It is of course understood in (14) that $\mathbf{h} \in R^n$. If $|\mathbf{h}|$ is small enough, then $\mathbf{x} + \mathbf{h} \in E$, since E is open. Thus $\mathbf{f}(\mathbf{x} + \mathbf{h})$ is defined, $\mathbf{f}(\mathbf{x} + \mathbf{h}) \in R^m$, and since $A \in L(R^n, R^m)$, $A\mathbf{h} \in R^m$. Thus

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h} \in R^m.$$

The norm in the numerator of (14) is that of R^m. In the denominator we have the R^n-norm of $\mathbf{h}$.

There is an obvious uniqueness problem which has to be settled before we go any further.

9.12 Theorem *Suppose E and $\mathbf{f}$ are as in Definition* 9.11, $\mathbf{x} \in E$, *and* (14) *holds with $A = A_1$ and with $A = A_2$. Then $A_1 = A_2$.*

Proof If $B = A_1 - A_2$, the inequality

$$|B\mathbf{h}| \le |\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - A_1\mathbf{h}| + |\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) - A_2\mathbf{h}|$$

shows that $|B\mathbf{h}|/|\mathbf{h}| \to 0$ as $\mathbf{h} \to \mathbf{0}$. For fixed $\mathbf{h} \neq \mathbf{0}$, it follows that

$$\frac{|B(t\mathbf{h})|}{|t\mathbf{h}|} \to 0 \qquad \text{as} \quad t \to 0. \tag{16}$$

The linearity of B shows that the left side of (16) is independent of t. Thus $B\mathbf{h} = 0$ for every $\mathbf{h} \in R^n$. Hence $B = 0$.

9.13 Remarks

(*a*) The relation (14) can be rewritten in the form

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{h} + \mathbf{r}(\mathbf{h}) \tag{17}$$

where the remainder $\mathbf{r}(\mathbf{h})$ satisfies

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{|\mathbf{r}(\mathbf{h})|}{|\mathbf{h}|} = 0. \tag{18}$$

We may interpret (17), as in Sec. 9.10, by saying that for fixed $\mathbf{x}$ and small $\mathbf{h}$, the left side of (17) is approximately equal to $\mathbf{f}'(\mathbf{x})\mathbf{h}$, that is, to the value of a linear transformation applied to $\mathbf{h}$.

(*b*) Suppose $\mathbf{f}$ and E are as in Definition 9.11, and $\mathbf{f}$ is differentiable in E. For every $\mathbf{x} \in E$, $\mathbf{f}'(\mathbf{x})$ is then a function, namely, a linear transformation of R^n into R^m. But $\mathbf{f}'$ is also a function: $\mathbf{f}'$ maps E into $L(R^n, R^m)$.

(*c*) A glance at (17) shows that $\mathbf{f}$ is continuous at any point at which $\mathbf{f}$ is differentiable.

(*d*) The derivative defined by (14) or (17) is often called the *differential* of $\mathbf{f}$ at $\mathbf{x}$, or the *total derivative* of $\mathbf{f}$ at $\mathbf{x}$, to distinguish it from the partial derivatives that will occur later.

9.14 Example We have defined derivatives of functions carrying R^n to R^m to be linear transformations of R^n into R^m. What is the derivative of such a linear transformation? The answer is very simple.

If $A \in L(R^n, R^m)$ and if $\mathbf{x} \in R^n$, then

$$A'(\mathbf{x}) = A. \tag{19}$$

Note that $\mathbf{x}$ appears on the left side of (19), but not on the right. Both sides of (19) are members of $L(R^n, R^m)$, whereas $A\mathbf{x} \in R^m$.

The proof of (19) is a triviality, since

$$A(\mathbf{x} + \mathbf{h}) - A\mathbf{x} = A\mathbf{h}, \tag{20}$$

by the linearity of A. With $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$, the numerator in (14) is thus 0 for every $\mathbf{h} \in R^n$. In (17), $\mathbf{r}(\mathbf{h}) = \mathbf{0}$.

We now extend the chain rule (Theorem 5.5) to the present situation.

9.15 Theorem *Suppose E is an open set in R^n, $\mathbf{f}$ maps E into R^m, $\mathbf{f}$ is differentiable at $\mathbf{x}_0 \in E$, $\mathbf{g}$ maps an open set containing $\mathbf{f}(E)$ into R^k, and $\mathbf{g}$ is differentiable at $\mathbf{f}(\mathbf{x}_0)$. Then the mapping $\mathbf{F}$ of E into R^k defined by*

$$\mathbf{F}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$$

is differentiable at $\mathbf{x}_0$, and

$$\mathbf{F}'(\mathbf{x}_0) = \mathbf{g}'(\mathbf{f}(\mathbf{x}_0))\mathbf{f}'(\mathbf{x}_0). \tag{21}$$

On the right side of (21), we have the product of two linear transformations, as defined in Sec. 9.6.

