## Chapter Seven

## Continuity, Derivatives, and All That

### 7.1 Limits and Continuity

Let $\boldsymbol{x}_{\boldsymbol{0}} \in \boldsymbol{R}^{n}$ and $r>0$. The set $B(\boldsymbol{a} ; r)=\left\{\boldsymbol{x} \in \boldsymbol{R}^{n}:|\boldsymbol{x}-\boldsymbol{a}|<r\right\}$ is called the open ball of radius $r$ centered at $x_{0}$. The closed ball of radius $\boldsymbol{r}$ centered at $\boldsymbol{x}_{0}$ is the set $\bar{B}(\boldsymbol{a} ; r)=\left\{\boldsymbol{x} \in \boldsymbol{R}^{n}:|\boldsymbol{x}-\boldsymbol{a}| \leq r\right\}$. Now suppose $\boldsymbol{D} \subset \boldsymbol{R}^{\boldsymbol{n}}$. A point $\boldsymbol{a} \in \boldsymbol{D}$ is called an interior point of $\boldsymbol{D}$ if there is an open ball $B(\boldsymbol{a} ; r) \subset \boldsymbol{D}$. The collection of all interior points of $\boldsymbol{D}$ is called the interior of D , and is usually denoted int $\boldsymbol{D}$. A set $\boldsymbol{U}$ is said to be open if $\boldsymbol{U}=\operatorname{int} \boldsymbol{U}$.

Suppose $f: \boldsymbol{D} \rightarrow \boldsymbol{R}^{\boldsymbol{p}}$, where $\boldsymbol{D} \subset \boldsymbol{R}^{\boldsymbol{n}}$ and suppose $\boldsymbol{a} \in \boldsymbol{R}^{\boldsymbol{n}}$ is such that every open ball centered at $\boldsymbol{a}$ meets the domain $\boldsymbol{D}$. If $\boldsymbol{y} \in \boldsymbol{R}^{\boldsymbol{p}}$ is such that for every $\varepsilon>0$, there is a $\delta>0$ so that $|f(\boldsymbol{x})-\boldsymbol{y}|<\varepsilon$ whenever $0<|\boldsymbol{x}-\boldsymbol{a}|<\delta$, then we say that $\boldsymbol{y}$ is the limit of $f$ at $\boldsymbol{a}$. This is written

$$
\lim _{x \rightarrow a} f(\boldsymbol{x})=\boldsymbol{y},
$$

and $\boldsymbol{y}$ is called the limit offat a.
Notice that this agrees with our previous definitions in case $\boldsymbol{n}=1$ and $\boldsymbol{p}=1,2$, or 3 . The usual properties of limits are relatively easy to establish:

$$
\begin{gathered}
\lim _{x \rightarrow a}(f(\boldsymbol{x})+g(\boldsymbol{x}))=\lim _{x \rightarrow a} f(\boldsymbol{x})+\lim _{x \rightarrow a} g(\boldsymbol{x}), \text { and } \\
\lim _{x \rightarrow a} a f(\boldsymbol{x})=a \lim _{x \rightarrow a} f(\boldsymbol{x}) .
\end{gathered}
$$

Now we are ready to say what we mean by a continuous function $f: \boldsymbol{D} \rightarrow \boldsymbol{R}^{p}$, where $\boldsymbol{D} \subset \boldsymbol{R}^{\boldsymbol{n}}$. Again this definition will not contradict our previous lower dimensional
definitions. Specifically, we say that $f$ is continuous at $\boldsymbol{a} \in \boldsymbol{D}$ if $\lim _{\boldsymbol{x} \rightarrow \boldsymbol{a}} f(\boldsymbol{x})=f(\boldsymbol{a})$. If $f$ is continuous at each point of its domain $\boldsymbol{D}$, we say simply that $f$ is continuous.

