborhood W in Y_1 such that the vector

(79) $$\mathbf{x} = \mathbf{x}_0 + S(\mathbf{y} - \mathbf{y}_0)$$

lies in V for all $\mathbf{y} \in W$. By (68),

$$A\mathbf{x} = A\mathbf{x}_0 + \mathbf{y} - \mathbf{y}_0 = \mathbf{y}.$$

Thus (73) and (79) give

(80) $$\varphi(\mathbf{y}) = \psi(\mathbf{x}_0 - S\mathbf{y}_0 + S\mathbf{y}) \qquad (\mathbf{y} \in W).$$

This formula shows that $\varphi \in \mathscr{C}'$ in W, hence in $A(V)$, since $\mathbf{y}_0$ was chosen arbitrarily in $A(V)$.

The proof is now complete.

Here is what the theorem tells us about the geometry of the mapping $\mathbf{F}$.

If $\mathbf{y} \in \mathbf{F}(U)$ then $\mathbf{y} = \mathbf{F}(\mathbf{H}(\mathbf{x}))$ for some $\mathbf{x} \in V$, and (66) shows that $P\mathbf{y} = A\mathbf{x}$. Therefore

(81) $$\mathbf{y} = P\mathbf{y} + \varphi(P\mathbf{y}) \qquad (\mathbf{y} \in \mathbf{F}(U)).$$

This shows that $\mathbf{y}$ is determined by its projection $P\mathbf{y}$, and that P, restricted to $\mathbf{F}(U)$, is a 1-1 mapping of $\mathbf{F}(U)$ onto $A(V)$. Thus $\mathbf{F}(U)$ is an "r-dimensional surface" with precisely one point "over" each point of $A(V)$. We may also regard $\mathbf{F}(U)$ as the graph of φ.

If $\Phi(\mathbf{x}) = \mathbf{F}(\mathbf{H}(\mathbf{x}))$, as in the proof, then (66) shows that the level sets of Φ (these are the sets on which Φ attains a given value) are precisely the level sets of A in V. These are "flat" since they are intersections with V of translates of the vector space $\mathscr{N}(A)$. Note that $\dim \mathscr{N}(A) = n - r$ (Exercise 25).

The level sets of $\mathbf{F}$ in U are the images under $\mathbf{H}$ of the flat level sets of Φ in V. They are thus "$(n - r)$-dimensional surfaces" in U.

DETERMINANTS

Determinants are numbers associated to square matrices, and hence to the operators represented by such matrices. They are 0 if and only if the corresponding operator fails to be invertible. They can therefore be used to decide whether the hypotheses of some of the preceding theorems are satisfied. They will play an even more important role in Chap. 10.

9.33 Definition If $(j_1, \ldots, j_n)$ is an ordered n-tuple of integers, define

$$s(j_1, \ldots, j_n) = \prod_{p<q} \operatorname{sgn}(j_q - j_p), \tag{82}$$

where $\operatorname{sgn} x = 1$ if $x > 0$, $\operatorname{sgn} x = -1$ if $x < 0$, $\operatorname{sgn} x = 0$ if $x = 0$. Then $s(j_1, \ldots, j_n) = 1$, -1, or 0, and it changes sign if any two of the j's are interchanged.

Let $[A]$ be the matrix of a linear operator A on R^n, relative to the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, with entries $a(i,j)$ in the ith row and jth column. The determinant of $[A]$ is defined to be the number

$$\det [A] = \sum s(j_1, \ldots, j_n) a(1, j_1) a(2, j_2) \cdots a(n, j_n). \tag{83}$$

The sum in (83) extends over all ordered n-tuples of integers $(j_1, \ldots, j_n)$ with $1 \le j_r \le n$.

The column vectors $\mathbf{x}_j$ of $[A]$ are

$$\mathbf{x}_j = \sum_{i=1}^{n} a(i,j)\mathbf{e}_i \qquad (1 \le j \le n). \tag{84}$$

It will be convenient to think of $\det [A]$ as a function of the column vectors of $[A]$. If we write

$$\det (\mathbf{x}_1, \ldots, \mathbf{x}_n) = \det [A],$$

det is now a real function on the set of all ordered n-tuples of vectors in R^n.