Proof Put $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$, $A = \mathbf{f}'(\mathbf{x}_0)$, $B = \mathbf{g}'(\mathbf{y}_0)$, and define

$$\mathbf{u}(\mathbf{h}) = \mathbf{f}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{f}(\mathbf{x}_0) - A\mathbf{h},$$
$$\mathbf{v}(\mathbf{k}) = \mathbf{g}(\mathbf{y}_0 + \mathbf{k}) - \mathbf{g}(\mathbf{y}_0) - B\mathbf{k},$$

for all $\mathbf{h} \in R^n$ and $\mathbf{k} \in R^m$ for which $\mathbf{f}(\mathbf{x}_0 + \mathbf{h})$ and $\mathbf{g}(\mathbf{y}_0 + \mathbf{k})$ are defined. Then

$$|\mathbf{u}(\mathbf{h})| = \varepsilon(\mathbf{h})|\mathbf{h}|, \qquad |\mathbf{v}(\mathbf{k})| = \eta(\mathbf{k})|\mathbf{k}|, \tag{22}$$

where $\varepsilon(\mathbf{h}) \to 0$ as $\mathbf{h} \to \mathbf{0}$ and $\eta(\mathbf{k}) \to 0$ as $\mathbf{k} \to \mathbf{0}$.

Given $\mathbf{h}$, put $\mathbf{k} = \mathbf{f}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{f}(\mathbf{x}_0)$. Then

$$|\mathbf{k}| = |A\mathbf{h} + \mathbf{u}(\mathbf{h})| \leq [\|A\| + \varepsilon(\mathbf{h})]\,|\mathbf{h}|, \tag{23}$$

and

$$\begin{aligned}\mathbf{F}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{F}(\mathbf{x}_0) - BA\mathbf{h} &= \mathbf{g}(\mathbf{y}_0 + \mathbf{k}) - \mathbf{g}(\mathbf{y}_0) - BA\mathbf{h} \\ &= B(\mathbf{k} - A\mathbf{h}) + \mathbf{v}(\mathbf{k}) \\ &= B\mathbf{u}(\mathbf{h}) + \mathbf{v}(\mathbf{k}).\end{aligned}$$

Hence (22) and (23) imply, for $\mathbf{h} \neq \mathbf{0}$, that

$$\frac{|\mathbf{F}(\mathbf{x}_0 + \mathbf{h}) - \mathbf{F}(\mathbf{x}_0) - BA\mathbf{h}|}{|\mathbf{h}|} \leq \|B\|\, \varepsilon(\mathbf{h}) + [\|A\| + \varepsilon(\mathbf{h})]\eta(\mathbf{k}).$$

Let $\mathbf{h} \to \mathbf{0}$. Then $\varepsilon(\mathbf{h}) \to 0$. Also, $\mathbf{k} \to \mathbf{0}$, by (23), so that $\eta(\mathbf{k}) \to 0$. It follows that $\mathbf{F}'(\mathbf{x}_0) = BA$, which is what (21) asserts.

9.16 Partial derivatives We again consider a function $\mathbf{f}$ that maps an open set $E \subset R^n$ into R^m. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be the standard bases of R^n and R^m. The *components of* $\mathbf{f}$ are the real functions $f_1, \ldots, f_m$ defined by

$$\mathbf{f}(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{x})\mathbf{u}_i \qquad (\mathbf{x} \in E), \tag{24}$$

or, equivalently, by $f_i(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{u}_i$, $1 \leq i \leq m$.

For $\mathbf{x} \in E$, $1 \leq i \leq m$, $1 \leq j \leq n$, we define

$$(D_j f_i)(\mathbf{x}) = \lim_{t \to 0} \frac{f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x})}{t}, \tag{25}$$

provided the limit exists. Writing $f_i(x_1, \ldots, x_n)$ in place of $f_i(\mathbf{x})$, we see that $D_j f_i$ is the derivative of f_i with respect to x_j, keeping the other variables fixed. The notation

$$\frac{\partial f_i}{\partial x_j} \tag{26}$$

is therefore often used in place of $D_j f_i$, and $D_j f_i$ is called a *partial derivative.*

In many cases where the existence of a derivative is sufficient when dealing with functions of one variable, continuity or at least boundedness of the partial derivatives is needed for functions of several variables. For example, the functions f and g described in Exercise 7, Chap. 4, are not continuous, although their partial derivatives exist at every point of R^2. Even for continuous functions. the existence of all partial derivatives does not imply differentiability in the sense of Definition 9.11; see Exercises 6 and 14, and Theorem 9.21.