## Example

Every linear function is continuous. To see this, suppose $f: \boldsymbol{R}^{\boldsymbol{n}} \rightarrow \boldsymbol{R}^{p}$ is linear and $\boldsymbol{a} \in \boldsymbol{R}^{n}$. Let $\varepsilon>0$. Now let $M=\max \left\{\left|f\left(\boldsymbol{e}_{1}\right)\right|,\left|f\left(\boldsymbol{e}_{2}\right)\right|, \ldots,\left|f\left(\boldsymbol{e}_{n}\right)\right|\right\}$ and let $\delta=\frac{\boldsymbol{\varepsilon}}{n M}$. Then for $\boldsymbol{x}$ such that $0<|\boldsymbol{x}-\boldsymbol{a}|<\delta$, we have

$$
\begin{aligned}
|f(\boldsymbol{x})-f(\boldsymbol{a})| & =\left|f\left(x_{1} \boldsymbol{e}_{1}+x_{2} \boldsymbol{e}_{2}+\ldots+x_{n} \boldsymbol{e}_{n}\right)-f\left(a_{1} \boldsymbol{e}_{1}+a_{2} \boldsymbol{e}_{2}+\ldots+a_{n} \boldsymbol{e}_{n}\right)\right| \\
& =\left|\left(x_{1}-a_{1}\right) f\left(\boldsymbol{e}_{1}\right)+\left(x_{2}-a_{2}\right) f\left(\boldsymbol{e}_{2}\right)+\ldots+\left(x_{n}-a_{n}\right) f\left(\boldsymbol{e}_{n}\right)\right| \\
& \leq\left|x_{1}-a_{1}\right|\left|f\left(\boldsymbol{e}_{1}\right)\right|+\left|x_{2}-a_{2}\right|\left|f\left(\boldsymbol{e}_{2}\right)\right|+\ldots+\left|x_{n}-a_{n}\right|\left|f\left(\boldsymbol{e}_{n}\right)\right| \\
& \leq\left(\left|x_{1}-a_{1}\right|+\left|x_{2}-a_{2}\right|+\ldots+\left|x_{n}-a_{n}\right|\right) M \\
& \leq n|\boldsymbol{x}-\boldsymbol{a}| M \\
& <\varepsilon
\end{aligned}
$$

Thus $\lim _{x \rightarrow a} f(\boldsymbol{x})=f(\boldsymbol{a})$ and so $f$ is continuous.

## Another Example

Let $f: \boldsymbol{R}^{2} \rightarrow \boldsymbol{R}$ be defined by $f(\boldsymbol{x})=f\left(x_{1}, x_{2}\right)=\left\{\begin{array}{cc}\frac{x_{1} x_{2}}{x_{1}^{2}+x_{2}^{2}}, & \text { for } x_{1}^{2}+x_{2}^{2} \neq 0 \\ 0, & \text { otherwise }\end{array}\right.$.
Let's see about $\lim _{x \rightarrow(0,0)} f(x)$. Let $\boldsymbol{x}=\alpha(1,1)$. Then for all $\alpha \neq 0$, we have

$$
f(x)=f(\alpha, \alpha)=\frac{\alpha^{2}}{\alpha^{2}+\alpha^{2}}=\frac{1}{2} .
$$

Now. let $\boldsymbol{x}=\alpha(1,0)=(\alpha, 0)$. It follows that all $\alpha \neq 0, f(\boldsymbol{x})=0$. What does this tell us? It tells us that for any $\delta>0$, there are vectors $\boldsymbol{x}$ with $0<|\boldsymbol{x}-(0,0)|<\delta$ such that $f(\boldsymbol{x})=\frac{1}{2}$ and such that $f(\boldsymbol{x})=0$. This, of course, means that $\lim _{x \rightarrow(0,0)} f(\boldsymbol{x})$ does not exist.

### 7.2 Derivatives

Let $f: \boldsymbol{D} \rightarrow \boldsymbol{R}^{p}$, where $\boldsymbol{D} \subset \boldsymbol{R}^{n}$, and let $\boldsymbol{x}_{0} \in \operatorname{int} \boldsymbol{D}$. Then $f$ is differentiable at $\boldsymbol{x}_{\boldsymbol{0}}$ if there is a linear function $\boldsymbol{L}$ such that

$$
\lim _{h \rightarrow 0} \frac{1}{|\boldsymbol{h}|}\left[f\left(\boldsymbol{x}_{0}+\boldsymbol{h}\right)-f\left(\boldsymbol{x}_{0}\right)-\boldsymbol{L}(\boldsymbol{h})\right]=\boldsymbol{0} .
$$

The linear function $\boldsymbol{L}$ is called the derivative of $\boldsymbol{f}$ at $\boldsymbol{x}_{\boldsymbol{0}}$. It is usual to identify the linear function $\boldsymbol{L}$ with its matrix representation and think of the derivative at a $p \times n$ matrix. Note that in case $n=p=1$, the matrix $\boldsymbol{L}$ is simply the $1 \times 1$ matrix whose sole entry is the every day grammar school derivative of $f$.