9.34 Theorem

(*a*) *If I is the identity operator on R^n, then*

$$\det [I] = \det (\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1.$$

(*b*) *det is a linear function of each of the column vectors $\mathbf{x}_j$, if the others are held fixed.*

(*c*) *If $[A]_1$ is obtained from $[A]$ by interchanging two columns, then* $\det [A]_1 = -\det [A]$.

(*d*) *If $[A]$ has two equal columns, then* $\det [A] = 0$.

Proof If $A = I$, then $a(i, i) = 1$ and $a(i,j) = 0$ for $i \neq j$. Hence

$$\det [I] = s(1, 2, \ldots, n) = 1,$$

which proves (*a*). By (82), $s(j_1, \ldots, j_n) = 0$ if any two of the j's are equal. Each of the remaining $n!$ products in (83) contains exactly one factor from each column. This proves (*b*). Part (*c*) is an immediate consequence of the fact that $s(j_1, \ldots, j_n)$ changes sign if any two of the j's are interchanged, and (*d*) is a corollary of (*c*).

9.35 Theorem *If $[A]$ and $[B]$ are n by n matrices, then*

$$\det([B][A]) = \det[B] \det[A].$$

Proof If $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are the columns of $[A]$, define

$$\Delta_B(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \Delta_B[A] = \det([B][A]). \tag{85}$$

The columns of $[B][A]$ are the vectors $B\mathbf{x}_1, \ldots, B\mathbf{x}_n$. Thus

$$\Delta_B(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \det(B\mathbf{x}_1, \ldots, B\mathbf{x}_n). \tag{86}$$

By (86) and Theorem 9.34, Δ_B also has properties 9.34 (*b*) to (*d*). By (*b*) and (84),

$$\Delta_B[A] = \Delta_B\left(\sum_i a(i, 1)\mathbf{e}_i, \mathbf{x}_2, \ldots, \mathbf{x}_n\right) = \sum_i a(i, 1)\, \Delta_B(\mathbf{e}_i, \mathbf{x}_2, \ldots, \mathbf{x}_n).$$

Repeating this process with $\mathbf{x}_2, \ldots, \mathbf{x}_n$, we obtain

$$\Delta_B[A] = \sum a(i_1, 1)a(i_2, 2) \cdots a(i_n, n)\, \Delta_B(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}), \tag{87}$$

the sum being extended over all ordered n-tuples $(i_1, \ldots, i_n)$ with $1 \le i_r \le n$. By (*c*) and (*d*),

$$\Delta_B(\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}) = t(i_1, \ldots, i_n)\, \Delta_B(\mathbf{e}_1, \ldots, \mathbf{e}_n), \tag{88}$$

where $t = 1, 0$, or -1, and since $[B][I] = [B]$, (85) shows that

$$\Delta_B(\mathbf{e}_1, \ldots, \mathbf{e}_n) = \det[B]. \tag{89}$$

Substituting (89) and (88) into (87), we obtain

$$\det([B][A]) = \left\{\sum a(i_1, 1) \cdots a(i_n, n)t(i_1, \ldots, i_n)\right\} \det[B],$$

for all n by n matrices $[A]$ and $[B]$. Taking $B = I$, we see that the above sum in braces is $\det[A]$. This proves the theorem.

9.36 Theorem *A linear operator A on R^n is invertible if and only if* $\det[A] \ne 0$.

Proof If A is invertible, Theorem 9.35 shows that

$$\det[A] \det[A^{-1}] = \det[AA^{-1}] = \det[I] = 1,$$

so that $\det[A] \ne 0$.

If A is not invertible, the columns $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of $[A]$ are dependent (Theorem 9.5); hence there is one, say, $\mathbf{x}_k$, such that

$$\mathbf{x}_k + \sum_{j \ne k} c_j \mathbf{x}_j = 0 \tag{90}$$

for certain scalars c_j. By 9.34 (*b*) and (*d*), $\mathbf{x}_k$ can be replaced by $\mathbf{x}_k + c_j\mathbf{x}_j$ without altering the determinant, if $j \ne k$. Repeating, we see that $\mathbf{x}_k$ can