However, if $\mathbf{f}$ is known to be differentiable at a point $\mathbf{x}$, then its partial derivatives exist at $\mathbf{x}$, and they determine the linear transformation $\mathbf{f}'(\mathbf{x})$ completely:

9.17 Theorem *Suppose* $\mathbf{f}$ *maps an open set* $E \subset R^n$ *into* R^m, *and* $\mathbf{f}$ *is differentiable at a point* $\mathbf{x} \in E$. *Then the partial derivatives* $(D_j f_i)(\mathbf{x})$ *exist, and*

$$\mathbf{f}'(\mathbf{x})\mathbf{e}_j = \sum_{i=1}^{m} (D_j f_i)(\mathbf{x})\mathbf{u}_i \qquad (1 \leq j \leq n). \tag{27}$$

Here, as in Sec. 9.16, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ are the standard bases of R^n and R^m.

Proof Fix j. Since $\mathbf{f}$ is differentiable at $\mathbf{x}$,

$$\mathbf{f}(\mathbf{x} + t\mathbf{e}_j) - \mathbf{f}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})(t\mathbf{e}_j) + \mathbf{r}(t\mathbf{e}_j)$$

where $|\mathbf{r}(t\mathbf{e}_j)|/t \to 0$ as $t \to 0$. The linearity of $\mathbf{f}'(\mathbf{x})$ shows therefore that

$$\lim_{t\to 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{e}_j) - \mathbf{f}(\mathbf{x})}{t} = \mathbf{f}'(\mathbf{x})\mathbf{e}_j. \tag{28}$$

If we now represent $\mathbf{f}$ in terms of its components, as in (24), then (28) becomes

$$\lim_{t\to 0} \sum_{i=1}^{m} \frac{f_i(\mathbf{x} + t\mathbf{e}_j) - f_i(\mathbf{x})}{t} \mathbf{u}_i = \mathbf{f}'(\mathbf{x})\mathbf{e}_j. \tag{29}$$

It follows that each quotient in this sum has a limit, as $t \to 0$ (see Theorem 4.10), so that each $(D_j f_i)(\mathbf{x})$ exists, and then (27) follows from (29).

Here are some consequences of Theorem 9.17:

Let $[\mathbf{f}'(\mathbf{x})]$ be the matrix that represents $\mathbf{f}'(\mathbf{x})$ with respect to our standard bases, as in Sec. 9.9.

Then $\mathbf{f}'(\mathbf{x})\mathbf{e}_j$ is the jth column vector of $[\mathbf{f}'(\mathbf{x})]$, and (27) shows therefore that the number $(D_j f_i)(\mathbf{x})$ occupies the spot in the ith row and jth column of $[\mathbf{f}'(\mathbf{x})]$. Thus

$$[\mathbf{f}'(\mathbf{x})] = \begin{bmatrix} (D_1 f_1)(\mathbf{x}) & \cdots & (D_n f_1)(\mathbf{x}) \\ \cdots & \cdots & \cdots \\ (D_1 f_m)(\mathbf{x}) & \cdots & (D_n f_m)(\mathbf{x}) \end{bmatrix}.$$

If $\mathbf{h} = \Sigma h_j \mathbf{e}_j$ is any vector in R^n, then (27) implies that

$$\mathbf{f}'(\mathbf{x})\mathbf{h} = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n} (D_j f_i)(\mathbf{x}) h_j \right\} \mathbf{u}_i. \tag{30}$$

9.18 Example Let γ be a differentiable mapping of the segment $(a, b) \subset R^1$ into an open set $E \subset R^n$, in other words, γ is a differentiable curve in E. Let f be a real-valued differentiable function with domain E. Thus f is a differentiable mapping of E into R^1. Define

$$g(t) = f(\gamma(t)) \quad (a < t < b). \tag{31}$$

The chain rule asserts then that

$$g'(t) = f'(\gamma(t))\gamma'(t) \qquad (a < t < b). \tag{32}$$

Since $\gamma'(t) \in L(R^1, R^n)$ and $f'(\gamma(t)) \in L(R^n, R^1)$, (32) defines $g'(t)$ as a linear operator on R^1. This agrees with the fact that g maps (a, b) into R^1. However, $g'(t)$ can also be regarded as a real number. (This was discussed in Sec. 9.10.) This number can be computed in terms of the partial derivatives of f and the derivatives of the components of γ, as we shall now see.