Now, how do find the derivative of $f$ ? Suppose $f$ has a derivative at $\boldsymbol{x}_{\boldsymbol{0}}$. First, let $\boldsymbol{h}=\boldsymbol{t}_{j}=(0,0, \ldots, 0, t, 0, \ldots, 0)$. Then

$$
f(\boldsymbol{x}+\boldsymbol{h})=f\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)=\left[\begin{array}{c}
f_{1}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right) \\
f_{2}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right) \\
\vdots \\
f_{p}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)
\end{array}\right],
$$

and

$$
\boldsymbol{L} \boldsymbol{h}=\left[\begin{array}{cccc}
m_{11} & m_{12} & \cdots & m_{1 n} \\
m_{21} & m_{22} & \cdots & m_{2 n} \\
\vdots & & & \\
m_{p 1} & m_{p 2} & \cdots & m_{p n}
\end{array}\right]\left[\begin{array}{c}
0 \\
0 \\
\vdots \\
t \\
\vdots \\
0
\end{array}\right]=\left[\begin{array}{c}
m_{1 j} t \\
m_{2 j} t \\
\vdots \\
m_{p j} t
\end{array}\right],
$$

where $\boldsymbol{x}_{\boldsymbol{0}}=\left(x_{1}, x_{2}, \ldots, x_{n}\right)$, etc.
Now then,

$$
\begin{aligned}
& \frac{1}{|\mathbf{h}|}\left[f\left(\mathbf{x}_{\mathbf{0}}+\mathbf{h}\right)-f\left(\mathbf{x}_{\mathbf{0}}\right)-\mathbf{L}(\mathbf{h})\right] \\
&=\left.\sum_{t} \left\lvert\, \begin{array}{c}
\left.f_{1}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)-f_{1}\left(x_{1}, x_{2}, \ldots, x_{n}\right)-m_{1 j} t\right\rceil \\
f_{2}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)-f_{2}\left(x_{1}, x_{2}, \ldots, x_{n}\right)-m_{2 j} t \\
\vdots \\
f_{p}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)-f_{p}\left(x_{1}, x_{2}, \ldots, x_{n}\right)-m_{p j} t
\end{array}\right.\right] \\
&=\left|\begin{array}{l}
\frac{f_{1}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)-f_{1}\left(x_{1}, x_{2}, \ldots, x_{n}\right)}{t}-m_{1 j} \\
\frac{f_{2}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)-f_{2}\left(x_{1}, x_{2}, \ldots, x_{n}\right)}{t}-m_{2 j} \\
\vdots \\
\frac{f_{p}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)-f_{p}\left(x_{1}, x_{2}, \ldots, x_{n}\right)}{t}-m_{p j}
\end{array}\right|
\end{aligned}
$$

Meditate on this vector. For each component,

$$
\begin{array}{r}
\lim _{t \rightarrow 0} \frac{f_{i}\left(x_{1}, x_{2}, \ldots, x_{j}+t, \ldots, x_{n}\right)-f_{i}\left(x_{1}, x_{2}, \ldots, x_{n}\right)}{t} \\
=\left.\frac{d}{d s} f_{i}\left(x_{1}, x_{2}, \ldots, s, \ldots, x_{n}\right)\right|_{s=x_{j}}
\end{array}
$$

This derivative has a name. It is called the partial derivative of $\boldsymbol{f}_{\boldsymbol{i}}$ with respect to the $\boldsymbol{j}^{\text {th }}$ variable. There are many different notations for the partial derivatives of a function $g\left(x_{1}, x_{2}, \ldots, x_{n}\right)$. The two most common are:

$$
\begin{aligned}
& g_{, j}\left(x_{1}, x_{2}, \ldots, x_{n}\right) \\
& \frac{\partial}{\partial x_{j}} g\left(x_{1}, x_{2}, \ldots, x_{n}\right)
\end{aligned}
$$

The requirement that $\lim _{\boldsymbol{h} \rightarrow 0} \frac{1}{|\boldsymbol{h}|}\left[f\left(\boldsymbol{x}_{0}+\boldsymbol{h}\right)-f\left(\boldsymbol{x}_{0}\right)-\boldsymbol{L}(\boldsymbol{h})\right]=\mathbf{0}$ now translates into

$$
m_{i j}=\frac{\partial f_{i}}{\partial x_{j}},
$$

and, mirabile dictu, we have found the matrix $\boldsymbol{L}$ !

## Example

Let $f: \boldsymbol{R}^{2} \rightarrow \boldsymbol{R}^{2}$ be given by $f\left(x_{1} \cdot x_{2}\right)=\left[\begin{array}{l}3 x_{1} \sin x_{2} \\ x_{1}^{3}+x_{1} x_{2}^{2}\end{array}\right]$. Assume $f$ is differentiable and let's find the derivative (more precisely, the matrix of the derivative. This matrix will, of course, be $2 \times 2: \quad \boldsymbol{L}=\left[\begin{array}{ll}m_{11} & m_{12} \\ m_{21} & m_{22}\end{array}\right]$. Now

$$
\begin{aligned}
& f_{1}\left(x_{1}, x_{2}\right)=3 x_{1} \sin x_{2}, \text { and } \\
& f_{2}\left(x_{1}, x_{2}\right)=x_{1}^{3}+x_{1} x_{2}^{2}
\end{aligned}
$$

Compute the partial derivatives:

$$
\begin{aligned}
& \frac{\partial f_{1}}{\partial x_{1}}=3 \sin x_{2} \\
& \frac{\partial f_{2}}{\partial x_{1}}=3 x_{1}^{2}+x_{2}^{2}
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{\partial f_{1}}{\partial x_{2}}=3 x_{1} \cos x_{2} \\
& \frac{\partial f_{2}}{\partial x_{2}}=2 x_{1} x_{2}
\end{aligned} .
$$

The derivative is thus

$$
\boldsymbol{L}=\left[\begin{array}{cc}
3 \sin x_{2} & 3 x_{1} \cos x_{2} \\
3 x_{1}^{2}+x_{2}^{2} & 2 x_{1} x_{2}
\end{array}\right]
$$

We now know how to find the derivative of $f$ at $\boldsymbol{x}$ if we know the derivative exists; but how do we know when there is a derivative? The function $f$ is differentiable at $\boldsymbol{x}$ if the partial derivatives exist and are continuous. It should be noted that it is not sufficient just for the partial derivatives to exist.

## Exercises

1. Find all partial derivatives of the given functions:
a) $f(x, y)=x^{2} y^{3}$
b) $f(x, y, z)=x^{2} y z+z \cos (x y)$
c) $g\left(x_{1}, x_{2}, x_{3}\right)=x_{1} x_{2} x_{3}+x_{2}$
d) $h\left(x_{1}, x_{2}, x_{3}, x_{4}\right)=\frac{x_{3} \sin \left(e^{x_{1}}\right)}{x_{2}+x_{4}}$
2. Find the derivative of the linear function whose matrix is $\left[\begin{array}{ccc}1 & 3 & 2 \\ -2 & 7 & 0\end{array}\right]$.
3. What is the derivative a linear function whose matrix is $\boldsymbol{A}$ ?
4. Find the derivative of $\boldsymbol{R}(t)=\cos t \boldsymbol{i}+\sin t \boldsymbol{j}+t \boldsymbol{k}$.
5. Find the derivative of $f(x, y)=x^{2} y^{3}$.
6. Find the derivative of

$$
f\left(x_{1}, x_{2}, x_{3}\right)=\left[\begin{array}{c}
x_{1} x_{3}+e^{x_{2}} \\
x_{3} \log \left(x_{1}+x_{2}^{2}\right) \\
x_{2} \\
x_{1} x_{3}^{2}+5
\end{array}\right]
$$

### 7.3 The Chain Rule

Recall from elementary one dimensional calculus that if a function is differentiable at a point, it is also continuous there. The same is true here in the more general setting of functions $f: \boldsymbol{R}^{\boldsymbol{n}} \rightarrow \boldsymbol{R}^{p}$. Let's see why this is so. Suppose $f$ is differentiable at $\boldsymbol{a}$ with derivative $\boldsymbol{L}$. Let $\boldsymbol{h}=\boldsymbol{x}-\boldsymbol{a}$. Then $\lim _{\boldsymbol{x} \rightarrow a} f(\boldsymbol{x})=\lim _{h \rightarrow 0} f(\boldsymbol{a}+\boldsymbol{h})$. Now,

$$
f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a})=|\boldsymbol{h}|\left[\frac{f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a})-\boldsymbol{L}(h)}{|\boldsymbol{h}|}\right]-\boldsymbol{L}(\boldsymbol{h})
$$