With respect to the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of R^n, $[\gamma'(t)]$ is the n by 1 matrix (a "column matrix") which has $\gamma_i'(t)$ in the ith row, where $\gamma_1, \ldots, \gamma_n$ are the components of γ. For every $\mathbf{x} \in E$, $[f'(\mathbf{x})]$ is the 1 by n matrix (a "row matrix") which has $(D_j f)(\mathbf{x})$ in the jth column. Hence $[g'(t)]$ is the 1 by 1 matrix whose only entry is the real number

$$(33) \qquad g'(t) = \sum_{i=1}^{n} (D_i f)(\gamma(t))\gamma_i'(t).$$

This is a frequently encountered special case of the chain rule. It can be rephrased in the following manner.

Associate with each $\mathbf{x} \in E$ a vector, the so-called "gradient" of f at $\mathbf{x}$, defined by

$$(34) \qquad (\nabla f)(\mathbf{x}) = \sum_{i=1}^{n} (D_i f)(\mathbf{x})\mathbf{e}_i.$$

Since

$$(35) \qquad \gamma'(t) = \sum_{i=1}^{n} \gamma_i'(t)\mathbf{e}_i,$$

(33) can be written in the form

$$(36) \qquad g'(t) = (\nabla f)(\gamma(t)) \cdot \gamma'(t),$$

the scalar product of the vectors $(\nabla f)(\gamma(t))$ and $\gamma'(t)$.

Let us now fix an $\mathbf{x} \in E$, let $\mathbf{u} \in R^n$ be a unit vector (that is, $|\mathbf{u}| = 1$), and specialize γ so that

$$(37) \qquad \gamma(t) = \mathbf{x} + t\mathbf{u} \qquad (-\infty < t < \infty).$$

Then $\gamma'(t) = \mathbf{u}$ for every t. Hence (36) shows that

$$(38) \qquad g'(0) = (\nabla f)(\mathbf{x}) \cdot \mathbf{u}.$$

On the other hand, (37) shows that

$$g(t) - g(0) = f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x}).$$

Hence (38) gives

$$(39) \qquad \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t} = (\nabla f)(\mathbf{x}) \cdot \mathbf{u}.$$

The limit in (39) is usually called the *directional derivative* of f at $\mathbf{x}$, in the direction of the unit vector $\mathbf{u}$, and may be denoted by $(D_{\mathbf{u}}f)(\mathbf{x})$.

If f and $\mathbf{x}$ are fixed, but $\mathbf{u}$ varies, then (39) shows that $(D_{\mathbf{u}}f)(\mathbf{x})$ attains its maximum when $\mathbf{u}$ is a positive scalar multiple of $(\nabla f)(\mathbf{x})$. [The case $(\nabla f)(\mathbf{x}) = \mathbf{0}$ should be excluded here.]

If $\mathbf{u} = \Sigma u_i \mathbf{e}_i$, then (39) shows that $(D_{\mathbf{u}}f)(\mathbf{x})$ can be expressed in terms of the partial derivatives of f at $\mathbf{x}$ by the formula

$$(D_{\mathbf{u}}f)(\mathbf{x}) = \sum_{i=1}^{n} (D_i f)(\mathbf{x}) u_i. \tag{40}$$

Some of these ideas will play a role in the following theorem.

9.19 Theorem *Suppose* $\mathbf{f}$ *maps a convex open set* $E \subset R^n$ *into* R^m, $\mathbf{f}$ *is differentiable in* E, *and there is a real number* M *such that*

$$\|\mathbf{f}'(\mathbf{x})\| \le M$$

for every $\mathbf{x} \in E$. *Then*

$$|\mathbf{f}(\mathbf{b}) - \mathbf{f}(\mathbf{a})| \le M|\mathbf{b} - \mathbf{a}|$$

for all $\mathbf{a} \in E$, $\mathbf{b} \in E$.

Proof Fix $\mathbf{a} \in E$, $\mathbf{b} \in E$. Define

$$\gamma(t) = (1 - t)\mathbf{a} + t\mathbf{b}$$

for all $t \in R^1$ such that $\gamma(t) \in E$. Since E is convex, $\gamma(t) \in E$ if $0 \le t \le 1$. Put

$$\mathbf{g}(t) = \mathbf{f}(\gamma(t)).$$

Then

$$\mathbf{g}'(t) = \mathbf{f}'(\gamma(t))\gamma'(t) = \mathbf{f}'(\gamma(t))(\mathbf{b} - \mathbf{a}),$$

so that

$$|\mathbf{g}'(t)| \le \|\mathbf{f}'(\gamma(t))\|\,|\mathbf{b} - \mathbf{a}| \le M|\mathbf{b} - \mathbf{a}|$$

for all $t \in [0, 1]$. By Theorem 5.19,

$$|\mathbf{g}(1) - \mathbf{g}(0)| \le M|\mathbf{b} - \mathbf{a}|.$$

But $\mathbf{g}(0) = \mathbf{f}(\mathbf{a})$ and $\mathbf{g}(1) = \mathbf{f}(\mathbf{b})$. This completes the proof.