Now look at the limit of this as $|\boldsymbol{h}| \rightarrow 0$ :

$$
\lim _{\boldsymbol{h} \rightarrow 0}\left[\frac{f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a})-\boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right]=0
$$

because $f$ is differentiable at $\boldsymbol{a}$, and $\lim _{\boldsymbol{h} \rightarrow \boldsymbol{0}} \boldsymbol{L}(\boldsymbol{h})=\boldsymbol{L}(\boldsymbol{0})=\mathbf{0}$ because the linear function $\boldsymbol{L}$ is continuous. Thus $\lim _{\boldsymbol{h} \rightarrow 0}(f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a}))=\boldsymbol{0}$, or $\lim _{\boldsymbol{h} \rightarrow 0} f(\boldsymbol{a}+\boldsymbol{h})=f(\boldsymbol{a})$, which means $f$ is continuous at $\boldsymbol{a}$.

Next, let's see what the celebrated chain rule looks like in higher dimensions. Let $f: \boldsymbol{R}^{n} \rightarrow \boldsymbol{R}^{p}$ and $g: \boldsymbol{R}^{p} \rightarrow \boldsymbol{R}^{q}$. Suppose the derivative of $f$ at $\boldsymbol{a}$ is $\boldsymbol{L}$ and the derivative of $g$ at $f(\boldsymbol{a})$ is $\boldsymbol{M}$. We go on a quest for the derivative of the composition $g \circ f: \boldsymbol{R}^{n} \rightarrow \boldsymbol{R}^{q}$ at $\boldsymbol{a}$. Let $r=g \circ f$, and look at $r(\boldsymbol{a}+\boldsymbol{h})-r(\boldsymbol{a})=g(f(\boldsymbol{a}+\boldsymbol{h}))-g(f(\boldsymbol{a}))$. Next, let $\boldsymbol{k}=f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a})$. Then we may write

$$
\begin{aligned}
r(\boldsymbol{a}+\boldsymbol{h})-r(\boldsymbol{a})-\boldsymbol{M L}(\boldsymbol{h}) & =g(f(\boldsymbol{a}+\boldsymbol{h}))-g(f(\boldsymbol{a}))-\boldsymbol{M L}(\boldsymbol{h}) \\
& =g(f(\boldsymbol{a})+\boldsymbol{k})-g(f(\boldsymbol{a}))-\boldsymbol{M}(\boldsymbol{k})+\boldsymbol{M}(\boldsymbol{k})-\boldsymbol{M L}(\boldsymbol{h}) . \\
& =g(f(\boldsymbol{a})+\boldsymbol{k})-g(f(\boldsymbol{a}))-\boldsymbol{M}(\boldsymbol{k})+\boldsymbol{M}(\boldsymbol{k}-\boldsymbol{L}(\boldsymbol{h}))
\end{aligned}
$$

Thus,

$$
\frac{r(\boldsymbol{a}+\boldsymbol{h})-r(\boldsymbol{a})-\boldsymbol{M L}(\boldsymbol{h})}{|\boldsymbol{h}|}=\frac{g(f(\boldsymbol{a})+\boldsymbol{k})-g(f(\boldsymbol{a}))-\boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{h}|}+\boldsymbol{M}\left(\frac{\boldsymbol{k}-\boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right)
$$

Now we are ready to see what happens as $|\boldsymbol{h}| \rightarrow 0$. look at the second term first:

$$
\begin{aligned}
\lim _{\boldsymbol{h} \rightarrow 0} \boldsymbol{M}\left(\frac{\boldsymbol{k}-\boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right) & =\lim _{\boldsymbol{h} \rightarrow 0} \boldsymbol{M}\left(\frac{f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a})-\boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right)=\boldsymbol{M}\left(\lim _{\boldsymbol{h} \rightarrow 0}\left(\frac{f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a})-\boldsymbol{L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right)\right. \\
& =\boldsymbol{M}(\boldsymbol{0})=\mathbf{0}
\end{aligned}
$$

since $\boldsymbol{L}$ is the derivative of $f$ at $\boldsymbol{a}$ and $\boldsymbol{M}$ is linear, and hence continuous.
Now we need to see what happens to the term

$$
\lim _{\boldsymbol{h} \rightarrow 0}\left(\frac{g(f(\boldsymbol{a})+\boldsymbol{k})-g(f(\boldsymbol{a}))-\boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{h}|}\right)
$$

This is a bit tricky. Note first that because $f$ is differentiable at $\boldsymbol{a}$, we know that

$$
\frac{|\boldsymbol{k}|}{|\boldsymbol{h}|}=\frac{|f(\boldsymbol{a}+\boldsymbol{h})-f(\boldsymbol{a})|}{|\boldsymbol{h}|}
$$

behaves nicely as $|\boldsymbol{h}| \rightarrow 0$. Next,

$$
\begin{aligned}
& \lim _{h \rightarrow 0}\left(\frac{g(f(\boldsymbol{a})+\boldsymbol{k})-g(f(\boldsymbol{a}))-\boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{h}|} \cdot \frac{|\boldsymbol{k}|}{|\boldsymbol{k}|}\right) \\
& =\lim _{h \rightarrow 0}\left(\frac{g(f(\boldsymbol{a})+\boldsymbol{k})-g(f(\boldsymbol{a}))-\boldsymbol{M}(\boldsymbol{k})}{|\boldsymbol{k}|}\right) \frac{|\boldsymbol{k}|}{|\boldsymbol{h}|}=\boldsymbol{0}
\end{aligned}
$$

since the derivative of $g$ at $f(\boldsymbol{a})$ is $\boldsymbol{M}$, and $\frac{|\boldsymbol{k}|}{|\boldsymbol{h}|}$ is well-behaved. Finally at last, we have shown that

$$
\lim _{h \rightarrow 0}\left(\frac{r(\boldsymbol{a}+\boldsymbol{h})-r(\boldsymbol{a})-\boldsymbol{M L}(\boldsymbol{h})}{|\boldsymbol{h}|}\right)=\mathbf{0},
$$

which means the derivative of the composition $r=g \circ f$ is simply the composition, or matrix product, of the derivatives. What could be more pleasing from an esthetic point of view!

## Example

Let $f(t)=\left(t^{2}, 1+t^{3}\right)$ and $g\left(x_{1}, x_{2}\right)=\left(2 x_{1}-x_{2}\right)^{3}$, and let $r=g \circ f$. First, we shall find the derivative of $r$ at $t=2$ using the Chain Rule. The derivative of $f$ is

$$
\boldsymbol{L}=\left[\begin{array}{c}
2 t \\
3 t^{2}
\end{array}\right]
$$

and the derivative of $g$ is

$$
\boldsymbol{M}=\left[\begin{array}{ll}
6\left(2 x_{1}-x_{2}\right)^{2} & -3\left(2 x_{1}-x_{2}\right)^{2}
\end{array}\right] .
$$

At $t=2, \quad \boldsymbol{L}=\left[\begin{array}{c}4 \\ 12\end{array}\right]$; and at $g(f(2))=g(4,9), \boldsymbol{M}=\left[\begin{array}{ll}6 & -3\end{array}\right]$. Thus the derivative of the composition is $\boldsymbol{M L}=\left[\begin{array}{ll}6 & -3\end{array}\right]\left[\begin{array}{c}4 \\ 12\end{array}\right]=[-12]=-12$.

Now for fun, let's find an explicit recipe for $r$ and differentiate:
$r(t)=g(f(t))=g\left(t^{2}, 1+t^{3}\right)=\left(2 t^{2}-1-t^{3}\right)^{3}$. Thus $\quad r^{\prime}(t)=3\left(2 t^{2}-1-t^{3}\right)^{2}\left(4 t-3 t^{2}\right)$, and so $r^{\prime}(2)=3(1)(8-12)=-12$. It is, of course, very comforting to get the same answer as before.

There are several different notations for the matrix of the derivative of $f: \boldsymbol{R}^{n} \rightarrow \boldsymbol{R}^{p}$ at $\boldsymbol{x} \in \boldsymbol{R}^{\boldsymbol{n}} \quad$ The most usual is simply $f^{\prime}(\boldsymbol{x})$.

## Exercises

7. Let $g\left(x_{1}, x_{2}, x_{3}\right)=\left(x_{1} x_{3}, x_{2} x_{3}+1\right)$ and $f\left(x_{1}, x_{2}\right)=\left(x_{1} x_{2} \sin x_{1}, x_{1}+3 x_{2}, x_{2}-2 x_{1}^{2}\right)$.