Corollary *If, in addition,* $\mathbf{f}'(\mathbf{x}) = \mathbf{0}$ *for all* $\mathbf{x} \in E$, *then* $\mathbf{f}$ *is constant.*

Proof To prove this, note that the hypotheses of the theorem hold now with $M = 0$.

9.20 Definition A differentiable mapping $\mathbf{f}$ of an open set $E \subset R^n$ into R^m is said to be *continuously differentiable* in E if $\mathbf{f}'$ is a continuous mapping of E into $L(R^n, R^m)$.

More explicitly, it is required that to every $\mathbf{x} \in E$ and to every $\varepsilon > 0$ corresponds a $\delta > 0$ such that

$$\|\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\| < \varepsilon$$

if $\mathbf{y} \in E$ and $|\mathbf{x} - \mathbf{y}| < \delta$.

If this is so, we also say that $\mathbf{f}$ is a $\mathscr{C}'$-mapping, or that $\mathbf{f} \in \mathscr{C}'(E)$.

9.21 Theorem *Suppose* $\mathbf{f}$ *maps an open set* $E \subset R^n$ *into* R^m. *Then* $\mathbf{f} \in \mathscr{C}'(E)$ *if and only if the partial derivatives* $D_j f_i$ *exist and are continuous on* E *for* $1 \le i \le m$, $1 \le j \le n$.

Proof Assume first that $\mathbf{f} \in \mathscr{C}'(E)$. By (27),

$$(D_j f_i)(\mathbf{x}) = (\mathbf{f}'(\mathbf{x})\mathbf{e}_j) \cdot \mathbf{u}_i$$

for all i, j, and for all $\mathbf{x} \in E$. Hence

$$(D_j f_i)(\mathbf{y}) - (D_j f_i)(\mathbf{x}) = \{[\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})]\mathbf{e}_j\} \cdot \mathbf{u}_i$$

and since $|\mathbf{u}_i| = |\mathbf{e}_j| = 1$, it follows that

$$\begin{aligned}|(D_j f_i)(\mathbf{y}) - (D_j f_i)(\mathbf{x})| &\le |[\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})]\mathbf{e}_j| \\ &\le \|\mathbf{f}'(\mathbf{y}) - \mathbf{f}'(\mathbf{x})\|.\end{aligned}$$

Hence $D_j f_i$ is continuous.

For the converse, it suffices to consider the case $m = 1$. (Why?) Fix $\mathbf{x} \in E$ and $\varepsilon > 0$. Since E is open, there is an open ball $S \subset E$, with center at $\mathbf{x}$ and radius r, and the continuity of the functions $D_j f$ shows that r can be chosen so that

$$(41) \qquad |(D_j f)(\mathbf{y}) - (D_j f)(\mathbf{x})| < \frac{\varepsilon}{n} \qquad (\mathbf{y} \in S,\ 1 \le j \le n).$$

Suppose $\mathbf{h} = \Sigma h_j \mathbf{e}_j$, $|\mathbf{h}| < r$, put $\mathbf{v}_0 = \mathbf{0}$, and $\mathbf{v}_k = h_1\mathbf{e}_1 + \cdots + h_k\mathbf{e}_k$, for $1 \le k \le n$. Then

$$(42) \qquad f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \sum_{j=1}^{n} [f(\mathbf{x} + \mathbf{v}_j) - f(\mathbf{x} + \mathbf{v}_{j-1})].$$

Since $|\mathbf{v}_k| < r$ for $1 \le k \le n$ and since S is convex, the segments with end points $\mathbf{x} + \mathbf{v}_{j-1}$ and $\mathbf{x} + \mathbf{v}_j$ lie in S. Since $\mathbf{v}_j = \mathbf{v}_{j-1} + h_j\mathbf{e}_j$, the mean value theorem (5.10) shows that the jth summand in (42) is equal to

$$h_j(D_j f)(\mathbf{x} + \mathbf{v}_{j-1} + \theta_j h_j \mathbf{e}_j)$$