Find the derivative of $g \circ f$ at $(2,-4)$.
8. Let $u(x, y, z)=\left(x+y^{2}, 2 x y, x \sin y, x^{3} y^{2}\right)$ and $v(r, s, t, q)=\left(r+s-q^{3},(r-t) e^{s}\right)$. a)Which, if either, of the composition functions $u \circ v$ or $v \circ u$ is defined? Explain. b)Find the derivative of your answer to part a).
9. Let $f(x, y)=\left(e^{(x+y)}, e^{(x-y)}\right)$ and $g(x, y)=\left(x-y^{3}, x^{2}+y\right)$.
a)Find the derivative of $f \circ g$ at the point $(1,-2)$.
b)Find the derivative of $g \circ f$ at the point $(1,-2)$.
c) Find the derivative of $f \circ f$ at the point $(1,-2)$.
d) Find the derivative of $g \circ g$ at the point $(1,-2)$.
10. Suppose $r=t^{2} \cos t$ and $t=x^{2}-3 y^{2}$. Find the partial derivatives $\frac{\partial r}{\partial x}$ and $\frac{\partial r}{\partial y}$.

### 7.4 More Chain Rule Stuff

In the everyday cruel world, we seldom compute the derivative of the composition of two functions by explicitly multiplying the two derivative matrices. Suppose, as usual, we have $r=g \circ f: \boldsymbol{R}^{\boldsymbol{n}} \rightarrow \boldsymbol{R}^{q}$. The the derivative is, as we now know,

$$
r^{\prime}(\boldsymbol{x})=r^{\prime}\left(x_{1}, x_{2}, \ldots, x_{n}\right)=\left[\begin{array}{cccc}
\frac{\partial r_{1}}{\partial \mathrm{x}_{1}} & \frac{\partial r_{1}}{\partial \mathrm{x}_{2}} & \cdots & \frac{\partial r_{1}}{\partial \mathrm{x}_{n}} \\
\frac{\partial r_{2}}{\partial x_{1}} & \frac{\partial r_{2}}{\partial x_{2}} & \cdots & \frac{\partial r_{2}}{\partial x_{n}} \\
\vdots & & & \\
\frac{\partial r_{p}}{\partial \mathrm{x}_{1}} & \frac{\partial r_{p}}{\partial \mathrm{x}_{2}} & \cdots & \frac{\partial r_{p}}{\partial \mathrm{x}_{n}}
\end{array}\right] .
$$

We can thus find the derivative using the Chain Rule only in the very special case in which the compsite function is real valued. Specifically, suppose $g: \boldsymbol{R}^{p} \rightarrow \boldsymbol{R}$ and $f: \boldsymbol{R}^{\boldsymbol{n}} \rightarrow \boldsymbol{R}^{p}$. Let $r=g \circ f . \quad$ Then $r$ is simply a real-valued function of $\boldsymbol{x}=\left(x_{1}, x_{2}, \ldots, x_{n}\right)$. Let's use the Chain Rule to find the partial derivatives.

$$
r^{\prime}(\boldsymbol{x})=\left[\begin{array}{llll}
\frac{\partial r}{\partial x_{1}} & \frac{\partial r}{\partial x_{2}} & \cdots & \frac{\partial r}{\partial x_{n}}
\end{array}\right]=\left[\begin{array}{llll}
\frac{\partial g}{\partial y_{1}} & \frac{\partial g}{\partial y_{2}} & \cdots & \frac{\partial g}{\partial y_{p}}
\end{array}\right]\left[\begin{array}{cccc}
\frac{\partial \mathrm{f}_{1}}{\partial x_{1}} & \frac{\partial \mathrm{f}_{1}}{\partial x_{2}} & \cdots & \frac{\partial \mathrm{f}_{1}}{\partial x_{n}} \\
\frac{\partial \mathrm{f}_{2}}{\partial x_{1}} & \frac{\partial \mathrm{f}_{2}}{\partial x_{2}} & \cdots & \frac{\partial \mathrm{f}_{2}}{\partial x_{n}} \\
\vdots & & \\
\frac{\partial \mathrm{f}_{\mathrm{p}}}{\partial x_{1}} & \frac{\partial \mathrm{f}_{\mathrm{p}}}{\partial x_{2}} & \cdots & \frac{\partial \mathrm{f}_{\mathrm{p}}}{\partial x_{n}}
\end{array}\right]
$$

Thus makes it clear that

$$
\frac{\partial r}{\partial x_{j}}=\frac{\partial \mathrm{g}}{\partial y_{1}} \frac{\partial f_{1}}{\partial x_{j}}+\frac{\partial \mathrm{g}}{\partial y_{2}} \frac{\partial f_{2}}{\partial x_{j}}+\cdots+\frac{\partial \mathrm{g}}{\partial y_{p}} \frac{\partial f_{p}}{\partial x_{j}} .
$$

Frequently, engineers and other malefactors do not use a different name for the composition $g \circ f$, and simply use the name $g$ to denote both the composition $g \circ f\left(x_{1}, x_{2}, \ldots, x_{n}\right)=g\left(f_{1}\left(x_{1}, x_{2}, \ldots, x_{n}\right), f_{2}\left(x_{1}, x_{2}, \ldots, x_{n}\right), \ldots, f_{p}\left(x_{1}, x_{2}, \ldots, x_{n}\right)\right)$ and the function $g$ given by $g(\boldsymbol{y})=g\left(y_{1}, y_{2}, \ldots, y_{p}\right)$. Since $y_{j}=f_{j}\left(x_{1}, x_{2}, \ldots, x_{n}\right)$, these same folks also frequently just use $y_{j}$ to denote the function $f_{j}$. The Chain Rule given above then looks even nicer:

$$
\frac{\partial \mathrm{g}}{\partial x_{j}}=\frac{\partial \mathrm{g}}{\partial y_{1}} \frac{\partial y_{1}}{\partial x_{j}}+\frac{\partial \mathrm{g}}{\partial y_{2}} \frac{\partial y_{2}}{\partial x_{j}}+\cdots+\frac{\partial \mathrm{g}}{\partial y_{p}} \frac{\partial y_{p}}{\partial x_{j}} .
$$

## Example

Suppose $g(x, y, z)=x^{2} y+y e^{z}$ and $x=s+t, y=s t^{3}$, and $z=s^{2}+3 t^{2}$. Let us find the partial derivatives $\frac{\partial g}{\partial r}$ and $\frac{\partial g}{\partial t}$. We know that

$$
\begin{aligned}
\frac{\partial g}{\partial s} & =\frac{\partial g}{\partial x} \frac{\partial x}{\partial s}+\frac{\partial g}{\partial y} \frac{\partial y}{\partial s}+\frac{\partial g}{\partial z} \frac{\partial z}{\partial s} \\
& =2 x y(1)+\left(x^{2}+e^{z}\right) t^{3}+y e^{z}(2 s) \\
& =2 x y+\left(x^{2}+e^{z}\right) t^{3}+2 s y e^{z}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\frac{\partial g}{\partial t} & =\frac{\partial g}{\partial x} \frac{\partial x}{\partial \mathrm{t}}+\frac{\partial g}{\partial y} \frac{\partial y}{\partial t}+\frac{\partial g}{\partial z} \frac{\partial z}{\partial t} \\
& =2 x y(1)+\left(x^{2}+e^{z}\right) 3 s t^{2}+y e^{z}(6 t) \\
& =2 x y+3\left(x^{2}+e^{z}\right) s t^{2}+6 t y e^{z}
\end{aligned}
$$

These notational shortcuts are fine and everyone uses them; you should, however, be aware that it is a practice sometimes fraught with peril. Suppose, for instance, you have $g(x, y, z)=x^{2}+y^{2}+z^{2}$, and $x=t+z, y=t^{2}+2 z$, and $z=t^{3}$. Now it is not at all clear what is meant by the symbol $\frac{\partial \mathrm{g}}{\partial z}$. Meditate on this.

## Exercises

11. Suppose $g(x, y)=f(x-y, y-s)$. Find $\frac{\partial g}{\partial x}+\frac{\partial g}{\partial y}$.
12. Suppose the temperature $T$ at the point $(x, y, z)$ in space is given by the function $T(x, y, z)=x^{2}+x y z-z y^{2}$. Find the derivative with respect to $t$ of a particle moving along the curve described by $\boldsymbol{r}(t)=\cos \boldsymbol{i}+\sin t \boldsymbol{j}+3 \boldsymbol{t} \boldsymbol{k}$.
13. Suppose the temperature $T$ at the point $(x, y, z)$ in space is given by the function $T(x, y, z)=x^{2}+y^{2}+z^{2}$. A particle moves along the curve described by $\boldsymbol{r}(t)=\sin \pi t i+\cos \pi t \boldsymbol{j}+\left(t^{2}-2 t+2\right) \boldsymbol{k}$. Find the coldest point on the trajectory.
14. Let $r(x, y)=f(x) g(y)$, and suppose $x=t$ and $y=t$. Use the Chain Rule to find $\frac{d r}{d t}$.
